

Oracle® Exadata Database Machine Maintenance Guide



22.1
F29250-25
September 2022

The Oracle logo, consisting of a solid red square with the word "ORACLE" in white, uppercase, sans-serif font centered within it.

ORACLE®

Oracle Exadata Database Machine Maintenance Guide, 22.1

F29250-25

Copyright © 2008, 2022, Oracle and/or its affiliates.

Primary Author: Peter Fusek

Contributing Authors: James Spiller

Contributors: Doug Archambault, Leo Agranonik, Andrew Babb, Dib Chatterjee, Craig Brown, Nilesh Choudhury, Henry Chow, Ravindra Dani, Boris Erlikhman, Jaime Figueroa, Roger Hansen, Leslie Keller, Frank Kobylanski, René Kundersma, Sue Lee, Holger Leister, Jennifer Leung, Yang Liu, Juan Loaiza, Barb Lundhild, Catherine Luu, Takeshi Maruyama, Philip Newlan, Dan Norris, Michael Nowak, Bharat Paliwal, Umesh Panchaksharaiah, Gavin Parish, Hector Pujol, Darryl Presley, Ashish Ray, Richard Scales, Oliver Sharwood, Jia Shi, Kesavan Srinivasan, Mahesh Subramaniam, Krishnadev Telikicherla, Cliff Thomas, Lawrence To, Alex Tsukerman, Shreyas Udgaonkar, Kothanda Umamageswaran, Doug Utzig, Jim Viscusi, Vern Wagman, Zheren Zhang, Alex Blyth

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, then the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs (including any operating system, integrated software, any programs embedded, installed or activated on delivered hardware, and modifications of such programs) and Oracle computer documentation or other Oracle data delivered to or accessed by U.S. Government end users are "commercial computer software" or "commercial computer software documentation" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, reproduction, duplication, release, display, disclosure, modification, preparation of derivative works, and/or adaptation of i) Oracle programs (including any operating system, integrated software, any programs embedded, installed or activated on delivered hardware, and modifications of such programs), ii) Oracle computer documentation and/or iii) other Oracle data, is subject to the rights and limitations specified in the license contained in the applicable contract. The terms governing the U.S. Government's use of Oracle cloud services are defined by the applicable contract for such services. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle, Java, and MySQL are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Inside are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Epyc, and the AMD logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.

Contents

Preface

Audience	xxiii
Documentation Accessibility	xxiii
Diversity and Inclusion	xxiii
Related Documentation	xxiv
Conventions	xxvi

1 General Maintenance Information

1.1 Overview of Roles and Responsibilities	1-2
1.1.1 Common Administrative Tasks for Exadata Database Machine Management	1-2
1.1.2 Understanding the Administrative Differences with Exadata Database Machine	1-4
1.2 Powering On and Off Oracle Exadata Rack	1-5
1.2.1 Non-emergency Power Procedures	1-5
1.2.1.1 Powering On Oracle Exadata Rack	1-5
1.2.1.2 Powering On Servers Remotely using ILOM	1-6
1.2.1.3 Powering Off Oracle Exadata Rack	1-6
1.2.1.4 Powering On and Off Network Switches	1-9
1.2.2 Emergency Power-off Considerations	1-9
1.2.2.1 Emergency Power-off Procedure	1-9
1.2.2.2 Emergency Power-off Switch	1-9
1.2.3 Cautions and Warnings	1-9
1.3 Using Auto Service Request to Manage Hardware Faults	1-9
1.3.1 Understanding Auto Service Request	1-10
1.3.2 Installing and Configuring ASR	1-12
1.4 Monitoring the System Using Oracle Enterprise Manager Cloud Control	1-13
1.5 Monitoring the System Using Oracle Configuration Manager	1-14
1.6 Changing Component Passwords	1-14
1.6.1 Changing the Database Server Passwords	1-15
1.6.1.1 Changing the User Account Password on the Database Server	1-15
1.6.1.2 Changing the GRUB Account Password on the Database Server	1-15
1.6.2 Changing the Exadata Storage Server Passwords	1-16
1.6.3 Changing the Power Distribution Unit Password	1-16

1.6.4	Changing the ILOM Password	1-16
1.6.5	Changing the InfiniBand Switch Password	1-17
1.6.6	Changing the Cisco Management Network Switch or RoCE Network Fabric Switch Password	1-17
1.6.6.1	Changing the Password for Cisco 93xx Switches	1-17
1.6.6.2	Changing the Cisco 4948 Ethernet Switch Password	1-18
1.6.7	Changing the KVM Password	1-20
1.7	Determining the Server Model	1-21
1.8	Monitoring Ambient Temperature of Servers	1-22
1.9	Replacing a Disk Controller Battery Backup Unit	1-23
1.9.1	Replacing a Disk Controller BBU on a Database Server	1-23
1.9.1.1	Step 1: Prepare the Disk Controller BBU for Removal	1-24
1.9.1.2	Step 2: Replace the Disk Controller BBU	1-28
1.9.1.3	Step 3: Enable and Verify the New Disk Controller BBU	1-29
1.9.2	Replacing a Disk Controller BBU on an Exadata Storage Server	1-33
1.9.2.1	Step 1: Prepare the Disk Controller BBU for Removal	1-33
1.9.2.2	Step 2: Replace the Disk Controller BBU	1-37
1.9.2.3	Step 3: Enable the New Disk Controller BBU	1-38
1.10	Overview of the dbmsrv Service	1-40
1.10.1	Using a Script to Change User IDs and Group IDs for dbmsrv	1-41
1.10.2	Manually Changing User IDs and Group IDs for dbmsrv	1-42
1.11	Configuring Password Expiration for Users Accessing the Server Remotely	1-44
1.12	State of Storage Server and Database Servers During Configuration Changes	1-45
1.13	Rescue Plan	1-46
1.14	Using ExaWatcher Charts to Monitor Exadata Database Machine Performance	1-48
1.14.1	About ExaWatcher Charts	1-48
1.14.2	Requirements for Using ExaWatcher Charts	1-50
1.14.3	IO Charts	1-50
1.14.3.1	IO Stat Summary	1-50
1.14.3.2	I/O Stat Detail	1-52
1.14.4	CPU Charts	1-53
1.14.5	CPU Detail	1-54
1.14.6	Cell Server Charts	1-55
1.14.7	Alert History	1-56

2 Maintaining Exadata Database Servers

2.1	Management Server on Database Servers	2-2
2.2	Maintaining the Local Storage on Exadata Database Servers	2-2
2.2.1	Verifying the Database Server Configuration	2-3
2.2.1.1	About the RAID Storage Configuration	2-3

2.2.1.2	Verifying Disk Controller Configuration on Oracle Exadata Database Machine X7-8 or Later Systems	2-4
2.2.1.3	Verifying Disk Controller Configuration on Oracle Exadata Database Machine X6-8 and Earlier	2-5
2.2.1.4	Verifying Virtual Drive Configuration	2-7
2.2.1.5	Verifying Physical Drive Configuration	2-8
2.2.2	Monitoring a Database Server RAID Set Rebuilding	2-8
2.2.3	Reclaiming a Hot Spare Drive After Upgrading to Oracle Exadata System Software Release 12.1.2.1.0 or Later	2-9
2.2.4	Understanding Automated File Deletion Policy	2-14
2.3	Maintaining Flash Disks on Exadata Database Servers	2-15
2.3.1	Monitoring the Status of Flash Disks	2-15
2.3.2	Performing a Hot-Pluggable Replacement of a Flash Disk	2-16
2.4	Adding the Disk Expansion Kit to Database Servers	2-17
2.4.1	Adding the Disk Expansion Kit to Database Servers: X9M-2	2-17
2.4.2	Adding the Disk Expansion Kit to Database Servers: X8M-2 and Prior	2-18
2.5	Adding Memory Expansion Kit to Database Servers	2-23
2.6	Verifying and Modifying the Link Speed on the Client Network Ports for X7 and Later Systems	2-25
2.7	Adding and Configuring an Extra Network Card on Exadata Database Machine X6-2 and Later	2-27
2.7.1	Viewing the Network Interfaces	2-29
2.7.2	Configuring the Additional Network Card for a Non-Oracle VM Environment	2-31
2.7.3	Configuring the Additional Network Card for an Oracle VM Environment	2-34
2.8	Increasing the Number of Active Cores on Database Servers	2-36
2.9	Extending LVM Partitions	2-38
2.9.1	Extending the root LVM Partition	2-39
2.9.1.1	Extending the root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later	2-39
2.9.1.2	Extending the root LVM Partition on Systems Running Oracle Exadata System Software Earlier than Release 11.2.3.2.1	2-42
2.9.2	Resizing a Non-root LVM Partition	2-45
2.9.2.1	Extending a Non-root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later	2-45
2.9.2.2	Extending a Non-root LVM Partition on Systems Running Oracle Exadata System Software Earlier than Release 11.2.3.2.1	2-47
2.9.2.3	Reducing a Non-root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later	2-51
2.9.3	Extending the Swap Partition	2-55
2.10	Creating a Snapshot-Based Backup of Oracle Linux Database Server	2-57
2.10.1	Creating a Snapshot-Based Backup of Exadata Database Servers X8M and Later with Uncustomized Partitions	2-57
2.10.2	Creating a Snapshot-Based Backup of Exadata X8 or Earlier Database Servers with Uncustomized Partitions	2-61

2.10.3	Creating a Snapshot-Based Backup of Oracle Linux Database Server with Customized Partitions	2-64
2.11	Recovering Oracle Linux Database Servers Using a Snapshot-Based Backup	2-64
2.11.1	Overview of Snapshot-Based Recovery of Database Servers	2-65
2.11.2	Recovering Oracle Linux Database Server with Uncustomized Partitions	2-66
2.11.3	Recovering Exadata X9M Database Servers with Customized Partitions	2-70
2.11.4	Recovering Exadata X8M Database Servers with Customized Partitions	2-76
2.11.5	Recovering Exadata Database Servers X7 or X8 with Customized Partitions	2-82
2.11.6	Recovering Exadata X6 or Earlier Database Servers with Customized Partitions	2-88
2.11.7	Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery	2-93
2.11.7.1	Configuring Eighth Rack On X3-2 or Later Machines Running Oracle Exadata Storage Server Release 12.1.2.3.0 or Later	2-94
2.11.7.2	Configuring Eighth Rack On X3-2 Machines Running Oracle Exadata Storage Server Release 12.1.2.2.3 or Earlier	2-95
2.12	Re-Imaging the Oracle Exadata Database Server	2-96
2.12.1	Contact Oracle Support Services	2-97
2.12.2	Download Latest Release of Cluster Verification Utility	2-97
2.12.3	Remove the Database Server from the Cluster	2-97
2.12.4	Image the Database Server	2-100
2.12.5	Configure the Re-imaged Database Server	2-100
2.12.6	Prepare the Re-imaged Database Server for the Cluster	2-101
2.12.7	Apply Oracle Exadata System Software Patch Bundles to the Replacement Database Server	2-103
2.12.8	Clone Oracle Grid Infrastructure to the Replacement Database Server	2-104
2.12.9	Clone Oracle Database Homes to the Replacement Database Server	2-108
2.13	Changing Existing Elastic Configurations for Database Servers	2-110
2.13.1	Adding a New Database Server to the Cluster	2-110
2.13.2	Moving an Existing Database Server to a Different Cluster	2-111
2.13.3	Dropping a Database Server from an Oracle RAC Cluster	2-112
2.14	Managing Quorum Disks for High Redundancy Disk Groups	2-112
2.14.1	Using Quorum Disks to Increase Fault Tolerance	2-113
2.14.2	Overview of Quorum Disk Manager	2-115
2.14.3	Software Requirements for Quorum Disk Manager	2-115
2.14.4	quorumdiskmgr Reference	2-115
2.14.4.1	Syntax for the Quorum Disk Manager Utility	2-116
2.14.4.2	quorumdiskmgr Objects	2-117
2.14.4.3	Creating a Quorum Disk Configuration (--create --config)	2-117
2.14.4.4	Creating a Target (--create --target)	2-118
2.14.4.5	Creating a Device (--create --device)	2-119
2.14.4.6	Listing Quorum Disk Configurations (--list --config)	2-119
2.14.4.7	Listing Targets (--list --target)	2-120

2.14.4.8	Listing Devices (--list --device)	2-121
2.14.4.9	Deleting Configurations (--delete --config)	2-122
2.14.4.10	Deleting Targets (--delete --target)	2-122
2.14.4.11	Deleting Devices (--delete --device)	2-122
2.14.4.12	Changing Owner and Group Values (--alter --config)	2-123
2.14.4.13	Changing the RDMA Network Fabric IP Addresses (--alter --target)	2-124
2.14.5	Add Quorum Disks to Database Nodes	2-124
2.14.6	Recreate Quorum Disks	2-132
2.14.7	Use Cases	2-132
2.14.7.1	New Deployments on Oracle Exadata 12.1.2.3.0 or Later	2-132
2.14.7.2	Upgrading to Oracle Exadata Release 12.1.2.3.0 or Later	2-133
2.14.7.3	Downgrading to a Pre-12.1.2.3.0 Oracle Exadata Release	2-133
2.14.7.4	Managing Quorum Disks When Changing Elastic Configurations	2-133
2.14.8	Reconfigure Quorum Disk After Restoring a Database Server	2-141
2.15	Using vmetrics	2-143
2.15.1	About the vmetrics Package	2-144
2.15.2	Installing and Starting the vmetrics Service	2-144
2.15.3	Files in the vmetrics Package	2-145
2.15.4	Displaying the Statistics	2-145
2.15.5	Adding Metrics to vmetrics	2-146
2.16	Using FIPS mode	2-147
2.17	LED Status Descriptions	2-148
2.17.1	Exadata Database Server X5-2 and Later LEDs	2-148
2.17.2	Sun Server X4-2 Oracle Database Server LEDs	2-149
2.17.3	Sun Server X3-2 Oracle Database Server LEDs	2-150
2.17.4	Sun Fire X4170 M2 Oracle Database Server LEDs	2-151
2.17.5	Sun Fire X4170 Oracle Database Server LEDs	2-152
2.17.6	Exadata Database Server X5-8 and Later LEDs	2-153
2.17.7	Sun Server X4-8 Oracle Database Server LEDs	2-154
2.17.8	Sun Server X2-8 Oracle Database Server LEDs	2-156
2.17.9	Sun Fire X4800 Oracle Database Server LEDs	2-157
2.18	Exadata Database Server Images	2-159
2.18.1	Oracle Server X9-2 Database Server Images	2-159
2.18.2	Oracle Server X8-2 Database Server Images	2-160
2.18.3	Oracle Server X7-2 Oracle Database Server Images	2-160
2.18.4	Oracle Server X6-2 Oracle Database Server Images	2-161
2.18.5	Oracle Server X5-2 Oracle Database Server Images	2-161
2.18.6	Sun Server X4-2 Oracle Database Server Images	2-162
2.18.7	Sun Server X3-2 Oracle Database Server Images	2-162
2.18.8	Sun Fire X4170 M2 Oracle Database Server Images	2-163
2.18.9	Sun Fire X4170 Oracle Database Server Images	2-163

2.18.10	Oracle Server X8-8 Database Server Images	2-164
2.18.11	Oracle Server X7-8 Oracle Database Server Images	2-165
2.18.12	Oracle Server X5-8 and X6-8 Oracle Database Server Images	2-166
2.18.13	Sun Server X4-8 Oracle Database Server Images	2-167
2.18.14	Sun Server X2-8 Oracle Database Server Images	2-168
2.18.15	Sun Fire X4800 Oracle Database Server Images	2-169

3 Maintaining Oracle Exadata Storage Servers

3.1	Maintaining Oracle Exadata Storage Servers	3-2
3.1.1	Shutting Down Exadata Storage Server	3-2
3.1.2	Checking Status of a Rebalance Operation	3-4
3.1.3	Enabling Network Connectivity with the Diagnostic ISO	3-4
3.2	Using Exadata Extended (XT) Storage Servers	3-5
3.2.1	About Oracle Exadata Extended (XT) Storage Servers	3-5
3.2.2	What Data Can Be Stored on Oracle Exadata Extended (XT) Storage Servers?	3-6
3.2.3	Enabling Smart Scan on Exadata Extended (XT) Storage Servers	3-6
3.3	Maintaining the Hard Disks of Oracle Exadata Storage Servers	3-7
3.3.1	Monitoring the Status of Hard Disks	3-8
3.3.2	Monitoring Hard Disk Controller Write-through Caching Mode	3-10
3.3.3	Replacing a Hard Disk Due to Disk Failure	3-12
3.3.4	Replacing a Hard Disk Due to Disk Problems	3-14
3.3.5	Replacing a Hard Disk Due to Bad Performance	3-16
3.3.6	Replacing a Hard Disk Proactively	3-18
3.3.7	Moving All Drives to Another Exadata Storage Server	3-20
3.3.8	Repurposing a Hard Disk	3-22
3.3.9	Removing and Replacing the Same Hard Disk	3-24
3.3.10	Re-Enabling a Hard Disk That Was Rejected	3-24
3.4	Maintaining Flash Disks on Oracle Exadata Storage Servers	3-24
3.4.1	Replacing a Flash Disk Due to Flash Disk Failure	3-25
3.4.2	About Flash Disk Degraded Performance Statuses	3-28
3.4.3	Replacing a Flash Disk Due to Flash Disk Problems	3-31
3.4.4	Performing a Hot Pluggable Replacement of a Flash Disk	3-33
3.4.5	Enabling Write Back Flash Cache	3-34
3.4.5.1	Enable Write Back Flash Cache for 11.2.3.3.1 or Higher	3-35
3.4.5.2	Enabling Write Back Flash Cache on a Rolling Basis for Software Versions Lower Than 11.2.3.3.1	3-36
3.4.5.3	Enabling Write Back Flash Cache in a Non-Rolling Basis for Software Versions Lower Than 11.2.3.3.1	3-37
3.4.6	Disabling Write Back Flash Cache	3-38
3.4.6.1	Disable Write Back Flash Cache for Exadata X8M or Later Servers	3-39
3.4.6.2	Disable Write Back Flash Cache for 11.2.3.3.1 or Higher	3-41

3.4.6.3	Disabling Write Back Flash Cache on a Rolling Basis for Software Versions Lower Than 11.2.3.3.1	3-42
3.4.6.4	Disabling Write Back Flash Cache on a Non-Rolling Basis for Software Versions Lower Than 11.2.3.3.1	3-44
3.4.7	Enabling Flash Cache Compression	3-45
3.4.8	Monitoring Exadata Smart Flash Cache Usage Statistics	3-47
3.4.9	Disabling Flash Cache Compression	3-48
3.5	Maintaining PMEM Devices on Oracle Exadata Storage Servers	3-50
3.5.1	Replacing a PMEM Device Due to Device Failure	3-51
3.5.2	Replacing a PMEM Device Due to Degraded Performance	3-52
3.5.3	Enabling and Disabling Write-Back PMEM Cache	3-54
3.5.3.1	Enable Write Back PMEM Cache	3-54
3.5.3.2	Disable Write-Back PMEM Cache	3-57
3.6	Maintaining the M.2 Disks of Oracle Exadata Storage Server	3-58
3.6.1	Monitoring the Status of M.2 Disks	3-58
3.6.2	Replacing a M.2 Disk Due to Failure or Other Problems	3-59
3.7	Managing the RAM Cache on the Storage Servers	3-61
3.7.1	About the Cell RAM Cache	3-62
3.7.2	Sizing Recommendations for the Cell RAM Cache	3-62
3.7.3	Enabling the Cell RAM Cache	3-64
3.7.4	Viewing the Cell RAM Cache Size	3-64
3.7.5	Changing the Size of the Cell RAM Cache	3-65
3.7.6	Monitoring Cell RAM Cache Usage Statistics	3-65
3.7.7	Disabling the Cell RAM Cache	3-66
3.8	Resizing Grid Disks	3-66
3.8.1	Determine the Amount of Available Space	3-67
3.8.2	Shrink the Oracle ASM Disks in the Donor Disk Group	3-72
3.8.3	Shrink the Grid Disks in the Donor Disk Group	3-74
3.8.4	Increase the Size of the Grid Disks Using Available Space	3-75
3.8.5	Increase the Size of the Oracle ASM Disks	3-77
3.9	Using the Oracle Exadata System Software Rescue Procedure	3-79
3.9.1	About the Oracle Exadata System Software Rescue Procedure	3-79
3.9.2	Performing the Rescue Procedure	3-81
3.9.3	Configuring Exadata Database Machine Storage Servers After Rescue	3-83
3.9.4	Configuring Oracle Exadata Database Machine Eighth Rack Storage Servers After Rescue	3-85
3.9.5	Re-creating a Damaged CELLBOOT USB Flash Drive	3-86
3.10	Changing Existing Elastic Configurations for Storage Cells	3-87
3.10.1	Adding a Cell Node	3-87
3.10.2	Adding a New Storage Server to an Eighth Rack Cluster	3-92
3.10.3	Adding Storage Cells using OEDACLI	3-96
3.10.4	Expanding an Existing Exadata Storage Grid	3-98

3.10.5	Dropping a Storage Server from an Existing Disk Group or Storage Grid	3-98
3.10.6	Dropping Storage Servers using OEDACLI	3-101
3.11	Managing Disk Controller Batteries	3-102
3.11.1	About Disk Controller Batteries	3-102
3.11.2	Monitoring Batteries in the Database Servers	3-103
3.11.3	Replacing Batteries in Disk Controllers	3-104
3.12	Managing F20 PCIe Energy Storage Modules	3-104
3.12.1	About F20 PCIe Energy Storage Modules	3-104
3.12.2	Replacing Flash ESM	3-105
3.13	LED Status Descriptions	3-105
3.13.1	Exadata Storage Server X7-2 and Later LEDs	3-105
3.13.2	Exadata Storage Server X5-2 Servers	3-107
3.13.3	Exadata Storage Server X4-2L Servers	3-108
3.13.4	Exadata Storage Server X3-2 Servers	3-108
3.13.5	Exadata Storage Server with Sun Fire X4270 M2 Servers	3-109
3.13.6	Exadata Storage Server with Sun Fire X4275 Servers LEDs	3-110
3.13.7	Sun Flash Accelerator F20 PCIe Card LEDs	3-111
3.14	Exadata Storage Server Images	3-112
3.14.1	Oracle Exadata Storage Server X9M-2 Extreme Flash Server Images	3-112
3.14.2	Oracle Exadata Storage Server X9M-2 High Capacity Server Images	3-113
3.14.3	Oracle Exadata Storage Server X9M-2 Extended Server Images	3-114
3.14.4	Oracle Exadata Storage Server X8M-2 and X8-2 High Capacity and Extended (XT) Server Images	3-114
3.14.5	Oracle Exadata Storage Server X8M-2 and X8-2 Extreme Flash Server Images	3-115
3.14.6	Oracle Exadata Storage Server X7-2 High Capacity Server Images	3-116
3.14.7	Oracle Exadata Storage Server X7-2 Extreme Flash Server Images	3-116
3.14.8	High Capacity Exadata Storage Server X6-2 Images	3-117
3.14.9	Extreme Flash Exadata Storage Server X6-2 Images	3-117
3.14.10	High Capacity Exadata Storage Server X5-2 Images	3-118
3.14.11	Extreme Flash Exadata Storage Server X5-2 Images	3-118
3.14.12	Exadata Storage Server X4-2L Images	3-119
3.14.13	Exadata Storage Server X3-2L Images	3-119
3.14.14	Exadata Storage Server with Sun Fire X4270 M2 Images	3-120
3.14.15	Exadata Storage Server with Sun Fire X4275 Images	3-121

4 Maintaining Other Exadata Database Machine Components

4.1	Replacing a Power Distribution Unit	4-2
4.1.1	Reviewing the PDU Replacement Guidelines	4-2
4.1.2	Replacing a PDU	4-3
4.2	Resetting a Non-Responsive ILOM	4-4

4.2.1	Resetting the ILOM Using SSH	4-5
4.2.2	Resetting the ILOM Using the ILOM Remote Console	4-5
4.2.3	Resetting the ILOM Using IPMItool	4-6
4.2.4	Resetting the ILOM Using the SP Reset Pin on Oracle Exadata Database Machine X2-2 Servers and Exadata Storage Servers	4-6
4.2.5	Removing the SP from Sun Fire X4800 Oracle Database Servers and Sun Server X2-8 Oracle Database Servers	4-6
4.2.6	Unplugging the ILOM Power Supply	4-7
4.3	Configuring Service Processor and ILOM Network Settings	4-7
4.4	Verifying and Modifying the Link Speed on the Client Network Ports for X7 and Later Systems	4-9
4.5	Verify the Link Speed on Network Ports	4-10
4.6	Changing from 1 GbE Connections to 10 GbE Connections	4-11
4.6.1	Task 1: Verify ping Functionality	4-12
4.6.2	Task 2: Back up the Current Interface Files	4-12
4.6.3	Task 3: Edit the 10 GbE Interface Settings	4-13
4.6.4	Task 4: Verify the 10 GbE Interfaces	4-15
4.7	Maintaining the RoCE Network Fabric	4-17
4.7.1	Backing Up Settings on the Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch	4-17
4.7.2	Applying Golden Configuration Settings on Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches	4-18
4.7.3	Restoring Settings on a Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch	4-21
4.7.4	Using Access VLANs with Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches	4-23
4.7.5	Replacing a Failed RoCE Network Fabric Switch	4-25
4.7.6	Verifying the RoCE Network Fabric Configuration	4-26
4.7.7	Verifying RoCE Network Fabric Operation	4-26
4.7.8	Upgrading the Switch Firmware for RoCE Network Fabric	4-28
4.7.9	Downgrading the RoCE Network Fabric Switch Software	4-29
4.8	Maintaining the InfiniBand Network Fabric Network	4-29
4.8.1	Backing Up and Restoring InfiniBand Switch Settings	4-29
4.8.1.1	Backing Up Settings on a Switch with 2.1.3-4 Firmware	4-30
4.8.1.2	Backing Up Settings on a Switch with 1.1.3-2 Firmware	4-30
4.8.1.3	Backing Up Settings on a Switch with 1.0.1 Firmware	4-31
4.8.1.4	Restoring Settings on a Switch with 2.1.3-4 Firmware	4-32
4.8.1.5	Restoring Settings on a Switch with 1.1.3-2 Firmware	4-32
4.8.1.6	Restoring Settings on a Switch with 1.0.1 Firmware	4-33
4.8.2	Verifying the InfiniBand Network Fabric Configuration	4-34
4.8.3	Using the verify-topology Utility	4-34
4.8.4	Verifying InfiniBand Network Fabric Operation	4-36
4.8.5	Understanding the Network Subnet Manager Master	4-38
4.8.6	Upgrading the Switch Firmware for InfiniBand Network Fabric	4-39

4.8.7	Downgrading the InfiniBand Network Fabric Switch Software	4-40
4.9	Modifying the InfiniBand Network Fabric Configuration	4-40
4.9.1	Configuring InfiniBand Partitioning	4-40
4.9.2	Changing InfiniBand IP Addresses and Host Names	4-40
4.9.3	Changing InfiniBand Network Information	4-40
4.10	Configuring Network Routing on Database Servers	4-52
4.10.1	About Network Routing on Database Servers	4-53
4.10.2	Task 1: Configure for Boot-Time Routing	4-54
4.10.3	Task 2: Configure for Real-Time Routing	4-54
4.10.4	Task 3: Verify Network Routing Rules and Routes	4-54
4.10.5	Removing Network Routing Configuration for Troubleshooting	4-55
4.10.6	Returning to Default Routing	4-55
4.11	Changing the DNS Servers	4-56
4.11.1	Change the DNS Server Address on the Database Server	4-56
4.11.2	Change the DNS Server on Oracle Exadata Storage Server	4-58
4.11.3	Change the DNS Server Address on the Cisco RoCE Network Fabric Switches	4-61
4.11.4	Change the DNS Server Address on the Cisco 9300 Series Management Network Switch	4-62
4.11.5	Change the DNS Server Address on the Cisco 4948 Ethernet Switch	4-64
4.11.6	Change the DNS Server Address on the InfiniBand Network Fabric Switch	4-66
4.11.7	Change the DNS Server on the KVM Switch	4-66
4.12	Changing the NTP Servers	4-67
4.12.1	Set the NTP Server Address on the Database Servers	4-68
4.12.2	Set the NTP Server on Oracle Exadata Storage Servers	4-69
4.12.3	Set the NTP Server Address on the Cisco RoCE Network Fabric Switches	4-73
4.12.4	Set the NTP Server Address on the Cisco 9300 Series Management Network Switch	4-74
4.12.5	Set the NTP Server Address on the Cisco 4948 Ethernet Switch	4-76
4.12.6	Set the NTP Server Address on the InfiniBand Network Fabric Switch	4-78
4.13	Changing the Time Zone Settings	4-78
4.13.1	Change Time Zone Settings on Storage Servers	4-79
4.13.2	Change Time Zone Settings on the Database Servers	4-81
4.13.3	Change Time Zone Settings on the InfiniBand Network Fabric Switches	4-82
4.13.4	Change Time Zone Settings on the Cisco RoCE Network Fabric Switches	4-84
4.13.5	Change Time Zone Settings on the Cisco Management Network Switch	4-85
4.14	Managing the KVM Switch	4-86
4.14.1	Configuring the KVM Switch	4-87
4.14.2	Configuring the KVM Switch to Access a Server	4-88
4.14.3	Accessing a Server Using the KVM Switch	4-89
4.15	LED Status Descriptions	4-89
4.15.1	Sun Datacenter InfiniBand Switch 36 Switch LEDs	4-90

5 Managing Oracle VM User Domains

5.1	Oracle VM and Exadata Database Machine	5-2
5.1.1	About Oracle VM	5-3
5.1.2	Oracle VM Deployment Specifications and Limits	5-4
5.1.3	Supported Operations in the Management Domain (dom0)	5-5
5.1.4	Oracle VM Resources	5-6
5.1.4.1	Storage Configuration for Management Domain	5-7
5.1.4.2	Storage Configuration for User Domain	5-7
5.2	Migrating a Bare Metal Oracle RAC Cluster to an Oracle RAC Cluster in Oracle VM	5-9
5.3	Showing Running Domains	5-9
5.4	Monitoring a User Domain Console	5-10
5.5	Monitoring Oracle VMs with Oracle Enterprise Manager	5-10
5.6	Starting a User Domain	5-10
5.7	Managing Automatic Startup of Oracle VM User Domains	5-11
5.7.1	Enabling User Domain Automatic Start	5-11
5.7.2	Disabling User Domain Automatic Start	5-11
5.8	Shutting Down a User Domain From Within the User Domain	5-12
5.9	Shutting Down a User Domain From Within the Management Domain	5-12
5.10	Backing Up and Restoring Oracle Databases on Oracle VM User Domains	5-13
5.11	Modifying the Memory Allocated to a User Domain	5-13
5.12	Modifying the Number of Virtual CPUs Allocated to a User Domain	5-15
5.13	Increasing the Disk Space in a User Domain	5-17
5.13.1	Adding a New LVM Disk to a User Domain	5-17
5.13.2	Increasing the Size of the root File System	5-20
5.13.3	Increasing the Size of the /u01 File System	5-22
5.13.4	Increasing the Size of the Grid Infrastructure Home or Database Home File System	5-23
5.13.5	Increasing the Size of the Swap Area	5-25
5.14	Expanding /EXAVMIMAGES After Adding the Database Server Disk Expansion Kit	5-26
5.14.1	Expanding /EXAVMIMAGES on Management Domain on Release 18.1.x or Later	5-27
5.14.2	Expanding /EXAVMIMAGES on Management Domain on Release 12.2.x	5-31
5.14.3	Expanding /EXAVMIMAGES on Management Domain on Releases Earlier than 12.2.x	5-35
5.15	Adding an Oracle VM Cluster	5-39
5.16	Expanding an Oracle RAC Cluster on Oracle VM Using OEDACLI	5-40
5.17	Creating a User Domain Without Oracle Grid Infrastructure and Oracle Database	5-44
5.18	Moving a User Domain to a Different Database Server	5-46

5.19	Backing up the Management Domain and User Domains in an Oracle VM Deployment	5-47
5.19.1	Backing up the Management Domain dom0 Using Snapshot-Based Backup	5-47
5.19.2	Backing up the User Domains	5-49
5.19.2.1	Method 1: Back up All the User Domains	5-50
5.19.2.2	Method 2: Back up Individual User Domains	5-51
5.19.2.3	Method 3: Back up a User Domain from Inside the User Domain	5-53
5.20	Recovering an Oracle VM Deployment	5-55
5.20.1	Overview of Snapshot-Based Recovery of Database Servers	5-55
5.20.2	Scenario 1: Recovering a Management Domain and Its User Domains from Backup	5-56
5.20.2.1	Recovering a Management Domain and Its User Domains (Releases Prior to 12.2.1.1.0)	5-57
5.20.2.2	Recovering a Management Domain and Its User Domains (Releases 12.2.1.1.0 and Later)	5-61
5.20.2.3	Recovering a Management Domain and Its User Domains (Release 18.1 and X7 and Later)	5-66
5.20.3	Scenario 2: Re-imaging the Management Domain and Restoring User Domains from Backups	5-73
5.20.4	Scenario 3: Restoring and Recovering User Domains from Snapshot Backups	5-77
5.21	Removing an Oracle RAC Cluster Running in Oracle VM	5-77
5.22	Removing a User Domain	5-81
5.22.1	Removing a User Domain from an Oracle VM Cluster Using OEDACLI	5-81
5.22.2	Removing a User Domain from an Oracle VM Cluster Using domu_maker	5-82
5.23	Implementing Tagged VLAN Interfaces	5-83
5.24	Implementing InfiniBand Partitioning across Oracle VM Oracle RAC Clusters on Exadata Database Machine	5-85
5.24.1	About InfiniBand Partitioning Across Oracle RAC Clusters Running in Oracle VM	5-85
5.24.2	Requirements for Implementing InfiniBand Partitioning across OVM RAC Clusters	5-87
5.24.3	About InfiniBand Partitioning Network Configuration	5-88
5.24.4	Configuring InfiniBand Partitioning across Oracle VM RAC Clusters	5-88
5.24.5	Implementing InfiniBand Partitioning across OVM RAC Clusters: Setting up Limited Membership	5-97
5.25	Running Oracle EXAchk in Oracle VM Environments	5-101

6 Managing Oracle Linux KVM Guests

6.1	Oracle Linux KVM and Exadata Database Machine	6-2
6.1.1	About Oracle Linux KVM	6-3
6.1.2	Oracle Linux KVM Deployment Specifications and Limits	6-4
6.1.3	Supported Operations in the KVMHost	6-5
6.1.4	Oracle Linux KVM Resources	6-6

6.2	Migrating a Bare Metal Oracle RAC Cluster to an Oracle RAC Cluster in Oracle Linux KVM	6-6
6.3	Showing Running Domains	6-7
6.4	Starting a Guest	6-7
6.5	Starting a Guest using the Diagnostic ISO File	6-8
6.6	Monitoring a Guest Console During Startup	6-10
6.7	Managing Automatic Startup of Oracle Linux KVM Guests	6-10
6.7.1	Enabling Guest Automatic Start	6-11
6.7.2	Disabling Guest Automatic Start	6-11
6.8	Shutting Down a User Domain From Within the User Domain	6-11
6.9	Shutting Down a Guest From Within the KVM host	6-11
6.10	Backing up the KVM host and Guests	6-12
6.10.1	Backing up the KVM host Using Snapshot-Based Backup	6-12
6.10.2	Backing up the Oracle Linux KVM Guests	6-15
6.10.2.1	Method 1: Back Up All of the KVM Guests	6-16
6.10.2.2	Method 2: Back Up an Individual Guest	6-19
6.10.2.3	Method 3: Back Up a Guest Internally	6-21
6.11	Backing Up and Restoring Oracle Databases on KVM Guests	6-25
6.12	Modifying the Memory Allocated to a Guest	6-26
6.13	Modifying the Number of Virtual CPUs Allocated to a Guest	6-27
6.14	Increasing the Disk Space in a Guest	6-27
6.14.1	Adding a New LVM Disk to a Guest	6-28
6.14.2	Increasing the Size of the root File System	6-32
6.14.3	Increasing the Size of the /u01 File System	6-33
6.14.4	Increasing the Size of the Grid Infrastructure Home or Database Home File System	6-35
6.14.5	Increasing the Size of the Swap Area	6-39
6.15	Expanding /EXAVMIMAGES on the KVM host	6-42
6.16	Adding an Oracle Linux KVM Cluster	6-43
6.17	Expanding an Oracle RAC Cluster in Oracle Linux KVM Using OEDACLI	6-44
6.18	Moving a Guest to a Different Database Server	6-48
6.19	Recovering a KVM Deployment	6-49
6.19.1	Overview of Snapshot-Based Recovery of KVM Hosts	6-49
6.19.2	KVM System Recovery Scenarios	6-50
6.19.2.1	Recovering a KVM Host and the Guests from Backup	6-50
6.19.2.2	Re-imaging a KVM Host and Restoring the Guests from Backup	6-67
6.19.2.3	Recover and Restart a KVM Guest	6-69
6.19.2.4	Restoring and Recovering Guests from Snapshot Backups	6-70
6.20	Removing a Guest	6-70
6.20.1	Removing a Guest from an Oracle RAC Cluster Using OEDACLI	6-70
6.20.2	Removing a Guest Using vm_maker	6-72
6.21	Using Client Network VLAN Tagging with Oracle Linux KVM	6-72

6.22	Using Exadata Secure RDMA Fabric Isolation with Oracle Linux KVM	6-75
6.23	Adding a Bonded Network Interface to an Oracle Linux KVM Guest	6-76
6.24	Using Oracle EXAchk in Oracle Linux KVM Environments	6-78
6.24.1	Running Oracle EXAchk in Oracle Linux KVM Environments	6-78
6.24.2	Audit Checks Performed by Oracle EXAchk	6-78
6.24.3	Oracle EXAchk Command Line Options for Exadata Database Machine	6-79

7 VM_MAKER Command Reference

7.1	About the vm_maker Utility	7-2
7.2	attach Command	7-3
7.3	add-bonded-bridge Command	7-4
7.4	add-single-bridge Command	7-4
7.5	allocate-bridge Command	7-5
7.6	autostart Command	7-6
7.7	boot-from-hd Command	7-6
7.8	boot-from-iso Command	7-6
7.9	check Command	7-7
7.10	cleanup-ip Command	7-8
7.11	configure-system Command	7-9
7.12	create Command	7-9
7.13	detach Command	7-10
7.14	dumpxml Command	7-10
7.15	list Command	7-11
7.16	list-domains Command	7-12
7.17	network-discovery Command	7-12
7.18	reboot Command	7-12
7.19	remove-bridge Command	7-12
7.20	remove-domain Command	7-13
7.21	set Command	7-13
7.22	setup-ip Command	7-14
7.23	start-domain Command	7-16
7.24	stop-domain Command	7-16
7.25	validate-ip Command	7-17
7.26	vhostmd Command	7-18

8 Updating Exadata Software

8.1	About Updating Exadata Software	8-2
8.2	Planning for Software Maintenance	8-3
8.2.1	Understanding Exadata Database Machine Software and Updates	8-3

8.2.1.1	What Software Do You Update on Exadata Database Machine?	8-3
8.2.1.2	Software Release Types	8-4
8.2.1.3	Software Release Availability	8-5
8.2.1.4	Software Update Frequency	8-6
8.2.1.5	Software Update Utilities	8-8
8.2.2	Configuration and Operational Best Practices for Software Maintenance	8-12
8.2.2.1	Understanding Rolling and Non-Rolling Updates	8-12
8.2.2.2	Online Updates for Oracle Linux Kernel and Oracle Database Interim Fixes	8-14
8.2.2.3	Configuration Practices for Optimal Software Maintenance	8-15
8.2.2.4	Operational Practices for Optimal Software Maintenance	8-17
8.2.2.5	Version Compatibility and Mixed Version Support	8-18
8.2.2.6	Ordering of Updates	8-20
8.2.3	Understanding the Exadata Software Image Version	8-20
8.2.4	Rules for Updating to Newer Exadata Version	8-21
8.3	About Upgrading to Oracle Linux 7 on Exadata Servers	8-21
8.4	Overview of Performing Exadata Software Updates	8-22
8.4.1	Actions to Perform Before Any Software Maintenance	8-22
8.4.2	Overview of Performing Exadata Storage Server Updates	8-23
8.4.3	Overview of Performing Exadata Database Servers Updates	8-24
8.4.3.1	Steps for Updating Non-Virtualized Configurations	8-24
8.4.3.2	Steps for Updating Virtualized Configurations	8-25
8.4.4	Overview of Performing RoCE Network Fabric Switch Updates	8-26
8.4.5	Overview of Performing InfiniBand Network Fabric Switch Updates	8-26
8.4.6	Overview of Performing Oracle Grid Infrastructure and Oracle Database Updates	8-27
8.4.7	Using sudo When Performing Software Updates	8-29
8.4.7.1	Running patchmgr (and dbnodeupdate Orchestration) Using sudo	8-29
8.4.7.2	Running dbnodeupdate.sh Using sudo	8-30
8.5	Exadata Patchmgr Update Utility	8-31
8.5.1	About the Exadata Patchmgr Update Utility	8-31
8.5.2	Obtaining Patchmgr	8-32
8.5.3	Patchmgr Syntax	8-33
8.5.3.1	Patchmgr Syntax for Storage Servers	8-34
8.5.3.2	Patchmgr Syntax for Database Servers	8-37
8.5.3.3	Patchmgr Syntax for RoCE Network Fabric Switches	8-41
8.5.3.4	Patchmgr Syntax for InfiniBand Network Fabric Switches	8-46
8.5.3.5	Patchmgr Syntax for the Management Network Switch	8-48
8.6	Updating Exadata Database Machine Database Servers	8-49
8.6.1	Overview of Exadata Database Machine Database Server Updates	8-50
8.6.2	Update Paths for Exadata Database Servers	8-51
8.6.3	Installing, Updating, and Managing Non-Oracle Software	8-52

8.6.4	Customization Levels and Impact	8-53
8.6.5	Update Utility for Exadata Database Servers	8-54
8.6.6	Update Tool Execution Host	8-55
8.6.7	Running the Update Utility as a Non-root User and Running Multiple Invocations Concurrently	8-55
8.6.8	Recommended Timeline for Updating Exadata Database Servers	8-56
8.6.9	Preparing and Populating the YUM Repository with the Oracle Exadata Channel Content	8-57
8.6.9.1	Downloading the ISO Image as a YUM Repository from My Oracle Support and Passing the Location of the Compressed Image to the Update Utility	8-58
8.6.9.2	Setting Up a Local Mirror as a YUM HTTP Repository from the ULN Oracle Exadata Channel on a Separate Server	8-58
8.6.9.3	Downloading the ISO Image as a YUM Repository from My Oracle Support and Making It Available as a YUM HTTP Repository on a Web Server	8-60
8.6.10	Managing Exadata Obsolete Packages	8-61
8.6.11	Updating Individual Packages	8-62
8.6.12	Running Prerequisite Checks	8-62
8.6.13	Backing up Exadata Database Servers Before Planned Maintenance	8-64
8.6.14	Running the Update	8-66
8.6.15	Rolling Back Updates	8-68
8.7	Updating Database Servers Running Oracle Exadata System Software Release 11.2.2.4.2	8-69
8.7.1	Preparing to Use the dbnodeupdate.sh Utility on Database Servers with Release 11.2.2.4.2	8-69
8.7.2	Running the dbnodeupdate.sh Utility	8-70
8.8	Updating Software on Oracle Exadata Storage Servers	8-71
8.8.1	Overview of Exadata Database Machine Storage Server Updates	8-71
8.8.2	Scheduling Automated Updates of Storage Servers	8-72
8.8.3	Update Utility for Oracle Exadata Storage Server	8-73
8.8.4	Recommended Timeline for Updating Oracle Exadata Storage Server	8-74
8.8.5	Preparing Exadata Storage Servers for Update	8-75
8.8.6	Running the Update for Exadata Storage Servers	8-79
8.8.7	Rolling Back Updates for Exadata Storage Servers	8-80
8.9	Upgrading and Downgrading RoCE Network Fabric Switch Firmware	8-81
8.9.1	Preparing for RoCE Network Fabric Switch Firmware Upgrades or Downgrades	8-82
8.9.2	Upgrading RoCE Network Fabric Switch Firmware Software	8-85
8.9.3	Downgrading RoCE Network Fabric Switch Firmware	8-86
8.10	Updating InfiniBand Network Fabric Switch Firmware	8-87
8.10.1	Preparing for InfiniBand Network Fabric Switch Firmware Updates	8-88
8.10.2	Updating InfiniBand Network Fabric Switch Firmware Software	8-90
8.10.3	Downgrading InfiniBand Network Fabric Switch Firmware	8-91

8.11	Upgrading Oracle Java SE on Oracle Linux	8-92
8.11.1	Stop the MS Process	8-93
8.11.2	Download and Update the Java JDK Package	8-93
8.11.2.1	Using YUM and ULN to Update the JDK Package on Database Servers	8-94
8.11.2.2	Manually Update the JDK Package on Oracle Exadata System Software versions 12.1.2.1.0 to 12.1.2.2.0	8-99
8.11.2.3	Manually Update the JDK Package on Oracle Exadata System Software Release 12.1.2.2.1 and Later	8-100
8.11.3	Reconfigure Management Server (MS)	8-101
8.12	Setting up SSH Equivalence	8-102
8.13	Troubleshooting Software Updates on Exadata Database Machine	8-103
8.13.1	Troubleshooting Exadata Database Server Updates	8-103
8.13.2	Monitoring, Validating, and Troubleshooting Exadata Storage Server Updates	8-105
8.13.3	Troubleshooting a Failing Prerequisite Check Due to Dependency Problems on Exadata Database Server	8-106
8.13.4	Troubleshooting a Multilib Problem on Exadata Database Server	8-108

9 Using the DBMCLI Utility

9.1	Overview of the DBMCLI Utility	9-1
9.1.1	Starting DBMCLI	9-2
9.1.2	Understanding Command Syntax and Options for DBMCLI	9-2
9.1.3	Reserved Words	9-3
9.1.4	DBMCLI Command-Line Editing	9-4
9.1.5	DBMCLI Input and Output Options	9-4
9.1.6	Comments in DBMCLI Scripts	9-4
9.1.7	Line Continuation in DBMCLI Commands	9-4
9.2	About DBMCLI Administration Commands	9-4
9.3	About DBMCLI Object Commands	9-5
9.4	About DBMCLI Object Types	9-5
9.5	DBMCLI Object Attributes	9-6
9.5.1	Restrictions on Values of Common Attributes	9-7
9.5.2	Attribute Lists in LIST Command	9-7
9.5.3	Attribute Filters in LIST and ALTER Commands	9-7
9.6	Monitoring Database Server with Metrics	9-8
9.6.1	About Metrics	9-9
9.6.2	Displaying Metrics	9-9
9.6.2.1	Displaying Specific Attributes for Metrics	9-10
9.6.2.2	Displaying Metric Definitions	9-10
9.6.2.3	Displaying Current Metrics	9-11
9.6.2.4	Displaying Metric History	9-11
9.6.3	Database Server Metrics	9-12

9.6.4	Real-Time Insight	9-13
9.6.5	Using Real-Time Insight	9-13
9.6.5.1	Using Fine-Grained Metrics	9-14
9.6.5.2	Uploading Metric Observations to a Collection Platform	9-15
9.6.5.3	Downloading Metric Observations from an Exadata Server	9-17
9.6.5.4	Tagging Metric Observations	9-17
9.6.5.5	Understanding the Metric Stream Format	9-18
9.7	DBMCLI Command Reference	9-20
9.7.1	ALTER	9-20
9.7.1.1	ALTER ALERTHISTORY	9-21
9.7.1.2	ALTER DBSERVER	9-22
9.7.1.3	ALTER IBPORT	9-37
9.7.1.4	ALTER METRICDEFINITION	9-37
9.7.1.5	ALTER SOFTWAREUPDATE	9-39
9.7.1.6	ALTER THRESHOLD	9-42
9.7.1.7	ALTER USER	9-43
9.7.2	CREATE	9-43
9.7.2.1	CREATE DBSERVER	9-44
9.7.2.2	CREATE DIAGPACK	9-46
9.7.2.3	CREATE ROLE	9-50
9.7.2.4	CREATE THRESHOLD	9-50
9.7.2.5	CREATE USER	9-51
9.7.3	DESCRIBE	9-52
9.7.3.1	DESCRIBE ALERTDEFINITION	9-53
9.7.3.2	DESCRIBE ALERTHISTORY	9-53
9.7.3.3	DESCRIBE DBSERVER	9-55
9.7.3.4	DESCRIBE IBPORT	9-59
9.7.3.5	DESCRIBE LUN	9-61
9.7.3.6	DESCRIBE METRICCURRENT	9-62
9.7.3.7	DESCRIBE METRICDEFINITION	9-63
9.7.3.8	DESCRIBE METRICHISTORY	9-64
9.7.3.9	DESCRIBE PHYSICALDISK	9-66
9.7.3.10	DESCRIBE ROLE	9-68
9.7.3.11	DESCRIBE SOFTWAREHISTORY	9-68
9.7.3.12	DESCRIBE SOFTWAREUPDATE	9-69
9.7.3.13	DESCRIBE THRESHOLD	9-69
9.7.3.14	DESCRIBE USER	9-70
9.7.4	DROP	9-71
9.7.4.1	DROP ALERTHISTORY	9-72
9.7.4.2	DROP DBSERVER	9-72
9.7.4.3	DROP ROLE	9-72

9.7.4.4	DROP SOFTWAREHISTORY	9-73
9.7.4.5	DROP THRESHOLD	9-73
9.7.4.6	DROP USER	9-73
9.7.5	GRANT	9-74
9.7.5.1	GRANT PRIVILEGE	9-75
9.7.5.2	GRANT ROLE	9-76
9.7.6	HELP	9-77
9.7.7	LIST	9-77
9.7.7.1	LIST ALERTDEFINITION	9-79
9.7.7.2	LIST ALERTHISTORY	9-79
9.7.7.3	LIST DBSERVER	9-80
9.7.7.4	LIST DIAGPACK	9-82
9.7.7.5	LIST IBPORT	9-83
9.7.7.6	LIST LUN	9-84
9.7.7.7	LIST METRICCURRENT	9-85
9.7.7.8	LIST METRICDEFINITION	9-87
9.7.7.9	LIST METRICHISTORY	9-87
9.7.7.10	LIST METRICSTREAM	9-89
9.7.7.11	LIST PHYSICALDISK	9-89
9.7.7.12	LIST ROLE	9-90
9.7.7.13	LIST SOFTWAREHISTORY	9-91
9.7.7.14	LIST SOFTWAREUPDATE	9-91
9.7.7.15	LIST THRESHOLD	9-92
9.7.7.16	LIST USER	9-93
9.7.8	REVOKE	9-93
9.7.8.1	REVOKE PRIVILEGE	9-94
9.7.8.2	REVOKE ROLE	9-95
9.7.9	SET	9-95
9.7.10	SPOOL	9-96
9.7.11	START	9-97

10 Using the ExaCLI Utility

10.1	ExaCLI Command Reference	10-1
10.2	ExaCLI Security	10-6
10.3	Creating Users for Use with ExaCLI	10-6
10.4	Differences in Commands Between ExaCLI and DBMCLI/CellCLI	10-7
10.5	Certificates for ExaCLI	10-8
10.5.1	Using the Default Self-Signed Certificate	10-8
10.5.2	Using a CA-Certified Security Certificate	10-9

11 Using the exadcli Utility

11.1	exadcli Command Reference	11-1
11.2	Using exadcli for the First Time	11-4
11.3	exadcli Examples	11-6

Preface

This guide describes the maintenance procedures for Oracle Exadata Database Machine. It includes information about maintenance and repair, as well as replacement parts information.

- [Audience](#)
- [Documentation Accessibility](#)
- [Diversity and Inclusion](#)
- [Related Documentation](#)
- [Conventions](#)

Audience

This guide is intended for Oracle Exadata Database Machine customers and those responsible for data center site planning, configuration, and maintenance of Oracle Exadata Database Machine.

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc>.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info> or visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs> if you are hearing impaired.

Diversity and Inclusion

Oracle is fully committed to diversity and inclusion. Oracle respects and values having a diverse workforce that increases thought leadership and innovation. As part of our initiative to build a more inclusive culture that positively impacts our employees, customers, and partners, we are working to remove insensitive terms from our products and documentation. We are also mindful of the necessity to maintain compatibility with our customers' existing technologies and the need to ensure continuity of service as Oracle's offerings and industry standards evolve. Because of these technical constraints, our effort to remove insensitive terms is ongoing and will take time and external cooperation.

Related Documentation

In addition to the Exadata Database Machine documentation set, the following guides contain hardware information for Exadata Database Machine:

- *Oracle Exadata Database Machine System Overview*
- *Oracle Exadata Database Machine Security Guide*
- *Oracle Engineered System Safety and Compliance Guide*
- *Oracle Exadata Database Machine Extending and Multi-Rack Cabling Guide*
- *Oracle Exadata Database Machine Maintenance Guide*
- *Oracle Exadata System Software User's Guide*
- *Oracle Exadata X9-2 Database Server Service Manual* at <https://docs.oracle.com/en/servers/x86/x9-2/exa-database/index.html>
- *Oracle Exadata Storage Server X9-2 EF, HC, and XT Service Manual* at <https://docs.oracle.com/en/servers/x86/x9-2/exa-storage/index.html>
- *Oracle Exadata X8-2 Database Server Service Manual* at <https://docs.oracle.com/en/servers/x86/x8-2/exadata-db-service-manual/>
- *Oracle Exadata Storage Server X8-2 EF, HC, and XT Service Manual* at <https://docs.oracle.com/en/servers/x86/x8-2/exadata-storage-service-manual/>
- *Oracle Exadata Database Server X8-8 Service Manual* at <https://docs.oracle.com/en/servers/x86/x8-8/exadata-storage-service-manual/index.html>
- *Oracle Server X7-2 Service Manual* at http://docs.oracle.com/cd/E72435_01/html/E72445/index.html
- *Oracle Server X7-2L Service Manual* at http://docs.oracle.com/cd/E72463_01/html/E72474/index.html
- *Oracle Server X7-8 Service Manual* at http://docs.oracle.com/cd/E71925_01/html/E71936/index.html
- *Oracle Server X6-2 Service Manual* at http://docs.oracle.com/cd/E62159_01/html/E62171/index.html
- *Oracle Server X6-2L Service Manual* at http://docs.oracle.com/cd/E62172_01/html/E62184/index.html
- *Oracle Server X5-2 Service Manual* at http://docs.oracle.com/cd/E41059_01/html/E48312/napsm.html
- *Oracle Server X5-2L Service Manual* at http://docs.oracle.com/cd/E41033_01/html/E48325/cnpsm.html#scrolltoc
- *Sun Server X4-8 Service Manual* at http://docs.oracle.com/cd/E40591_01/html/E40317/index.html
- *Sun Server X4-2 Service Manual* at http://docs.oracle.com/cd/E36975_01/html/E38045/gentextid-14757.html#scrolltoc
- *Sun Server X4-2L Service Manual* at http://docs.oracle.com/cd/E36974_01/html/E38145/gentextid-14728.html#scrolltoc

- *Sun Server X3-2 (formerly Sun Fire X4170 M3) Service Manual* at http://docs.oracle.com/cd/E22368_01/html/E27242/gentextid-14840.html#scrolltoc
- *Sun Server X3-2L (formerly Sun Fire X4270 M3) Service Manual* at http://docs.oracle.com/cd/E23393_01/html/E27229/gentextid-14804.html#scrolltoc
- *Sun Server X2-8 (formerly Sun Fire X4800 M2) Service Manual* at http://docs.oracle.com/cd/E20815_01/html/E20819/index.html
- *Sun Fire X4800 Server Service Manual* at <http://docs.oracle.com/cd/E19140-01/html/821-0282/index.html>
- *Sun Fire X4270 M2 Server Service Manual* at <http://docs.oracle.com/cd/E19245-01/E21671/index.html>
- *Sun Fire X4170 M2 Server Service Manual* at <http://docs.oracle.com/cd/E19762-01/E22369-02/index.html>
- *Sun Fire X4170, X4270, and X4275 Servers Service Manual* at <http://docs.oracle.com/cd/E19477-01/820-5830-13/index.html>
- *Sun Datacenter InfiniBand Switch 36 Firmware Version 2.1 Documentation* at http://docs.oracle.com/cd/E36265_01/index.html
- *Sun Datacenter InfiniBand Switch 36 Firmware Version 2.2 Documentation* at http://docs.oracle.com/cd/E76424_01/index.html
- *Sun Flash Accelerator F20 PCIe Card User's Guide* at <http://docs.oracle.com/cd/E19682-01/E21358/index.html>
- *Sun Flash Accelerator F40 PCIe Card User's Guide* at http://docs.oracle.com/cd/E29748_01/html/E29741/index.html
- *Sun Flash Accelerator F80 PCIe Card User's Guide* at http://docs.oracle.com/cd/E41278_01/html/E41251/index.html
- *Oracle Flash Accelerator F160 PCIe Card User Guide* at http://docs.oracle.com/cd/E54943_01/html/E54947/index.html
- *Oracle Flash Accelerator F320 PCIe Card User Guide* at http://docs.oracle.com/cd/E65386_01/html/E65387/index.html
- *Oracle Flash Accelerator F640 PCIe Card User Guide* at https://docs.oracle.com/cd/E87231_01/html/E87233/index.html
- *Sun Storage 6 Gb SAS PCIe RAID HBA Documentation* at <http://docs.oracle.com/cd/E19221-01/>
- *Oracle Storage 12 Gb/s SAS PCIe RAID HBA, Internal Documentation Library* at http://docs.oracle.com/cd/E52363_01/index.html
- *Oracle Integrated Lights Out Manager (ILOM) Documentation* at <http://www.oracle.com/goto/ilom/docs>
- "Cisco Catalyst 4948E and 4948E-F Ethernet Switches Data Sheet" at https://www.cisco.com/c/en/us/products/collateral/switches/catalyst-4948e-ethernet-switch/data_sheet_c78-598933.html
- "Cisco Nexus 9300-EX and 9300-FX Platform Switches Data Sheet" at <https://www.cisco.com/c/en/us/products/collateral/switches/nexus-9000-series-switches/datasheet-c78-736651.html>

Conventions

The following text conventions are used in this document:

Convention	Meaning
boldface	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
<i>italic</i>	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
monospace	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.
\$ prompt	The dollar sign (\$) prompt indicates a command run as the <code>oracle</code> user.
# prompt	The pound (#) prompt indicates a command that is run as the <code>root</code> user.

1

General Maintenance Information

This chapter contains topics about general maintenance of Oracle Exadata Database Machine.

Note:

- All procedures in this chapter are applicable to Exadata Database Machine and Oracle Exadata Storage Expansion Rack.
- For ease of reading, the name "Oracle Exadata Rack" is used when information refers to both Exadata Database Machine and Oracle Exadata Storage Expansion Rack.

- [Overview of Roles and Responsibilities](#)
You should determine which individuals or groups are responsible for resolving any issue that arises.
- [Powering On and Off Oracle Exadata Rack](#)
This section contains the procedures for powering on and off the components of an Oracle Exadata Rack.
- [Using Auto Service Request to Manage Hardware Faults](#)
Auto Service Request (ASR) is designed to automatically open service requests when specific Oracle Exadata Rack hardware faults occur.
- [Monitoring the System Using Oracle Enterprise Manager Cloud Control](#)
- [Monitoring the System Using Oracle Configuration Manager](#)
Oracle Configuration Manager collects configuration information and uploads it to the Oracle repository.
- [Changing Component Passwords](#)
The passwords for the components can be changed after initial configuration.
- [Determining the Server Model](#)
Use the `exadata.img.hw` command to determine the model of the cell or database server.
- [Monitoring Ambient Temperature of Servers](#)
- [Replacing a Disk Controller Battery Backup Unit](#)
- [Overview of the dbmsrv Service](#)
- [Configuring Password Expiration for Users Accessing the Server Remotely](#)
You can configure `DBSERVER` attributes to expire user passwords.
- [State of Storage Server and Database Servers During Configuration Changes](#)
Before a change to a configuration, determine if the database and storage servers need to be offline or online.
- [Rescue Plan](#)

- [Using ExaWatcher Charts to Monitor Exadata Database Machine Performance](#)
ExaWatcher is a utility that collects performance data on the storage servers and database servers of an Exadata system. The data collected includes operating system statistics, such as iostat, cell statistics (cellsvrstat), and network statistics.

1.1 Overview of Roles and Responsibilities

You should determine which individuals or groups are responsible for resolving any issue that arises.

Most IT organizations have teams of database administrators, system administrators, network administrators, and storage administrators. These administrators are responsible for system implementation and ongoing operations. In an Exadata Database Machine environment, it is usually more efficient and effective to have the database administrator to take the lead role for Exadata Database Machine management, with assistance from the system administrator. This is because Exadata Database Machine is engineered to run Oracle Database, and administration is specific to Oracle Database and Oracle Exadata System Software. The other teams may have distinct responsibilities or be a second level of support to provide assistance.

Usually there is one individual or group that has primary responsibility for any issue that arises. This individual or group receives the first contact from Oracle Enterprise Manager Cloud Control, the help desk, or operations team when there is an issue on the system. For Exadata Database Machine, the primary contact is typically the database administrator. If the database administrator needs assistance from another team to resolve the issue, then they collaborate to resolve the issue. Ownership of the issue should remain clear.

- [Common Administrative Tasks for Exadata Database Machine Management](#)
Initial system deployment is usually performed by Oracle engineers. The primary responsibilities for the database administrator begin with typical operational tasks.
- [Understanding the Administrative Differences with Exadata Database Machine](#)
Most administration tasks are similar on Exadata Database Machine servers as on traditional database servers and storage servers, but there are some differences.

1.1.1 Common Administrative Tasks for Exadata Database Machine Management

Initial system deployment is usually performed by Oracle engineers. The primary responsibilities for the database administrator begin with typical operational tasks.

Table 1-1 Common Administrative Tasks for Exadata Database Machine Management

Task or Event	Administrator	Actions
Slow performance	Database administrator System Administrator	Receive alerts from Oracle Enterprise Manager Cloud Control that performance thresholds have been exceeded. Review system performance, CPU, memory and I/O on all servers for unusual trends. Review database performance, wait events, locking, parallelism, and execution plans.
Patch application or upgrades	Database administrator System Administrator	Apply Oracle Exadata System Software patches or upgrades, and RDMA Network Fabric switch firmware upgrades. Apply Oracle Database patches, Oracle Grid Infrastructure patches or upgrades.
System outage or failure	Database administrator System Administrator	Connect to Integrated Lights Out Manager (ILOM), verify current system state, identify hardware issue or restart system, and review logs for root cause analysis. Check surviving instances for errors, monitor performance on surviving instances, and verify application functionality has not been disrupted.
Suspected network issues	Database administrator System Administrator	Inspect network interfaces for errors or dropped packets, check if any switches have restarted, and escalate to network administration team, as needed. Inspect database-side performance to assess impact, if any.
Backup database	Database administrator System Administrator	Run database backup routines, and ensure database server backups are completed.
Failed disk replacement	Database administrator System Administrator	Receive alerts about hardware replacement, verify Oracle Auto Service Request (ASR) has opened a service request, and verify operators will allow field service technician in the data center to replace drive or provide spare drive.

1.1.2 Understanding the Administrative Differences with Exadata Database Machine

Most administration tasks are similar on Exadata Database Machine servers as on traditional database servers and storage servers, but there are some differences.

The following list shows the differences and exceptions for Exadata Database Machine servers:

- Configuration settings for Exadata Database Machine database servers, RDMA Network Fabric switches, and other components have settings based on testing and performance criteria. Changing the configuration settings, such as database server firmware or kernel parameters, based on company policy or other reasons should be reviewed for the potential impact to Exadata Database Machine.
- Restarting a server incorrectly can disrupt the database. The storage servers have special procedures and guidelines that must be followed to minimize disruption, such as off-lining grid disks before restarting the server, and not restarting more than one server at a time.
- Storage servers cannot be modified the same way as the database servers. Network changes, such as those for the NTP servers or DNS servers, are done using the `ipconf` utility. Network changes cannot be done manually by editing the configuration files. In addition, no software or additional packages can be installed on the storage servers. This restriction includes monitoring software. Storage server system updates are provided by Oracle Exadata System Software upgrades.
- Storage servers do not require backups. A self-maintained internal USB drive or M.2 device that can be used for cell recovery. Backup clients cannot be installed on the storage servers.
- Oracle wait events in Oracle Real Application Clusters (Oracle RAC) databases using storage servers may include events with `%cell%` in the name. These events are related to the storage servers.
- The Oracle Database `V$CELL` views include rows for any database using Oracle Exadata Storage Server.
- Oracle Automatic Storage Management (Oracle ASM) disk path names are of the format `o/cell_ip_address/cell_griddisk_name`, such as the following:

```
o/192.168.10.1/data_CD_01_dm01cel01
```

- SQL plans may include `storage` to indicate that some operations may be off-loaded to the storage servers.
- Operations such as backup and recovery use Oracle Recovery Manager (RMAN), and all data for backup and recovery continues to pass through the database instances. The backup clients for RMAN should be installed on the database servers in Exadata Database Machine to facilitate integration with enterprise backup solutions in the same way as in traditional environments.
- The practice of deploying one or more non-production environments for development, testing and quality assurance still apply for Exadata Database Machine environments.

Related Topics

- [Shutting Down Exadata Storage Server](#)
When performing maintenance on Exadata Storage Servers, it may be necessary to power down or restart the cell.
- Using the ipconf Utility

1.2 Powering On and Off Oracle Exadata Rack

This section contains the procedures for powering on and off the components of an Oracle Exadata Rack.

- [Non-emergency Power Procedures](#)
When the outage is planned, use these procedures for powering on and off the components of Oracle Exadata Rack in an orderly fashion.
- [Emergency Power-off Considerations](#)
- [Cautions and Warnings](#)

1.2.1 Non-emergency Power Procedures

When the outage is planned, use these procedures for powering on and off the components of Oracle Exadata Rack in an orderly fashion.

- [Powering On Oracle Exadata Rack](#)
- [Powering On Servers Remotely using ILOM](#)
Servers can be powered on remotely using the Integrated Lights Out Manager (ILOM) interface.
- [Powering Off Oracle Exadata Rack](#)
Power off the components of the Oracle Exadata Rack in the correct order.
- [Powering On and Off Network Switches](#)

1.2.1.1 Powering On Oracle Exadata Rack

Oracle Exadata Rack is powered on by either pressing the power button on the front of the servers, or by logging in to the ILOM interface, and applying power to the system. When a database server is powered on and the operating system boots, Oracle Clusterware is automatically started, if it is installed. Oracle Clusterware then starts all resources that are configured to start automatically.

The power on sequence is as follows:

1. Rack, including switches.
Ensure the switches have had power applied for a few minutes to complete power-on configuration before starting Exadata Storage Servers.
2. Exadata Storage Servers.
Ensure all Exadata Storage Servers complete the boot process before starting the database servers. This may take five to ten minutes before all services start.
3. Database servers.

1.2.1.2 Powering On Servers Remotely using ILOM

Servers can be powered on remotely using the Integrated Lights Out Manager (ILOM) interface.

The ILOM can be accessed using the Web console, the command-line interface (CLI), IPMI, or SNMP. For example, to apply power to server `dm01ce101` using IPMI, where `dm01ce101-ilom` is the host name of the ILOM for the server to be powered on, run the following command from a server that has IPMITool installed:

```
# ipmitool -I lanplus -H dm01ce101-ilom -U root chassis power on
```

The preceding command causes the system to prompt for the password.

Related Topics

- [Oracle Integrated Lights Out Manager \(ILOM\) Documentation](#)

1.2.1.3 Powering Off Oracle Exadata Rack

Power off the components of the Oracle Exadata Rack in the correct order.

The power off sequence for Oracle Exadata Rack is as follows:

1. Database servers (Exadata Database Machine only).
2. Exadata Storage Servers.
3. Rack, including switches.

- [Powering Off Database Servers](#)
- [Powering Off Oracle Exadata Storage Servers](#)
Oracle Exadata Storage Servers are powered off and restarted using the Linux `shutdown` command.
- [Powering Off Multiple Servers at the Same Time](#)

1.2.1.3.1 Powering Off Database Servers

When powering off database servers, Oracle Clusterware should be stopped prior to restarting or shutting down a database server. Oracle Clusterware is stopped using the following command:

```
crsctl stop cluster
```

The following procedure is the recommended shutdown procedure for database servers:

1. Stop Oracle Clusterware using the following command:

```
# GRID_HOME/grid/bin/crsctl stop cluster
```

If any resources managed by Oracle Clusterware are still running after issuing the `crsctl stop cluster` command, then the command fails. Use the `-f` option to unconditionally stop all resources, and stop Oracle Clusterware.

2. Shut down the operating system using the following command:

```
# shutdown -h now
```

1.2.1.3.2 Powering Off Oracle Exadata Storage Servers

Oracle Exadata Storage Servers are powered off and restarted using the Linux `shutdown` command.

The following command shuts down Oracle Exadata Storage Server immediately:

```
# shutdown -h now
```

When powering off Oracle Exadata Storage Servers, all storage services are automatically stopped.

If you use the `-r` option, then the `shutdown` command shuts down and then restarts Oracle Exadata Storage Server. The `-now` option indicates you want to stop the server immediately.

```
# shutdown -r now
```

Another system command to reboot a server is the `reboot` command. However, `shutdown -r now` is the preferred command to restart a server. You should never use the command `reboot -f` to shut down Oracle Exadata Storage Servers.



Note:

Do not perform successive shutdown or reboot commands, which is essentially the same as `reboot -f`.

Note the following when powering off Oracle Exadata Storage Server:

- All database and Oracle Clusterware processes should be shut down prior to shutting down more than one Oracle Exadata Storage Server.
- Powering off one Oracle Exadata Storage Server does not affect running database processes or Oracle Automatic Storage Management (Oracle ASM).
- Powering off or restarting Oracle Exadata Storage Servers can impact database availability.
- The `shutdown` commands can be used to power off or reboot Oracle Exadata Storage Server.

 **See Also:**

- "Shutting Down Exadata Storage Server" if the databases or Oracle Clusterware will remain operational while powering down Oracle Exadata Storage Server
- SHUTDOWN(8) manual page for details.

1.2.1.3.3 Powering Off Multiple Servers at the Same Time

The `dcli` utility can be used to run the `shutdown` command on multiple servers at the same time. Do not run the `dcli` utility from a server that will be shut down. For example, to shut down all Exadata Storage Servers using the `dcli` utility, run the command from a database server. The following command shows the command syntax:

```
# dcli -l root -g group_name shutdown -h now
```

In the preceding syntax, `group_name` is the file that contains a list of all Exadata Storage Servers, `cell_group`, or database servers, `db_group`.

The following command shows the syntax to shut down all Exadata Storage Servers at the same time:

```
# dcli -l root -g cell_group shutdown -h now
```

[Example 1-1](#) shows the power off procedure for Oracle Exadata Rack when using the `dcli` utility to shut down multiple servers at the same time. The commands are run from a database server.

Example 1-1 Powering Off Oracle Exadata Rack Using the `dcli` Utility

1. Stop Oracle Clusterware on all database servers using the following command:

```
# GRID_HOME/grid/bin/crsctl stop cluster -all
```

2. Shut down all remote database servers using the following command:

```
# dcli -l root -g remote_db_group shutdown -h now
```

In the preceding command, `remote_db_group` is the file that contains a list of all the remote database servers.

3. Shut down all Exadata Storage Servers using the following command:

```
# dcli -l root -g cell_group shutdown -h now
```

In the preceding command, `cell_group` is the file that contains a list of all Exadata Storage Servers.

4. Shut down the local database server using the following command:

```
shutdown -h now
```

5. Remove power from the rack.

1.2.1.4 Powering On and Off Network Switches

The network switches do not have power switches. They power off when power is removed, by way of the power distribution unit (PDU) or at the breaker in the data center.

1.2.2 Emergency Power-off Considerations

If there is an emergency, then power to Oracle Exadata Rack should be halted immediately. The following emergencies may require powering off Oracle Exadata Rack:

- Natural disasters such as earthquake, flood, hurricane, tornado or cyclone.
- Abnormal noise, smell or smoke coming from the machine.
- Threat to human safety.
- [Emergency Power-off Procedure](#)
- [Emergency Power-off Switch](#)

1.2.2.1 Emergency Power-off Procedure

To perform an emergency power-off procedure for Oracle Exadata Rack, turn off power at the circuit breaker or pull the emergency power-off switch in the computer room. After the emergency, contact Oracle Support Services to restore power to the machine.

1.2.2.2 Emergency Power-off Switch

Emergency power-off (EPO) switches are required when computer equipment contains batteries capable of supplying more than 750 volt-amperes for more than five minutes. Systems that have these batteries include internal EPO hardware for connection to a site EPO switch or relay. Use of the EPO switch removes power from Oracle Exadata Rack.

1.2.3 Cautions and Warnings

The following cautions and warnings apply to Oracle Exadata Rack:

- Do not touch the parts of this product that use high-voltage power. Touching them might result in serious injury.
- Do not power off Oracle Exadata Rack unless there is an emergency. In that case, follow the [Emergency Power-off Procedure](#).
- Keep the front and rear cabinet doors closed. Failure to do so might cause system failure or result in damage to hardware components.
- Keep the top, front, and back of the cabinets clear to allow proper airflow and prevent overheating of components.
- Use only the supplied hardware.

1.3 Using Auto Service Request to Manage Hardware Faults

Auto Service Request (ASR) is designed to automatically open service requests when specific Oracle Exadata Rack hardware faults occur.

- [Understanding Auto Service Request](#)
- [Installing and Configuring ASR](#)
Oracle recommends installing Oracle Auto Service Request (ASR) on a standalone server running Oracle Solaris or Oracle Linux.

1.3.1 Understanding Auto Service Request

When a hardware problem is detected, Oracle ASR Manager submits a service request to Oracle Support Services. In many cases, Oracle Support Services can begin work on resolving the issue before the database administrator is even aware the problem exists. Oracle Auto Service Request (ASR) is designed to automatically open service requests when specific Oracle Exadata Rack hardware faults occur.

To enable this feature, the Oracle Exadata Rack components must be configured to send hardware fault telemetry to the Oracle ASR Manager software. This service covers components in storage servers and Oracle Database servers, such as disks and flash cards.

Oracle ASR Manager must be installed on a server that has connectivity to Oracle Exadata Rack, and an outbound Internet connection using HTTPS or an HTTPS proxy. Oracle recommends that Oracle ASR Manager be installed on a server outside of Oracle Exadata Rack. The following are some of the reasons for the recommendation:

- If the server or the rack containing Oracle ASR Manager goes down, then Oracle ASR Manager is unavailable for all of the Exadata Database Machine components that it supports. This is very important to consider when several Exadata Database Machines use the Oracle ASR Manager.
- In order to submit a service request (SR), the server must be able to access the Internet.



Note:

Oracle ASR can only use the management network. Ensure the management network is set up to allow Oracle ASR to run.

Prior to using Oracle ASR, the following must be set up:

- Oracle Premier Support for Systems or Oracle/Sun Limited Warranty
- Technical contact responsible for Oracle Exadata Rack
- Valid shipping address for Oracle Exadata Rack parts

An e-mail message is sent to the technical contact for the activated asset to notify the creation of the service request. The following are examples of the disk failure Simple Network Management Protocol (SNMP) traps sent to Oracle ASR Manager.

 **Note:**

- Oracle ASR is applicable only for component faults. Not all component failures are covered, though the most common components such as disk, fan, and power supplies are covered.
- Oracle ASR is not a replacement for other monitoring mechanisms, such as SMTP, and SNMP alerts, within the customer data center. Oracle ASR is a complementary mechanism that expedites and simplifies the delivery of replacement hardware. Oracle ASR should not be used for downtime events in high-priority systems. For high-priority events, contact Oracle Support Services directly.
- There are occasions when a service request may not be automatically filed. This can happen because of the unreliable nature of the SNMP protocol, or loss of connectivity to the Oracle ASR Manager. Oracle recommends that customers continue to monitor their systems for faults, and call Oracle Support Services if they do not receive notice that a service request has been automatically filed.
- Oracle ASR can monitor Sun Datacenter InfiniBand Switch 36 switches that have firmware release 2.1.2 and later in Exadata Database Machines running Oracle Exadata System Software release 11.2.3.3.0 or later. Switches may need a field engineer to set the entitlement serial number.

Example 1-2 Example of Exadata Storage Server SNMP Trap

This example shows the SNMP trap for a storage server disk failure. The corresponding hardware alert code has been highlighted.

```
2011-09-07 10:59:54 server1.example.com [UDP: [192.85.884.156]:61945]:
RFC1213-MIB::sysUpTime.0 = Timeticks: (52455631) 6 days, 1:42:36.31
SNMPv2-SMI::snmpModules.1.1.4.1.0 = OID: SUN-HW-TRAP-
MIB::sunHwTrapHardDriveFault
SUN-HW-TRAP-MIB::sunHwTrapSystemIdentifier = STRING: Sun Oracle Database
Machine
1007AK215C
SUN-HW-TRAP-MIB::sunHwTrapChassisId = STRING: 0921XFG004
SUN-HW-TRAP-MIB::sunHwTrapProductName = STRING: SUN FIRE X4270 M2 SERVER
SUN-HW-TRAP-MIB::sunHwTrapSuspectComponentName = STRING: SEAGATE
ST32000SSSUN2.0T;
Slot: 0SUN-HW-TRAP-MIB::sunHwTrapFaultClass = STRING: NULL
SUN-HW-TRAP-MIB::sunHwTrapFaultCertainty = INTEGER: 0
SUN-HW-TRAP-MIB::sunHwTrapFaultMessageID = STRING: HALRT-02001
SUN-HW-TRAP-MIB::sunHwTrapFaultUUID = STRING:
acb0a175-70b8-435f-9622-38a9a55ee8d3
SUN-HW-TRAP-MIB::sunHwTrapAssocObjectId = OID: SNMPv2-SMI::zeroDotZero
SUN-HW-TRAP-MIB::sunHwTrapAdditionalInfo = STRING: Exadata Storage Server:
cellname Disk Serial Number: E06S8K
server1.example.com failure trap.
```

Example 1-3 Example of Oracle Database Server SNMP Trap

This example shows the SNMP trap from an Oracle database server disk failure. The corresponding hardware alert code has been highlighted.

```
2011-09-09 10:59:54 dbserv01.example.com [UDP: [192.22.645.342]:61945]:
RFC1213-MIB::sysUpTime.0 = Timeticks: (52455631) 6 days, 1:42:36.31
SNMPv2-SMI::snmpModules.1.1.4.1.0 = OID: SUN-HW-TRAP-
MIB::sunHwTrapHardDriveFault
SUN-HW-TRAP-MIB::sunHwTrapSystemIdentifier = STRING: Sun Oracle
Database Machine
1007AK215C
SUN-HW-TRAP-MIB::sunHwTrapChassisId = STRING: 0921XFG004
SUN-HW-TRAP-MIB::sunHwTrapProductName = STRING: SUN FIRE X4170 M2
SERVER
SUN-HW-TRAP-MIB::sunHwTrapSuspectComponentName = STRING: HITACHI
H103030SCSUN300G
Slot: 0SUN-HW-TRAP-MIB::sunHwTrapFaultClass = STRING: NULL
SUN-HW-TRAP-MIB::sunHwTrapFaultCertainty = INTEGER: 0
SUN-HW-TRAP-MIB::sunHwTrapFaultMessageID = STRING: HALRT-02007
SUN-HW-TRAP-MIB::sunHwTrapFaultUUID = STRING:
acb0a175-70b8-435f-9622-38a9a55ee8d3
SUN-HW-TRAP-MIB::sunHwTrapAssocObjectId = OID: SNMPv2-SMI::zeroDotZero
SUN-HW-TRAP-MIB::sunHwTrapAdditionalInfo = STRING: Exadata Database
Server: db03
Disk Serial Number: HITACHI H103030SCSUN300GA2A81019GGDE5E
dbserv01.example.com failure trap.
```

Related Topics

- About Oracle Auto Service Request

1.3.2 Installing and Configuring ASR

Oracle recommends installing Oracle Auto Service Request (ASR) on a standalone server running Oracle Solaris or Oracle Linux.

After installation is complete, configure fault telemetry destinations for the servers on Exadata Database Machine. The Exadata Database Machine servers can be set up during initial configuration. Oracle Exadata Deployment Assistant (OEDA) collects the configuration information, and then configures the servers.



Note:

When configuring the Integrated Lights Out Manager (ILOM) alert settings, do not remove the rules at the top of the rule list. To add a new rule, enter the rule at the bottom of the rule list.

- To install and configure Oracle ASR after initial configuration, refer to the installation and configuration information available in *Oracle Auto Service Request Quick Installation Guide for Oracle Exadata Database Machine*.

 **See Also:**

- *Oracle Auto Service Request Quick Installation Guide for Oracle Exadata Database Machine*
- *Oracle Exadata Database Machine Installation and Configuration Guide* for information about the Oracle ASR configuration page of OEDA
- Oracle Auto Service Request user documentation at <http://www.oracle.com/technetwork/systems/asr/documentation/index.html>
- Contacting Oracle Support Services for additional information about contacting Oracle Support Services

1.4 Monitoring the System Using Oracle Enterprise Manager Cloud Control

Oracle Exadata Database Machine can be monitored by Oracle Enterprise Manager Cloud Control agents using the Oracle Exadata Plug-in and the Oracle Systems Infrastructure Plug-in. The Oracle Exadata Database Machine is discovered and monitored as a system target in Oracle Enterprise Manager Cloud Control. Individual database servers, storage servers, and switches are grouped together under the system target for the Oracle Exadata Database Machine so they can be monitored as a group.

The Oracle Exadata Storage Server metrics are collected and managed by Management Server (MS). When used with Oracle Enterprise Manager Cloud Control, the metrics are presented as Oracle Enterprise Manager Cloud Control metrics.

All Exadata server alerts are delivered to Oracle Enterprise Manager Cloud Control using SNMP. The Exadata hardware and software components are monitored by Integrated Lights Out Manager (ILOM) and Oracle Exadata System Software in the following ways:

- Hardware components are monitored by ILOM. When a hardware component reports a failure or an exceeded threshold, ILOM reports the failure as an SNMP trap to MS. MS processes the trap, creates an alert, and delivers the alert to the Oracle Enterprise Manager Cloud Control agent.
- Hardware and software components are also monitored by MS directly. When a failure or threshold is exceeded, MS processes the trap, creates an alert, and delivers the alert to the Oracle Enterprise Manager Cloud Control agent.

From the end-user perspective, there is no difference between the two types of alerts. The alert message contains the corrective action to resolve the alert.

 **See Also:**

- *Oracle Exadata Database Machine Installation and Configuration Guide* for information about installing Oracle Enterprise Manager Cloud Control
- [Exadata Storage Software Versions Supported by the Oracle Enterprise Manager Exadata Plug-in \(My Oracle Support Doc ID 1626579.1\)](#)
- *Oracle Enterprise Manager Exadata Management Getting Started Guide*
- *Oracle Enterprise Manager Cloud Control Administrator's Guide* for additional information about using blackout to avoid unnecessary alerts when doing repairs
- The MAA Best Practices area for Enterprise Manager at <http://www.oracle.com/goto/maa> for Enterprise Manager and Exadata manageability best practices

1.5 Monitoring the System Using Oracle Configuration Manager

Oracle Configuration Manager collects configuration information and uploads it to the Oracle repository.

When the configuration information is uploaded daily, Oracle Support Services can analyze the data and provide better service. When a service request is logged, the configuration data is associated with the service request. The following are some of the benefits of Oracle Configuration Manager:

- Reduced time for problem resolution
- Proactive problem avoidance
- Improved access to best practices, and the Oracle knowledge base
- Improved understanding of the customer's business needs
- Consistent responses and services

The Oracle Configuration Manager software is installed and configured in each `ORACLE_HOME` directory on a server. For clustered databases, only one instance is configured for Oracle Configuration Manager. A configuration script is run on every database on the server. The Oracle Configuration Manager collectors then send their data to a centralized Oracle repository.

Related Topics

- *Oracle Configuration Manager Installation and Administration Guide*
- *Oracle Configuration Manager Collection Overview*

1.6 Changing Component Passwords

The passwords for the components can be changed after initial configuration.

- [Changing the Database Server Passwords](#)

- [Changing the Exadata Storage Server Passwords](#)
As the `root` user, use the `passwd` command to change operating system user passwords on Exadata Storage Server.
- [Changing the Power Distribution Unit Password](#)
- [Changing the ILOM Password](#)
- [Changing the InfiniBand Switch Password](#)
This procedure describes how to change a password for the InfiniBand switch.
- [Changing the Cisco Management Network Switch or RoCE Network Fabric Switch Password](#)
To change the password for the Cisco Management Network Switch or RDMA over Converged Ethernet (RoCE) switch, you must connect to the switch and run a command on the switch.
- [Changing the KVM Password](#)

1.6.1 Changing the Database Server Passwords

The user accounts and GRUB passwords can be changed on the database servers. The default user accounts on the database server are `root`, and the software owner account. Typically the software owner account is `oracle` or `grid`.

- [Changing the User Account Password on the Database Server](#)
As the `root` user, use the `passwd` command to change operating system user passwords.
- [Changing the GRUB Account Password on the Database Server](#)
Use the `host_access_control` script to change the GRUB account password on database servers.

1.6.1.1 Changing the User Account Password on the Database Server

As the `root` user, use the `passwd` command to change operating system user passwords.

At the operating system prompt, use the following command, where `user_name` identifies the user account that you want to change:

```
# passwd user_name
```

Related Topics

- [Default Security Setting Enacted by OEDA](#)

1.6.1.2 Changing the GRUB Account Password on the Database Server

Use the `host_access_control` script to change the GRUB account password on database servers.

Run the following command as the `root` user to change the GRUB account password on the database server:

```
# /opt/oracle.cellos/host_access_control grub-password
```

1.6.2 Changing the Exadata Storage Server Passwords

As the `root` user, use the `passwd` command to change operating system user passwords on Exadata Storage Server.

The default user accounts on Exadata Storage Servers are `root`, `celladmin`, and `cellmonitor`.

At the operating system prompt, use the following command, where `user_name` identifies the user account that you want to change:

```
# passwd user_name
```

1.6.3 Changing the Power Distribution Unit Password

The default account user for the power distribution unit (PDU) is `admin`. The following procedure describes how to change the password for the PDU:

1. Use a Web browser to access the PDU metering unit by entering the IP address for the unit in the address line of the browser. The Current Measurement page appears.
2. Click **Network Configuration** in the upper left of the page.
3. Log in as the `admin` user on the PDU metering unit.
4. Locate the Admin/User fields. Only letters and numbers are allowed for user names and passwords.
5. Enter up to five users and passwords in the Admin/Users fields.
6. Designate each user to be either an administrator or user.
7. Click **Submit** to set the users and passwords.

1.6.4 Changing the ILOM Password

The default user account for the Integrated Lights Out Manager (ILOM) is `root`. The following procedure describes how to change the password for the ILOM:

1. Log in to the ILOM using SSH.
2. Use the following command to change the password:

```
set /SP/users/user_name password
```

In the preceding command, `user_name` is the user account to be changed. The following is an example of the command:

```
set /SP/users/user1 password
```

```
Changing password for user /SP/users/user1/password...
Enter new password:*****
Enter new password again:*****
New password was successfully set for user /SP/users/user1
```

1.6.5 Changing the InfiniBand Switch Password

This procedure describes how to change a password for the InfiniBand switch.

The default user accounts for the InfiniBand switch are `root`, `ilom-admin`, `ilom-user`, `ilom-operator`, and `nm2user`.

1. Log in to the InfiniBand switch with SSH using the following command:

```
ssh user_name@switch_name
```

In the preceding command, `user_name` is the name of the user, and `switch_name` is the name of the InfiniBand switch.

2. Check the firmware version of the switch.
3. Use the ILOM to change the password using the following commands:

```
ssh -l ilom-admin switch_name  
  
set /SP/users/user_name password
```

See Also:

Sun Datacenter InfiniBand Switch 36 Administration Guide for Firmware Version 2.1 at http://docs.oracle.com/cd/E36265_01/html/E36266/gentextid-2626.html#scrolltoc

1.6.6 Changing the Cisco Management Network Switch or RoCE Network Fabric Switch Password

To change the password for the Cisco Management Network Switch or RDMA over Converged Ethernet (RoCE) switch, you must connect to the switch and run a command on the switch.

- [Changing the Password for Cisco 93xx Switches](#)
Use the `change-password` command to change the password for the Cisco Nexus 93108-1G, Cisco Nexus 9348, or Cisco Nexus 9336C-FX2 switch.
- [Changing the Cisco 4948 Ethernet Switch Password](#)
You can change the passwords for both serial port access and SSH access to the switch.

1.6.6.1 Changing the Password for Cisco 93xx Switches

Use the `change-password` command to change the password for the Cisco Nexus 93108-1G, Cisco Nexus 9348, or Cisco Nexus 9336C-FX2 switch.

1. Access the switch using `ssh` or via the serial port.

```
my_host$ ssh admin@my_switch
User Access Verification
Password:
```

2. Change the password.

```
switch# change-password
Enter old password:
Enter new password:
Confirm new password:
```

3. Copy the running configuration to the startup configuration.

```
switch# copy running-config startup-config
```

4. Exit from the session.

```
switch# exit
```

1.6.6.2 Changing the Cisco 4948 Ethernet Switch Password

You can change the passwords for both serial port access and SSH access to the switch.

There are two different methods to access the switch. One is through a serial port and the other is through `ssh`.

- **Serial port:** When using serial port access, there are no user accounts, so the `enable` password is all that is required.
- **ssh:** When the switch is accessed via `ssh`, you must supply a user account and password before you can change the password.

During the installation of the system, the `admin` user is created and you can use this user to access the switch using `ssh`.

- [Changing the Cisco 4948 Ethernet Switch Password for Serial Port Access](#)
You can change the passwords for both serial port access and SSH access to the switch.
- [Changing the Cisco 4948 Ethernet Switch Password for Telnet or SSH Access](#)
You can change the passwords for both serial port access and SSH access to the switch.

1.6.6.2.1 Changing the Cisco 4948 Ethernet Switch Password for Serial Port Access

You can change the passwords for both serial port access and SSH access to the switch.

1. Access the switch using `telnet`, `ssh`, or via the serial port.

If you use the serial port for access, you will not be prompted for a user name or password, you will just get the prompt.

```
my_host> ssh admin@my_switch
Using keyboard-interactive authentication.
Password:
```

 **Note:**

- In Oracle Linux 5 update 5 or higher, `telnet` was removed for security reasons. You may need to install `telnet` client package in the compute node before you can access the switch using `telnet`.
- Before using SSH to access the switch, you must enable SSH in the switch following the steps described in "Configuring SSH on Cisco Catalyst 4948 Ethernet Switch" (My Oracle Support Doc ID 1415044.1).

2. Change to enable mode.

```
Switch> enable
```

3. Set the password.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# no enable password
Switch(config)# enable secret new_password
Switch(config)# end
Switch# write memory
*Sep 15 14:25:05.893:%SYS-5-CONFIG_I:Configured from console by console
Building configuration...
Compressed configuration from 2502 bytes to 1085 bytes [OK ]
```

4. Save the current configuration.

```
Switch# copy running-config startup-config
```

5. Exit from the session.

```
Switch# exit
```

1.6.6.2.2 Changing the Cisco 4948 Ethernet Switch Password for Telnet or SSH Access

You can change the passwords for both serial port access and SSH access to the switch.

- 1. Access the switch using `telnet`, `ssh`, or via the serial port.**

If you use the serial port for access, you will not be prompted for a user name or password, you will just get the prompt.

```
my_host> ssh admin@my_switch
Using keyboard-interactive authentication.
Password:
```

2. Change to enable mode.

```
Switch> enable
Password:
```

3. Verify the password will be sent in encrypted format.

Use the following command to check that service password configuration is set to -encryption.

```
Switch# show running-config all | include service password-
encryption
service password-encryption
```

If this is set to `no service password-encryption`, then passwords will be sent in clear text. You can change this setting, as shown in Step 5.

4. Enter configuration mode.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
```

5. If password encryption is set to `no service password-encryption`, then change it to `service password-encryption`.

```
Switch(config)# service password-encryption
```

6. Change the password for a specific user.

```
Switch(config)# username user_name password new_password
```

7. Exit configuration mode and save the changes.

```
Switch(config)# end
Switch# write memory
```

1.6.7 Changing the KVM Password

The default user account for the KVM is `Admin`. The following procedure describes how to change the password for the KVM:

1. Pull the KVM tray out from the front of the rack, and open it using the handle.
2. Touch the touch pad.
3. Toggle between the host and KVM interface by pressing the Ctrl key on the left side twice, similar to a double-click on a mouse.

4. Select **Local** from User Accounts.
5. Click **Admin** under Users.
6. Set a password for the `Admin` account. Do not modify any other parameters.
7. Click **Save**.

1.7 Determining the Server Model

Use the `exadata.img.hw` command to determine the model of the cell or database server.

```
/usr/sbin/exadata.img.hw --get model
```

Reference the following table for the server model names and numbers for Exadata Database Machine.

Table 1-2 Exadata Database Machine Server Models

Exadata Database Machine	Database Server Model	Exadata Storage Server Model
Exadata Database Machine X9M-2	ORACLE SERVER X9-2	ORACLE SERVER X9-2L (High Capacity) ORACLE SERVER X9-2L_EXTREME_FLASH
Exadata Database Machine X9M-8	ORACLE SERVER X8-8	ORACLE SERVER X9-2L (High Capacity) ORACLE SERVER X9-2L_EXTREME_FLASH
Exadata Database Machine X8M-2	ORACLE SERVER X8-2	ORACLE SERVER X8-2L (High Capacity) ORACLE SERVER X8-2L_EXTREME_FLASH
Exadata Database Machine X8-2	ORACLE SERVER X8-2	ORACLE SERVER X8-2L (High Capacity) ORACLE SERVER X8-2L_EXTREME_FLASH
Exadata Database Machine X8M-8	ORACLE SERVER X8-8	ORACLE SERVER X8-2L (High Capacity) ORACLE SERVER X8-2L_EXTREME_FLASH
Exadata Database Machine X8-8	ORACLE SERVER X8-8	ORACLE SERVER X8-2L (High Capacity) ORACLE SERVER X8-2L_EXTREME_FLASH
Exadata Database Machine X7-2	ORACLE SERVER X7-2	ORACLE SERVER X7-2L (High Capacity) ORACLE SERVER X7-2L_EXTREME_FLASH

Table 1-2 (Cont.) Exadata Database Machine Server Models

Exadata Database Machine	Database Server Model	Exadata Storage Server Model
Exadata Database Machine X7-8	ORACLE SERVER X7-8	ORACLE SERVER X7-2L (High Capacity) ORACLE SERVER X7-2L_EXTREME_FLASH
Exadata Database Machine X6-2	ORACLE SERVER X6-2	ORACLE SERVER X6-2L ORACLE SERVER X6-2L_EXTREME_FLASH
Exadata Database Machine X6-8	ORACLE SERVER X5-8	ORACLE SERVER X6-2L ORACLE SERVER X6-2L_EXTREME_FLASH
Exadata Database Machine X5-2	ORACLE SERVER X5-2	ORACLE SERVER X5-2L
Exadata Database Machine X5-8	ORACLE SERVER X5-8	ORACLE SERVER X5-2L
Exadata Database Machine X4-2	SUN SERVER X4-2	SUN SERVER X4-2L
Exadata Database Machine X4-8 Full Rack	SUN SERVER X4-8	SUN SERVER X4-2L ORACLE SERVER X5-2L
Exadata Database Machine X3-2	SUN FIRE X4170 M3	SUN FIRE X4270 M3
Exadata Database Machine X3-8 Full Rack	Sun Fire X4800 M2	SUN FIRE X4270 M3
Exadata Database Machine X2-2	SUN FIRE X4170 M2 SERVER	SUN FIRE X4270 M2 SERVER
Exadata Database Machine X2-8 Full Rack	Sun Fire X4800 or Sun Fire X4800 M2	SUN FIRE X4270 M2 SERVER

1.8 Monitoring Ambient Temperature of Servers

Maintaining environmental temperature conditions within design specification for Oracle Exadata Rack helps achieve maximum efficiency and targeted component service lifetimes. The impact of validating the ambient temperature range is minimal. The impact of corrective actions will vary depending on the environmental conditions.

The ambient temperature range for operating Oracle Exadata Rack is 5 to 32 degrees Celsius (41 to 89.6 degrees Fahrenheit). However, for maximum efficiency and component longevity, the optimal temperature range is 21 to 23 degrees Celsius (70 to 74 degrees Fahrenheit). For details, see Temperature and Humidity Requirements.

Use the following command as the `root` user on the first database server in the cluster to verify the temperature on all servers in the cluster:

```
dcli -g /opt/oracle.SupportTools/onecommand/all_group -l root
'ipmitool sunoem cli "show /SYS/T_AMB" | grep value'
```

The following is an example of the output from the command:

```
dm01db01: value = 21.440 degree C
dm01db02: value = 21.440 degree C
dm01db03: value = 22.190 degree C
...
dm01db08: value = 21.940 degree C
dm01cel01: value = 22.000 degree C
dm01cel02: value = 22.000 degree C
dm01cel03: value = 23.000 degree C
...
dm01cel14: value = 22.080 degree C
```

If the output is outside the ambient temperature range, then investigate and correct the problem. The following items should be checked:

- Sufficient air flow into the rack
- Room temperature is within the specified range
- Rear of rack is clear of obstructions

1.9 Replacing a Disk Controller Battery Backup Unit

The disk controller battery backup unit (disk controller BBU) resides on a drive tray in the database and Exadata Storage Servers. The disk controller BBU can be replaced without downtime for the server or the applications running on the server.

Note:

The procedures in this section do not apply to on-controller battery backup units. Replacement of those units require a system shutdown because the system must be opened to access the controller card.

- [Replacing a Disk Controller BBU on a Database Server](#)
- [Replacing a Disk Controller BBU on an Exadata Storage Server](#)
This section describes how to replace a disk controller BBU on an Exadata Storage Server.

1.9.1 Replacing a Disk Controller BBU on a Database Server

This section describes how to replace a disk controller BBU on a database server.

Note:

After any maintenance procedure, Oracle recommends using the Oracle EXAchk tool. The tool is available with My Oracle Support note 1070954.1.

The high-level steps are:

- [Step 1: Prepare the Disk Controller BBU for Removal](#)
The method of removing the Disk Controller BBU depends on the Oracle Exadata Database Machine model.
- [Step 2: Replace the Disk Controller BBU](#)
- [Step 3: Enable and Verify the New Disk Controller BBU](#)

1.9.1.1 Step 1: Prepare the Disk Controller BBU for Removal

The method of removing the Disk Controller BBU depends on the Oracle Exadata Database Machine model.

On certain Oracle Exadata Database Machine X3-2, X4-2, and X4-8 database nodes, and Oracle Exadata Database Machine X3-2, X4-2, X3-8, and X4-8 storage servers, the BBU is remote mounted and does not require a system shutdown to be accessed. However you must still prepare it for removal from the RAID HBA to avoid the risk of data corruption to the disk volumes.



Note:

There is no remote mount BBU option for Oracle Exadata Database Machine X3-8 database nodes.

- [Preparing Systems with Remote Mount BBU](#)
Describes how to prepare to remove the disk controller BBU on systems with remote mount BBU.
- [Preparing Systems That Do Not Have a Remote Mount BBU](#)
This topic describes how to prepare to remove the disk controller BBU on systems without a remote mount BBU.

1.9.1.1.1 Preparing Systems with Remote Mount BBU

Describes how to prepare to remove the disk controller BBU on systems with remote mount BBU.

If your system does not have a remote mount BBU, see [Preparing Systems That Do Not Have a Remote Mount BBU](#).

1. Log in as the `root` user.
2. Get the version of the image that is running on the server in the rack that requires service.

```
# imageinfo -ver  
11.2.3.2.1.130302
```

The version is the first five parts, such as `11.2.3.2.1` in the example. The last part is the image date.

3. Prepare the disk controller BBU for removal.

If you are running version 12.1.2.1.0 or later:

- a. Drop the disk controller BBU for replacement:

```
DBMCLI> alter dbserver bbu drop for replacement
```

- b. Verify the BBU status has been updated.

```
DBMCLI> list dbserver attributes bbustatus -  
dropped for replacement
```

If you are running versions between 11.2.3.3.0 and 12.1.2.1.0:

- a. Drop the disk controller BBU for replacement.

```
# /opt/oracle.cellos/compmon/exadata_mon_hw_asr.pl -  
drop_bbu_for_replacement
```

- b. Verify the status has been updated.

```
# /opt/oracle.cellos/compmon/exadata_mon_hw_asr.pl -list_bbu_status  
BBU status: dropped for replacement.
```

If you are running version 11.2.3.2.x:

- a. Locate the server in the rack being serviced, and turn on the indicator light.

Exadata Storage Servers are identified by a number 1 through 18, where 1 is the lowest Storage Server in the rack installed in RU2, counting up to the top of the rack.

Exadata Database Nodes are identified by a number 1 through 8, where 1 is the lowest most database node in the rack installed in RU16.

Turn on the locate indicator light for easier identification of the server being serviced. If the server number has been identified, then the Locate Button on the front panel may be pressed. To turn on the indicator light remotely, use any of the following methods:

- From a login to the CellCli on Exadata Storage Servers:

```
CellCli> alter cell led on
```

- From a login to the server's ILOM:

```
-> set /SYS/LOCATE value=Fast_Blink
```

- From a login to the server's root account:

```
# ipmitool chassis identify force  
Chassis identify interval: indefinite
```

- b. Check that HBA can see the battery and its current status.

 **Note:**

If you are running on Solaris, use `/opt/MegaRAID/MegaCli` in place of `/opt/MegaRAID/MegaCli/MegaCli64` in the commands below.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -adpbucmd -a0
```

The default output should show that the battery is still visible and may show low voltage or other issues depending on the fault. It may return an error reading the BBU if it is hard failed and no longer accessible to the HBA.

- c. Verify the current cache policy for all logical volumes.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep BBU
```

The default cache policy should be `WriteBack` for all volumes. If the battery is functioning normally it will report as current cache policy `WriteBack`. However if it is failed it may report current cache policy as `WriteThrough`.

- d. Set the cache policy for all logical volumes to `WriteThrough` cache mode, which does not use the battery.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldsetprop wt -lall -a0
```

- e. Verify the current cache policy for all logical volumes is now `WriteThrough`.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep BBU
```

1.9.1.1.2 Preparing Systems That Do Not Have a Remote Mount BBU

This topic describes how to prepare to remove the disk controller BBU on systems without a remote mount BBU.

If the system does not have the remote mounted battery installed, you need to shut down the node for which the battery requires replacement.

If your system has a remote mount BBU, see [Preparing Systems with Remote Mount BBU](#).

1. Revert all the RAID disk volumes to `WriteThrough` mode

This ensures that all data in the RAID cache memory is flushed to disk and is not lost when the battery is replaced.

 **Note:**

If you are running Oracle Exadata System Software 19.1.0 or above substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands.

- a. Set all logical volumes cache policy to `WriteThrough` cache mode.

 **Note:**

If you are running on Solaris, use `/opt/MegaRAID/MegaCli` in place of `/opt/MegaRAID/MegaCli/MegaCli64` in the commands below.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldsetprop wt -lall -a0
```

- b. Verify the current cache policy for all logical volumes is now `WriteThrough`, which does not use the battery.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep BBU
```

2. Shut down the server operating system.

- a. Perform the steps in [Steps To Shutdown/Startup The Exadata & RDBMS Services and Cell/Compute Nodes On An Exadata Configuration \(My Oracle Support Doc ID 1093890.1\)](#).
- b. Change the environment to point to the Oracle Grid Infrastructure Home.

Run the following commands as the `root` user, where the `1` of `+ASM1` refers to the database node number:

```
# . oraenv
ORACLE_SID = [root] ? +ASM1
The Oracle base for ORACLE_HOME=/u01/app/11.2.0/grid is /u01/app/
oracle
```

For example, for database node 3, the value would be `+ASM3`.

- c. Shut down Oracle Clusterware Services prior to powering down the database node.

Run the following commands as the `root` user:

```
# $ORACLE_HOME/bin/crsctl stop crs
```

Or:

```
# Grid_home/bin/crsctl stop crs
```

`Grid_home` is typically set to `/u01/app/11.2.0/grid`, but this can vary depending on your installation configuration.

- d. Verify that Oracle Clusterware Services have been stopped.

There should be no Clusterware processes running.

```
# ps -ef | grep css
```

- e. Shut down the server operating system.

- Linux:

```
# shutdown -hP now
```
- Solaris:

```
# shutdown -y -i 5 -g 0
```

1.9.1.2 Step 2: Replace the Disk Controller BBU

In this step, you remove the old disk controller BBU and replace it with the new BBU.

Exadata X3-2, X4-2, or X4-8 Compute Nodes and X3-2, X3-8, X4-2, or X4-8 Storage Cell nodes with the Remote Battery

These steps apply to Exadata nodes based on X3-2, X4-2, X4-8 and X3-2L, X4-2L servers with the remote battery installed.

1. Locate the battery slot marked with an orange and white BBU label.

X3-2 and X4-2 Compute nodes: this is the upper right-most slot on the front of the chassis labelled BBU (previously designated "HDD7").

X4-8 Compute nodes: this is in the lower slot, second from the left, on the rear of the chassis labeled BBU.

X3-2L and X4-2L Storage cells: this is the right-hand slot on the rear of the chassis above PS1, labelled BBU (previously designated "REAR HDD 1").
2. Unlatch and carefully slide out the old BBU carrier.
3. Insert and carefully slide in the new BBU carrier, and latch it closed.

Exadata X3-2L or X4-2L Storage Cell nodes without the Remote Battery

Replace the existing HBA BBU with a remote-mounted battery kit (part 7060020) following the CAP detailed in Support Note 1561949.1.

Exadata X3-2 or X4-2 Database Machine Compute nodes without the Remote Battery

Replace the existing HBA BBU with a remote-mounted battery kit (part 7060020) following the CAP detailed in Support Note 1561949.1.

Exadata X3-8 Database Machine Compute nodes

These steps are relevant to Exadata nodes based on X2-8 servers (formerly x4800m2).



Note:

The Exadata X3-8 Database Machine Compute nodes are based on X2-8 servers. See [Table 1-2](#)

1. Remove CMOD0 from the server and set it on a flat, antistatic surface.
2. Remove the CMOD top cover.
3. Remove the HBA REM with BBU attached:

- a. Lift the REM ejector handle and rotate it to its fully open position.
 - b. Lift the connector end of the REM and pull the REM away from the retaining clip on the front support bracket.
4. Remove the old BBU from the REM:
 - a. Use a No. 1 Phillips screwdriver to remove the 3 retaining screws that secure the battery to the REM card. Do NOT attempt to remove any screws from the top side of the REM and battery pack; those screws hold the standoffs that provide the bottom screw holes and should remain with the battery pack.
 - b. Detach the battery pack including circuit board from the REM by gently lifting it from its circuit board connector.
5. Install the new BBU on the REM.
 - a. Attach the battery pack circuit board connector to mate with the REM's connector.
 - b. Use a No. 1 Phillips screwdriver to secure the battery to the REM. If the BBU comes with a package of new screws, then use those new screws - do not re-use the screws from the old BBU attachment.
6. Re-install the HBA REM with BBU attached.
 - a. Ensure that the REM ejector lever is in the closed position. The lever should be flat with the REM support bracket.
 - b. Position the REM so that the battery is facing downward and the connector is aligned with the connector on the motherboard.
 - c. Slip the opposite end of the REM under the retaining clips on the front support bracket and ensure that the notch on the edge of the REM is positioned around the alignment post on the bracket.
 - d. Carefully lower and position the connector end of the REM until the REM contacts the connector on the motherboard, ensuring that the connectors are aligned. To seat the connector, carefully push the REM downward until it is in a level position.
7. Install the cover on the CMOD.
8. Return the CMOD back into the unit in CMOD0 slot.

1.9.1.3 Step 3: Enable and Verify the New Disk Controller BBU

Similar to [Step 1: Prepare the Disk Controller BBU for Removal](#), this step has two subsections:

- [For Systems with Remote Mount BBU](#)
- [For Systems That Do Not Have a Remote Mount BBU](#)

For Systems with Remote Mount BBU

For systems with remote mount BBU, the system was not shut down at the end of [Step 1: Prepare the Disk Controller BBU for Removal](#).

If you are using Oracle Exadata System Software version 11.2.3.3.0 or later:

 **Note:**

If you are running Oracle Exadata System Software 19.1.0 or later, substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands:

1. Log in as the root user.
2. Verify the disk controller BBU battery state is present and seen by the RAID controller. It may take several minutes for the new BBU battery to be detected.

 **Note:**

If you are running on Solaris, use `/opt/MegaRAID/MegaCli` in place of `/opt/MegaRAID/MegaCli/MegaCli64` in the commands below.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -AdpAllInfo -a0 | grep BBU
BBU : Present
BBU : Yes
Cache When BBU Bad : Disabled
```

3. Re-enable the disk controller BBU and disk cache.
 - If you are running Oracle Exadata System Software version 12.1.2.1.0 or later:

```
DBMCLI> ALTER DBSERVER BBU REENABLE
```

- If you are running Oracle Exadata System Software version earlier than 12.1.2.1.0:

```
# /opt/oracle.cellos/compmon/exadata_mon_hw_asr.pl -reenable_bbu
HDD disk controller battery has been reenabled.
```

4. Verify the disk controller BBU battery state is operational.
 - If you are running Oracle Exadata System Software version 12.1.2.1.0 or later:

```
DBMCLI> LIST DBSERVER ATTRIBUTES bbustatus
```

- If you are running Oracle Exadata System Software version earlier than 12.1.2.1.0:

```
# /opt/oracle.cellos/compmon/exadata_mon_hw_asr.pl -
list_bbu_status
BBU status: present
```

5. Verify the current logical disk drive cache policy uses writeback mode:

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep -i bbu
Default Cache Policy: WriteBack, ReadAheadNone, Direct, No Write Cache if Bad BBU
```

Current Cache Policy: WriteBack, ReadAheadNone, Direct, No Write Cache if Bad BBU
... <repeated for each logical volume present>

6. If the current cache policy is WriteThrough mode, and not WriteBack, then check the status of the battery.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -adpbucmd -getbbustatus -a0|grep
Battery
BatteryType: iBBU08
Battery State : Operational
Battery Pack Missing : No
Battery Replacement required : No
```

If the "Battery State" is anything other than "Operational" or "Optimal" (exact term depends on image version), investigate and correct the problem before continuing.

The following shows which image version uses "Optimal" and "Operational".

Exadata image version	Battery State	Raid f/w version
-----	-----	-----
X4 12.1.2.1.0	Optimal	12.12.0-0178
X4 12.1.1.1.1	Optimal	12.12.0-0178
X3 11.2.3.3.0	Optimal	12.12.0-0178
X3 11.2.3.2.2	Optimal	12.12.0-0178
X3 11.2.3.2.1	Operational	12.12.0-0140

If you are using image version 11.2.3.2.x:

1. Log in as the root user.
2. Turn off the server's locate LED.

```
# ipmitool chassis identify off
Chassis identify interval: off
```

3. Wait approximately 5 minutes for the HBA to recognize and start communicating with the new BBU.
4. Verify the HBA battery status is Operational and charging.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -adpbucmd -a0
```

5. Set all logical drives cache policy to WriteBack cache mode.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldsetprop wb -lall -a0
```

6. Verify the current cache policy for all logical drives is now using WriteBack cache mode.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpinfo -a0 | grep -i bbu
Default Cache Policy: WriteBack, ReadAheadNone, Direct, No Write Cache if
Bad BBU
Current Cache Policy: WriteBack, ReadAheadNone, Direct, No Write Cache if
Bad BBU
... <repeated for each logical volume present>
```

For Systems That Do Not Have a Remote Mount BBU

For systems that do not have a remote mount BBU, you shut down the system at the end of [Step 1: Prepare the Disk Controller BBU for Removal](#). In this section you restart the system and enable the new BBU.

Note:

If you are running Oracle Exadata System Software 19.1.0 or later, substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands:

1. Power on the server by pressing the power button.
2. After ILOM has booted, power on the server by pressing the power button, and then connect to the server's console.
To connect to the console from the ILOM Web browser (preferred): Access the "Remote Control -> Redirection" tab and click the "Launch Remote Console" button. On ILOM 3.1.x systems, the console button can be launched from the initial Summary Information screen.

To connect to the console from the ILOM CLI:

```
> start /SP/console
```

3. From the server's console, monitor the system booting. Watch in particular the LSI controller BIOS while it is loading. If it gives a warning message regarding drives with preserved cache, then choose "D" to discard the cache and continue. This is not an issue as the disk will get re-synced after boot by ASM. If it gives a warning message regarding drives are in write-through mode due to a low battery, then choose to continue.

The Exadata boot should continue normally after that, showing the Exadata boot splash screen and continue with normal OS boot messages. Note that there may be a long pause between screen outputs on the ILOM serial console during subsequent boot steps as the default console is the graphics, and the Exadata boot splash screen will not display.

4. Once full boot is completed, log in as the root user and verify the new battery is seen and is charging.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -adpbucmd -a0
```

5. Set all logical drives cache policy to WriteBack cache mode using the battery.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldsetprop wb -lall -a0
```

6. Verify the current cache policy for all logical drives is now using WriteBack cache mode.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep BBU
```

7. Verify the database services were started automatically.

- a. Verify that CRS is running.

```
# . oraenv
ORACLE_SID = [root] ? +ASM1
The Oracle base for ORACLE_HOME=/u01/app/11.2.0/grid is /u01/app/
oracle

# crsctl check crs
CRS-4638: Oracle High Availability Services is online
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
```

In the above output the 1 of +ASM1 refers to the database node number. For example, for database node #3, the value would be +ASM3.

- b. Validate that instances are running.

```
# ps -ef |grep pmon
```

It should return a record for ASM instance and a record for each database.

1.9.2 Replacing a Disk Controller BBU on an Exadata Storage Server

This section describes how to replace a disk controller BBU on an Exadata Storage Server.



Note:

After any maintenance procedure, Oracle recommends using the Oracle EXAchk tool. The tool is available with My Oracle Support note 1070954.1.

The high-level steps are:

- [Step 1: Prepare the Disk Controller BBU for Removal](#)
- [Step 2: Replace the Disk Controller BBU](#)
- [Step 3: Enable the New Disk Controller BBU](#)

1.9.2.1 Step 1: Prepare the Disk Controller BBU for Removal

On certain X3-2, X4-2, and X4-8 database nodes, and X3-2, X4-2, and X3-8, X4-8 storage servers, the BBU is remote mounted and does not require a system shutdown to be accessed. However you must still prepare it for removal from the RAID HBA to avoid the risk of data corruption to the disk volumes. Note there is no remote mount BBU option for X3-8 database nodes.

For Systems with Remote Mount BBU

Perform the steps in this section if your system has a remote mount BBU. If your system does not have a remote mount BBU, perform the steps in "[For Systems That Do Not Have a Remote Mount BBU](#)".

1. Log in as the root user.

2. Get the version of the image that is running on the server in the rack that requires service.

```
# cellcli -e LIST CELL ATTRIBUTES releaseVersion
11.2.3.2.1
```

3. Drop the disk controller BBU.

If you are running version 11.2.3.3.0 or later:

- a. Drop the disk controller BBU for replacement. Run the following command as the `celladmin` or `root` user:

```
# cellcli -e ALTER CELL BBU DROP FOR REPLACEMENT
HDD disk controller battery has been dropped for replacement
```

- b. Verify that the BBU was dropped for replacement:

```
# cellcli -e LIST CELL ATTRIBUTES bbustatus
dropped for replacement.
```

If you are running version 11.2.3.2.x:

- a. Locate the server in the rack being serviced, and turn on the indicator light. Exadata Storage Servers are identified by a number 1 through 18, where 1 is the lowest Storage Server in the rack installed in RU2, counting up to the top of the rack.

Exadata Database Nodes are identified by a number 1 through 8, where 1 is the lowest most database node in the rack installed in RU16.

Turn on the locate indicator light for easier identification of the server being serviced. If the server number has been identified, then the Locate Button on the front panel may be pressed.

To turn on the indicator light remotely, use any of the following methods:

From a login to the CellCli on Exadata Storage Servers:

```
CellCli> ALTER CELL LED ON
```

From a login to the server's ILOM:

```
-> set /SYS/LOCATE value=Fast_Blink
```

From a login to the server's root account:

```
# ipmitool chassis identify force
Chassis identify interval: indefinite
```

- b. Check that HBA can see the battery and its current status.

 **Note:**

If you are running on Solaris, use `/opt/MegaRAID/MegaCli` in place of `/opt/MegaRAID/MegaCli/MegaCli64` in the commands below.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -adpbbucmd -a0
```

The default output should show that the battery is still visible and may show low voltage or other issues depending on the fault. It may return an error reading the BBU if it is hard failed and no longer accessible to the HBA.

- c. Verify the current cache policy for all logical volumes.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep BBU
```

The default cache policy should be `WriteBack` for all volumes. If the battery is functioning normally it will report as current cache policy `WriteBack`. However if it is failed it may report current cache policy as `WriteThrough`.

- d. Set the cache policy for all logical volumes to `WriteThrough` cache mode, which does not use the battery.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldsetprop wt -lall -a0
```

- e. Verify the current cache policy for all logical volumes is now `WriteThrough`.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep BBU
```

For Systems That Do Not Have a Remote Mount BBU

Perform the steps in this section if your system does not have a remote mount BBU. If your system has a remote mount BBU, see "[For Systems with Remote Mount BBU](#)".

If the system does not have the remote mounted battery installed, you need to shut down the node for which the battery requires replacement.

 **Note:**

If you are running Oracle Exadata System Software 19.0 or later, substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands:

1. Revert all the RAID disk volumes to `WriteThrough` mode to ensure all data in the RAID cache memory is flushed to disk and not lost when replacement of the battery occurs.

- a. Set all logical volumes cache policy to `WriteThrough` cache mode.

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldsetprop wt -lall -a0
```

- b. Verify the current cache policy for all logical volumes is now WriteThrough, which does not use the battery:

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep BBU
```

2. Shut down the server operating system.

Note the following when powering off Exadata Storage Servers:

- Verify there are no other storage servers with disk faults. Shutting down a storage server while another disk is failing may cause database processes and Oracle ASM to crash if it loses both disks in the partner pair when this server's disks go offline.
- Powering off one Exadata Storage Server with no disk faults in the rest of the rack will not affect running database processes or Oracle ASM.
- All database and Oracle Clusterware processes should be shut down prior to shutting down more than one Exadata Storage Server. Refer to the Exadata Owner's Guide for details if this is necessary.

ASM drops a disk shortly after it is taken offline. Powering off or restarting Exadata Storage Servers can impact database performance if the storage server is offline for longer than the ASM disk repair timer to be restored. The default `DISK_REPAIR_TIME` attribute value of 3.6hrs should be adequate for replacing components, but may need to be changed if you need more time.

- a. Check the disk repair time by logging into ASM and running the following query.

```
SQL> SELECT dg.name,a.value FROM v$asm_attribute a,
v$asm_diskgroup dg
WHERE a.name = 'disk_repair_time' AND a.group_number =
dg.group_number;
```

As long as the value is large enough to comfortably replace the components being replaced, there is no need to change it.

If you need to change it, you can use this statement:

```
SQL> ALTER DISKGROUP DATA SET ATTRIBUTE
'disk_repair_time'='8.5H';
```

- b. Check if ASM will be OK if the grid disks go offline. The following command should return Yes for the grid disks being listed.

```
# cellcli -e LIST GRIDDISK ATTRIBUTES name,asmmodestatus,asmdeactivationoutcome
...sample ...
DATA_CD_09_cel01 ONLINE Yes
DATA_CD_10_cel01 ONLINE Yes
DATA_CD_11_cel01 ONLINE Yes
RECO_CD_00_cel01 ONLINE Yes
RECO_CD_01_cel01 ONLINE Yes
...repeated for all griddisks....
```

If one or more disks does not return `asmdeactivationoutcome='Yes'`, check the respective disk group and restore the data redundancy for that disk group.

Once the disk group data redundancy is fully restored, re-run the command to verify that `asmdeactivationoutcome='Yes'` for all grid disks. Once all disks return `asmdeactivationoutcome='Yes'`, proceed to the next step.

 **Note:**

Shutting down the cell services when one or more grid disks does not return `asmdeactivationoutcome='Yes'` will cause Oracle ASM to dismount the affected disk group, causing the databases to shut down abruptly.

- c. Inactivate all grid disks on the cell that needs to be powered down for maintenance. This could take up to 10 minutes or longer.

```
# cellcli
...sample ...
CellCLI> ALTER GRIDDISK ALL INACTIVE
GridDisk DATA_CD_00_dmorlx8cel01 successfully altered
GridDisk DATA_CD_01_dmorlx8cel01 successfully altered
GridDisk DATA_CD_02_dmorlx8cel01 successfully altered
GridDisk RECO_CD_00_dmorlx8cel01 successfully altered
GridDisk RECO_CD_01_dmorlx8cel01 successfully altered
GridDisk RECO_CD_02_dmorlx8cel01 successfully altered
...repeated for all griddisks...
```

- d. Verify that the grid disks are now offline. The output should show `asmmodestatus='UNUSED'` or `'OFFLINE'` and `asmdeactivationoutcome=Yes` for all grid disks once the disks are offline and inactive in ASM.

```
CellCLI> LIST GRIDDISK ATTRIBUTES
name,status,asmmodestatus,asmdeactivationoutcome
DATA_CD_00_dmorlx8cel01 inactive OFFLINE Yes
DATA_CD_01_dmorlx8cel01 inactive OFFLINE Yes
DATA_CD_02_dmorlx8cel01 inactive OFFLINE Yes
RECO_CD_00_dmorlx8cel01 inactive OFFLINE Yes
RECO_CD_01_dmorlx8cel01 inactive OFFLINE Yes
RECO_CD_02_dmorlx8cel01 inactive OFFLINE Yes
...repeated for all griddisks...
```

- e. Once all disks are offline and inactive, you can shut down the cell.

```
# shutdown -hP now
```

When powering off Exadata Storage Servers, all storage services are automatically stopped.

1.9.2.2 Step 2: Replace the Disk Controller BBU

See "[Step 2: Replace the Disk Controller BBU](#)".

1.9.2.3 Step 3: Enable the New Disk Controller BBU

Similar to "Step 1: Prepare the Disk Controller BBU for Removal", this section has two subsections:

- [For Systems with Remote Mount BBU](#)
- [For Systems That Do Not Have Remote Mount BBU](#)

For Systems with Remote Mount BBU

Perform the steps in this section if your system has a remote mount BBU. In this scenario, the system was not shut down at the end of "Step 1: Prepare the Disk Controller BBU for Removal".

If you are running image version 11.2.3.3.0 or later:

1. Log in as the celladmin or root user.
2. Re-enable the BBU.

```
# cellcli -e alter cell bbu reenable
HDD disk controller battery has been reenabled
```

3. Verify the disk controller BBU battery state is operational.

```
# cellcli -e list cell attributes bbustatus
normal
```

If the "BBU status" is anything other than "normal", then investigate and correct the problem before continuing.

If you are running image version 11.2.3.2.x:

1. Log in as the root user.
 2. Turn off the server's locate LED.
- ```
ipmitool chassis identify off
Chassis identify interval: off
```
3. Wait approximately 5 minutes for the HBA to recognize and start communicating with the new BBU.

4. Verify the HBA battery status is Operational and charging.

```
/opt/MegaRAID/MegaCli/MegaCli64 -adpbucmd -a0
```

5. Set all logical drives cache policy to WriteBack cache mode.

```
/opt/MegaRAID/MegaCli/MegaCli64 -ldsetprop wb -lall -a0
```

6. Verify the current cache policy for all logical drives is now using WriteBack cache mode.

```
/opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep -i bbu
Default Cache Policy: WriteBack, ReadAheadNone, Direct, No Write Cache if
Bad BBU
Current Cache Policy: WriteBack, ReadAheadNone, Direct, No Write Cache if
Bad BBU
... <repeated for each logical volume present>
```

#### For Systems That Do Not Have Remote Mount BBU

At the end of "[Step 1: Prepare the Disk Controller BBU for Removal](#)", systems without a remote mount BBU were shut down. You now have to restart the system.

1. Power on the server by pressing the power button.
2. After ILOM has booted, power on the server by pressing the power button, and then connect to the server's console.

To connect to the console from the ILOM Web browser (preferred): Access the "Remote Control -> Redirection" tab and click the "Launch Remote Console" button. On ILOM 3.1.x systems, the console button can be launched from the initial Summary Information screen.

To connect to the console from the ILOM CLI:

```
> start /SP/console
```

3. From the server's console, monitor the system booting. Watch in particular the LSI controller BIOS while it is loading. If it gives a warning message regarding drives with preserved cache, then choose "D" to discard the cache and continue. This is not an issue as the disk will get re-synced after boot by ASM. If it gives a warning message regarding drives are in write-through mode due to a low battery, then choose to continue.

The Exadata boot should continue normally after that, showing the Exadata boot splash screen and continue with normal OS boot messages. Note that there may be a long pause between screen outputs on the ILOM serial console during subsequent boot steps as the default console is the graphics, and the Exadata boot splash screen will not display.

4. Once full boot is completed, log in as the root user and verify the new battery is seen and is charging.

```
/opt/MegaRAID/MegaCli/MegaCli64 -adpbucmd -a0
```

5. Set all logical drives cache policy to WriteBack cache mode using the battery.

```
/opt/MegaRAID/MegaCli/MegaCli64 -ldsetprop wb -lall -a0
```

6. Verify the current cache policy for all logical drives is now using WriteBack cache mode.

```
/opt/MegaRAID/MegaCli/MegaCli64 -ldpdinfo -a0 | grep BBU
```

7. Return the cell back to service.

- a. Activate the grid disks.

```
cellcli
CellCLI> alter griddisk all active
GridDisk DATA_CD_00_dmorlx8cel01 successfully altered
GridDisk DATA_CD_01_dmorlx8cel01 successfully altered
GridDisk DATA_CD_02_dmorlx8cel01 successfully altered
GridDisk RECO_CD_00_dmorlx8cel01 successfully altered
GridDisk RECO_CD_01_dmorlx8cel01 successfully altered
GridDisk RECO_CD_02_dmorlx8cel01 successfully altered
...etc...
```

- b. Verify that all disks are active.

```
CellCLI> list griddisk
DATA_CD_00_dmorlx8cel01 active
DATA_CD_01_dmorlx8cel01 active
DATA_CD_02_dmorlx8cel01 active
RECO_CD_00_dmorlx8cel01 active
RECO_CD_01_dmorlx8cel01 active
```

```
RECO_CD_02_dmorlx8cel01 active
...etc...
```

- c. Verify all grid disks have been successfully put online. Wait until 'asmmodestatus' is in status 'ONLINE' for all grid disks. The following is an example of the output early in the activation process.

```
CellCLI> list griddisk attributes
name,status,asmmodestatus,asmdeactivationoutcome
DATA_CD_00_dmorlx8cel01 active ONLINE Yes
DATA_CD_01_dmorlx8cel01 active ONLINE Yes
DATA_CD_02_dmorlx8cel01 active ONLINE Yes
RECO_CD_00_dmorlx8cel01 active SYNCING Yes
RECO_CD_01_dmorlx8cel01 active ONLINE Yes
...etc...
```

In the example above 'RECO\_CD\_00\_dmorlx8cel01' is still in the 'SYNCING' process. Oracle ASM synchronization is only complete when ALL grid disks show 'asmmodestatus=ONLINE'. This process can take some time depending on how busy the machine is, and has been while this individual server was down for repair.

## 1.10 Overview of the dbmsrv Service

Starting with Oracle Exadata System Software release 12.1.2.1.0:

- The database nodes now run the Management Server (MS). Previously MS ran only on the storage nodes.
- The database nodes now run a new service called Database Machine Service (`dbmsrv`). This new service is based on the MS that runs on the storage servers and provides enhanced management capabilities to the database nodes.
- Starting with Oracle Exadata System Software release 12.1.2.1.2, Management Server (MS) on the database nodes does not use `sudo` any more. This means that configuration for sudoers is no longer needed.

Prior to Oracle Exadata System Software release 12.1.2.1.2:

- For security reasons, Management Server on the database nodes is not run as `root`. However, it needs `root` permission to run certain utilities that monitor the system, such as disk status, ILOM, power supply unit, and to send Oracle Auto Service Request (ASR) messages and alerts. To achieve this, a sudoers configuration file, `dbmsvc_sudo_conf`, is added to enable the Management Server users on the database nodes to run the utilities with `root` privilege.

You should not disable the `dbmsrv` service or `dbserverd`, or edit the sudoers configuration file. If the entries in the file are removed, then the `dbmsrv` service may not be able to monitor some parts of the system. For example, if a disk fails, it might not be possible to send an Oracle ASR message in time, and this may cause a disruption on the database node and delay recovery.

To manage the new Management Server on the database node service in Oracle Exadata System Software release 12.1.2.1.0 and later, new users and groups were added.

The users and their IDs are:

**dbmsvc:** 12137  
**dbmadmin:** 12138  
**dbmmonitor:** 12139

The new groups and their IDs are:

**dbmsvc:** 11137  
**dbmadmin:** 11138  
**dbmmonitor:** 11139  
**dbmusers:** 11140

The following topics describe how to modify user IDs for this service:

- [Using a Script to Change User IDs and Group IDs for dbmsrv](#)  
Starting with Oracle Exadata System Software releases 18.1.12 and 19.1.2, you can use the `migrate_ids.sh` script to change the user and group IDs for the dbmsrv users.
- [Manually Changing User IDs and Group IDs for dbmsrv](#)  
Prior to Oracle Exadata System Software releases 18.1.12 and 19.1.2 when the `migrate_ids.sh` script was introduced, you have to manually change the user and group IDs for the dbmsrv users.

## 1.10.1 Using a Script to Change User IDs and Group IDs for dbmsrv

Starting with Oracle Exadata System Software releases 18.1.12 and 19.1.2, you can use the `migrate_ids.sh` script to change the user and group IDs for the dbmsrv users.

You can change the user ID and group ID of the dbmsrv service users if there are conflicts with the default values (for example, if you are using LDAP or if you are using session management tools that require different values from the default values).

These steps are specific to the dbmsrv service users and groups only. Do not use them to modify the user and group IDs for other Oracle products.

1. Navigate to the `opt/oracle.SupportTools` directory.
2. Run the `migrate_ids.sh` script.

The `migrate_ids.sh` script has the following syntax and options:

```
migrate_ids.sh [-uid username new_uid]
 [-gid group_name new_group_id]
 [-skipdirs directory_path [,directory_path]]
```

- `-uid`: Specify user name and the new uid to migrate the user to a new UID
- `-gid`: Specify group name and the new group ID to migrate the group to a new ID
- `-skipdirs`: Specify a list of absolute paths of directories to skip during the user or group ID migration.

The script searches all directories to find files that use the uid or gid being migrated so that the script can update the owner or group access to use the new uid or gid. The `-skipdirs` option allows you to specify which directories do not need to be searched. The specified directories and any files within them are skipped while changing the uid and gid values.

Using the `-skipdirs` option can be useful if you have large NFS directories that you want to skip to make the migration faster. However, if there are files in the directories being skipped that use the uid or gid being migrated, then those files are not updated. It is your

responsibility to make sure that the directories being skipped with this option do not contain such files to ensure successful migration of the IDs.

**Example 1-4 Migrate the dbmadmin user to a new user ID**

This example shows how to migrate only the uid of user dbmadmin to 3001.

```
migrate_ids.sh -uid dbmadmin 3001
```

**Example 1-5 Migrate the dbmusers group to a new group ID**

This example shows how to migrate only the gid of group dbmusers to 4001.

```
migrate_ids.sh -gid dbmusers 4001
```

**Example 1-6 Migrate all dbmsrv users and groups to new values**

This example shows how to migrate all the user and group IDs for dbmsrv to new values.

```
migrate_ids.sh -uid dbmsvc 3001 -gid dbmsvc 4001
migrate_ids.sh -uid dbmadmin 3002 -gid dbmadmin 4002
migrate_ids.sh -uid dbmmonitor 3003 -gid dbmmonitor 4003
migrate_ids.sh -gid dbmusers 4004
```

**Example 1-7 Migrate a user ID while skipping directories**

This example shows how to migrate the user ID of user dbmadmin to 3001 while not searching the files in the `/proc` or `/sys` directories.

```
migrate_ids.sh -uid dbmadmin 3001 -skipdirs /proc,/sys
```

**Related Topics**

- [Overview of the dbmsrv Service](#)

## 1.10.2 Manually Changing User IDs and Group IDs for dbmsrv

Prior to Oracle Exadata System Software releases 18.1.12 and 19.1.2 when the `migrate_ids.sh` script was introduced, you have to manually change the user and group IDs for the dbmsrv users.

You can change the user ID and group ID of the dbmsrv service users if there are conflicts with the default values (for example, if you are using LDAP or if you are using session management tools that require different values from the default values).

If possible, you should upgrade to the latest version of Oracle Exadata System Software and use the `migrate_ids.sh` script instead of using the manual procedure.

These steps are specific to the dbmsrv service users and groups only. Do not use them to modify the user and group IDs for other Oracle products.



1. Shut down the services on the database server. Run the following command as `root` or the `dbmadmin` user.

```
dbmcli -e alter dbserver shutdown services all
```

2. Change the group ID of the group.

- a. Change the assigned group ID for the group.

Run the following command as `root`, where *new\_group\_ID* is the new group ID, and *group\_name* the name of group you want to change:

```
groupmod -g new_group_ID group_name
```

For example:

```
groupmod -g 3001 dbmusers
```

- b. Update the files containing the old group ID.

Run the following command as `root`:

```
find / -gid old_group_ID -exec chgrp -h new_group_ID {} \;
```

For example:

```
find / -gid 11140 -exec chgrp -h 3001 {} \;
```

3. Change the user ID.

This step has to be done after changing the group ID or you will get a "GID does not exist" error.

- a. Change the user ID assigned to the user.

Run the following command as `c`, where *new\_user\_ID* is the new ID for the user, *new\_group\_ID* is the new group ID assigned in the previous step, and *username* is the name of the user you want to change.

```
usermod -u new_user_ID -g new_group_ID username
```

For example:

```
usermod -u 2998 -g 3001 dbmsvc
```

- b. Update the files containing the old user ID.

Run the following command as `root`:

```
find / -uid old_user_ID -exec chown -h new_user_ID {} \;
```

For example:

```
find / -uid 12137 -exec chown -h 2998 {} \;
```

#### 4. Reset the setuid bit on the executable files.

The setuid bit was changed by the `chgrp` and `chown` commands. Perform the following sub-steps as `root`.

##### a. Modify the permissions for the `dbrsMain` executable.

```
chmod 6550 /opt/oracle/dbserver/dbms/bin/dbrsMain
```

##### b. Modify the permissions for the `exaCmdHelper` executable.

```
chmod 4550 /opt/oracle/dbserver/dbms/bin/exaCmdHelper
```

#### 5. Restart the services on the database server.

Run the following command as the `root` or `dbmadmin` user:

```
dbmcli -e alter dbserver startup services all
```

### Related Topics

- [Overview of the dbmsrv Service](#)

## 1.11 Configuring Password Expiration for Users Accessing the Server Remotely

You can configure `DBSERVER` attributes to expire user passwords.

In Oracle Exadata System Software release 19.1.0, there are new `DBSERVER` attributes for configuring password security for users that access Oracle Exadata System Software servers remotely, such as with REST API or ExaCLI. These attributes determine if the user is able to change the password remotely, the amount of time before a user password expires, and the number of days prior to password expiration that the user receives warning messages. In the default configuration, user passwords do not expire.



### Note:

The `DBSERVER` attributes for password expiration apply only to users created with Oracle Exadata System Software. Password expiration applies only to users that are displayed with the `LIST USER` command and does not apply to operating system users like `dbmadmin` or `oracle`.

- To allow the user to change the password remotely, use the `ALTER DBSERVER` command to set the `remotePwdChangeAllowed` attribute to `true`.

If you set the value to `false`, then the user receives a message indicating that they must contact the server administrator to have their password changed.

```
DBMCLI> ALTER DBSERVER remotePwdChangeAllowed=true
```

- To change the length of time before a user password expires, use the `ALTER DBSERVER` command to modify the `pwdExpInDays` attribute.

Set the value *n* to the number of days before the password expires. If `pwdExpInDays` is set to 0 (the default value), then the user password does not expire.

```
DBMCLI> ALTER DBSERVER pwdExpInDays=n
```

- To configure the length of the warning period before the password expires, use the `ALTER DBSERVER` command to modify the `pwdExpWarnInDays` attribute.

Set the value *n* to the number of days to warn the user before the password expires. The default user account password expiration warning time is 7 days.

```
DBMCLI> ALTER DBSERVER pwdExpWarnInDays=n
```

- To specify the length of time before a user account is locked after the user password expires, use the `ALTER DBSERVER` command to modify the `accountLockInDays` attribute.

Set the value *n* to the number of days before the user account is locked. The default user account lock time is 7 days.

```
DBMCLI> ALTER DBSERVER accountLockInDays=n
```

## 1.12 State of Storage Server and Database Servers During Configuration Changes

Before a change to a configuration, determine if the database and storage servers need to be offline or online.

**Table 1-3 State of Storage Server and Database Servers for Operations**

| Operation                                                        | Storage Server                                                                | Database Server                                                               |
|------------------------------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| DNS server update                                                | Online                                                                        | Online                                                                        |
| NTP server update                                                | Online                                                                        | Online                                                                        |
| Time zone update                                                 | Offline                                                                       | Online                                                                        |
| Admin network IP address, netmask, gateway, or host name change  | Offline                                                                       | Online                                                                        |
| Client network IP address, netmask, gateway, or host name change | Offline                                                                       | Online                                                                        |
| Integrated Lights Out Manager (ILOM) IP address change           | Offline                                                                       | Online if the <code>ipmitool sunoem getval/setval</code> command is supported |
| Other ILOM parameter change                                      | Online if the <code>ipmitool sunoem getval/setval</code> command is supported | Online if the <code>ipmitool sunoem getval/setval</code> command is supported |

**Table 1-3 (Cont.) State of Storage Server and Database Servers for Operations**

| Operation                                                    | Storage Server | Database Server |
|--------------------------------------------------------------|----------------|-----------------|
| RDMA Network Fabric IP address, netmask, or host name change | Offline        | Online          |
| Partition key (pkey) change                                  | Offline        | Online          |

## 1.13 Rescue Plan

In Exadata releases earlier than 12.2.1.1.0, after a storage server or database server rescue, you need to re-run multiple commands to configure items such as IORM plans, thresholds, and storage server and database server notification setting.

In Oracle Exadata release 12.2.1.1.0, there is a new attribute called `rescuePlan` for the `cell` and `dbserver` objects. When you are done configuring your database servers and storage servers, you should save the value of the `rescuePlan` attribute to a file. The file should be saved to a remote server because the data on the rescued server will be erased in the event of a rescue. After you rescue the server, you can retrieve the file from the remote server and run the file to restore the settings. See Example 3 below.

For security reasons, the rescue plan does not include configurations that require a password.

### Example 1-8 Rescue Plan for a Storage Cell

The `rescuePlan` attribute for a storage server could look like this:

```
$ cellcli -e list cell attributes rescuePlan

CREATE ROLE "admin"

GRANT PRIVILEGE all actions ON diagpack all attributes WITH all
options TO ROLE "admin"

CREATE ROLE "diagRole"

GRANT PRIVILEGE download ON diagpack all attributes WITH all options
TO ROLE "diagRole"

GRANT PRIVILEGE create ON diagpack all attributes WITH all options TO
ROLE "diagRole"

GRANT PRIVILEGE list ON diagpack all attributes WITH all options TO
ROLE "diagRole"

ALTER CELL accessLevelPerm="remoteLoginEnabled", diagHistoryDays="7",
metricHistoryDays="7", notificationMethod="mail,snmp",
notificationPolicy="warning,critical,clear",
snmpSubscriber=((host="localhost", port=162, community="public",
type=asr)),
bbuLearnCycleTime="2016-10-17T02:00:00-07:00",
```

```
bbuLearnSchedule="MONTH 1 DATE 17 HOUR 2 MINUTE 0",
 alertSummaryStartTime="2016-09-21T17:00:00-07:00",
 alertSummaryInterval=weekly,
 hardDiskScrubInterval=biweekly, hardDiskScrubFollowupIntervalInDays="14"

ALTER IORMPLAN objective=basic
```

### Example 1-9 Rescue Plan for a Database Server

The `rescuePlan` attribute for a database server could look like this:

```
$ dbmcli -e list dserver attributes rescuePlan

CREATE ROLE "listdserverattrs"

GRANT PRIVILEGE list ON dserver ATTRIBUTES bbuStatus, coreCount WITH all
options TO ROLE "listdserverattrs"

ALTER DBSERVER diagHistoryDays="7", metricHistoryDays="7",
bbuLearnSchedule="MONTH 1 DATE 17 HOUR 2 MINUTE 0",
 alertSummaryStartTime="2016-09-26T08:00:00-07:00",
 alertSummaryInterval=weekly, pendingCoreCount="128" force
```

### Example 1-10 Creating a Rescue Plan script for a cell

The following command stores the commands in the `rescuePlan` attribute to a file called `rescue_cli` located on a remote server.

```
$ cellcli -e list cell attributes rescuePlan >& /location/on/remote/server/
rescue_cell.cli
```

If you need to rescue the server, you can run the script after the server rescue to restore the settings. The following command runs the `rescue_cell.cli` file using the CellCLI `start` command:

```
$ cellcli -e start /location/on/remote/server/rescue_cell.cli
```

### Example 1-11 Creating a Rescue Plan script for a database server

The following command stores the commands in the `rescuePlan` attribute to a file called `rescue_db.cli` located on a remote server.

```
$ dbmcli -e list dserver attributes rescuePlan >& /location/on/remote/
server/rescue_db.cli
```

If you need to rescue the server, you can run the script after the server rescue to restore the settings. The following command runs the `rescue_cell.cli` file using the CellCLI `start` command:

```
$ dbmcli -e start /location/on/remote/server/rescue_db.cli
```

## 1.14 Using ExaWatcher Charts to Monitor Exadata Database Machine Performance

ExaWatcher is a utility that collects performance data on the storage servers and database servers of an Exadata system. The data collected includes operating system statistics, such as `iostat`, cell statistics (`cellsrvstat`), and network statistics.

- [About ExaWatcher Charts](#)  
ExaWatcher collects and presents performance data on the storage servers and database servers of Exadata Database Machine for a specified period of time.
- [Requirements for Using ExaWatcher Charts](#)  
To view the HTML pages, the generated archive file must be moved to a machine with a local browser that has access to the internet.
- [IO Charts](#)  
IO charts show IO performance for an entire server or for individual disks in the storage server.
- [CPU Charts](#)  
The CPU charts show CPU utilization for the server. These statistics are from `iostat` (avg-cpu: %user, %system, %iowait).
- [CPU Detail](#)  
The CPU detail charts show detailed information for CPU usage, including the average CPU utilization per CPU ID. These statistics are from `mpstat`.
- [Cell Server Charts](#)  
Cell server statistics are useful for tracking features that are specific to Exadata storage servers. This page displays statistics related to Smart Flash Cache and Smart IOs.
- [Alert History](#)

### 1.14.1 About ExaWatcher Charts

ExaWatcher collects and presents performance data on the storage servers and database servers of Exadata Database Machine for a specified period of time.

To extract the data collected by ExaWatcher, run `GetExaWatcherResults.sh` and specify the start and end time of the desired time range. The results are then placed in a compressed archive file in a directory called `ExtractedResults`.

For example:

```
$ GetExaWatcherResults.sh --from 08/24/2016_17:00:00 --to
08/25/2016_17:00:00
```

 **Note:**

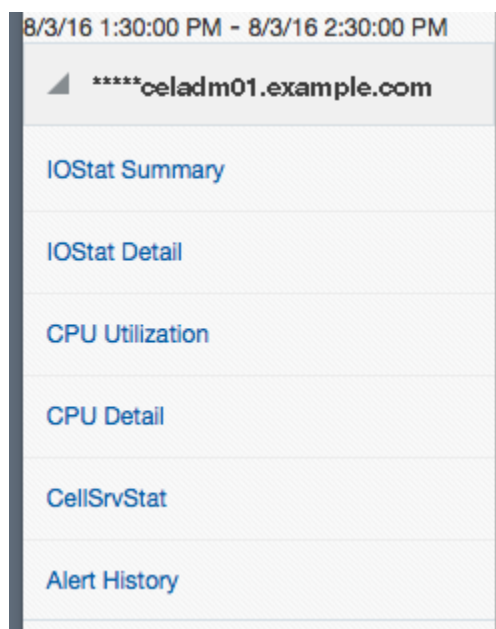
You can use the `-c` or `--scp` options with `GetExaWatcherResults.sh` to copy the resulting archive file to a different location.

In Oracle Exadata System Software release 12.2.1.1.0, `GetExaWatcherResults.sh` also generates HTML pages that contain charts for IO, CPU utilization, cell server statistics, and alert history. The IO and CPU utilization charts use data from `iostat`, CPU detail uses data from `mpstat`, and cell server statistics use data from `cellsrvstat`. Alert history will be retrieved for the specified time frame.

You can find the new charts in the resulting archive file. In the archive file, there is a subdirectory named: `Charts.ExaWatcher.<hostname>/<timestamp>_<duration>/`, for example, `Charts.ExaWatcher.xxxxxceladm13.oracle.com/2016_08_24_17_00_00_01h00m00s_0.`

To view the HTML pages, the archive file needs to be moved to a machine with a local browser that has access to the internet. The file needs to be uncompressed from a bz2 compressed file, then untar with `tar -xvf`. Then you can open `Charts.ExaWatcher.<hostname>/<timestamp>_<duration>/index.html` in a browser. The left panel on that page shows the following menu:

**Figure 1-1 ExaWatcher Menu in the Left Panel**



 **Note:**

For screen reader users, the menu items are navigated using the UP/DOWN arrow keys and activated using the SPACE bar. The TAB key will move you to the frame on the right side.

The CellSrvStat menu item is available only when run against a storage server. The Alert History menu item is available only if there were alerts during the requested time frame.

## 1.14.2 Requirements for Using ExaWatcher Charts

To view the HTML pages, the generated archive file must be moved to a machine with a local browser that has access to the internet.

Due to the complexity of the ExaWatcher charts, if the Oracle Exadata Rack resides in a restricted environment, and the generated HTML files or archive file cannot be moved to an environment that has access to the internet, then you will not be able to view the ExaWatcher charts.

## 1.14.3 IO Charts

IO charts show IO performance for an entire server or for individual disks in the storage server.

The following pages are available for IO statistics:

- [IO Stat Summary](#)
- [I/O Stat Detail](#)

### 1.14.3.1 IO Stat Summary

IOStat Summary shows a summary of IO performance for the entire server. The four charts shown in this page are:

**Table 1-4 Statistics for IOStat Summary**

| Statistic                    | Description                                                                                                                                                                 |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Flash IOPs<br>Hard Disk IOPs | Total reads per second, writes per second, and IO per second (reads per second + writes per second) for the server.<br>This uses r/s and w/s from <code>iostat</code> .     |
| Flash MB/s<br>Hard Disk MB/s | Total read MB per second, write MB per second, and IO MB per second.<br>This uses <code>rsec/s</code> and <code>wsec/s</code> from <code>iostat</code> , converted into MB. |

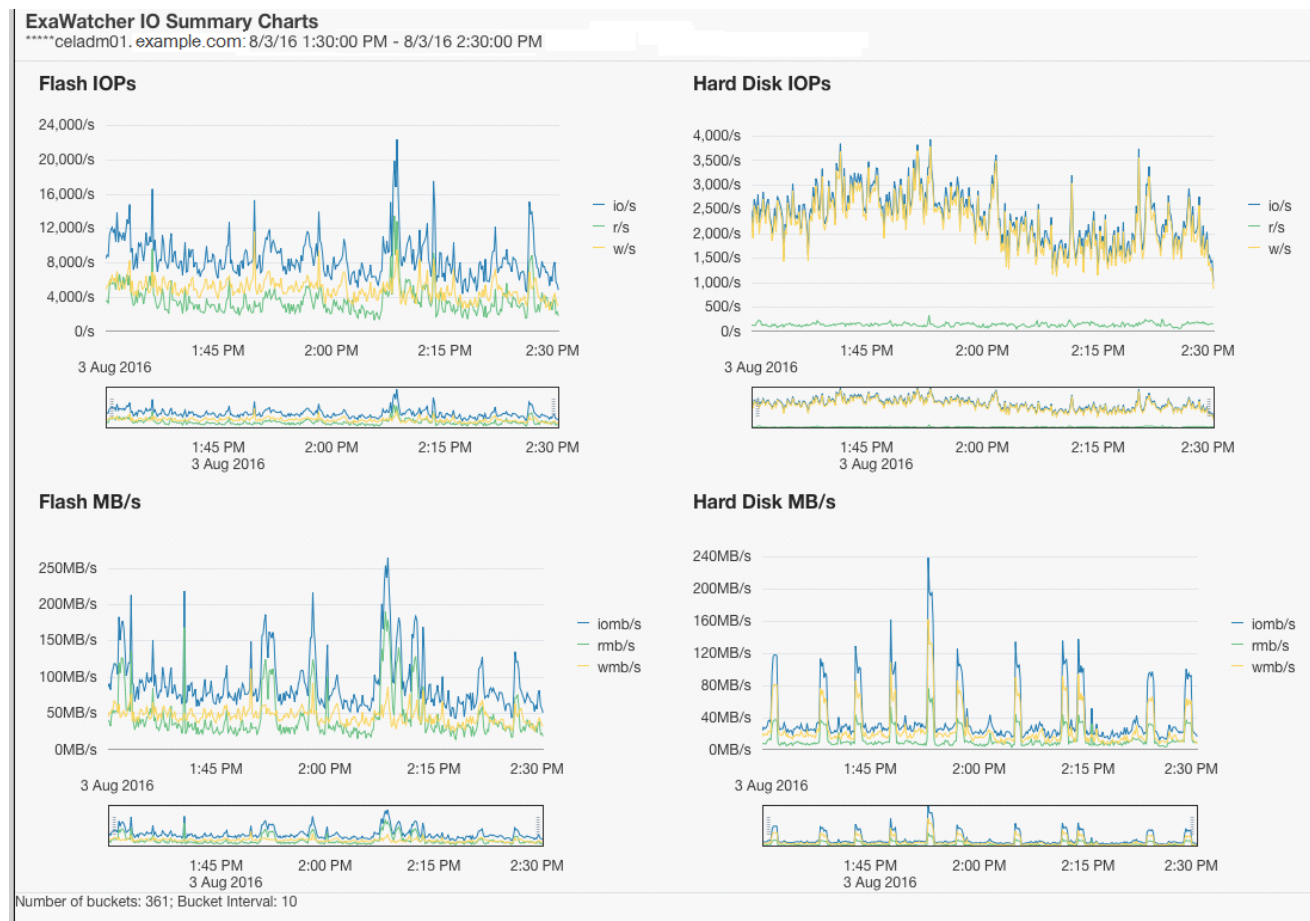
The statistics are shown for flash and hard disks, when applicable. On Exadata Extreme Flash, there are no hard disks. On database servers, there are no flash devices.

If there is a suspected I/O performance problem, the IOPs and the MB/s statistics for the storage servers can be compared to the data sheet to determine if the storage is at maximum capacity. High read times observed on the database can also be correlated to the service time and average wait time from `iostat`, to determine if the high times could potentially be due to the storage server. Note that the database times would typically include IOs that are satisfied from flash cache, as well as hard disk. In addition, these charts enable you to visualize any peaks during the time frame.

The partial screenshot below shows the IOPs and MB/s charts for flash and hard disk



Figure 1-2 IO Summary Charts



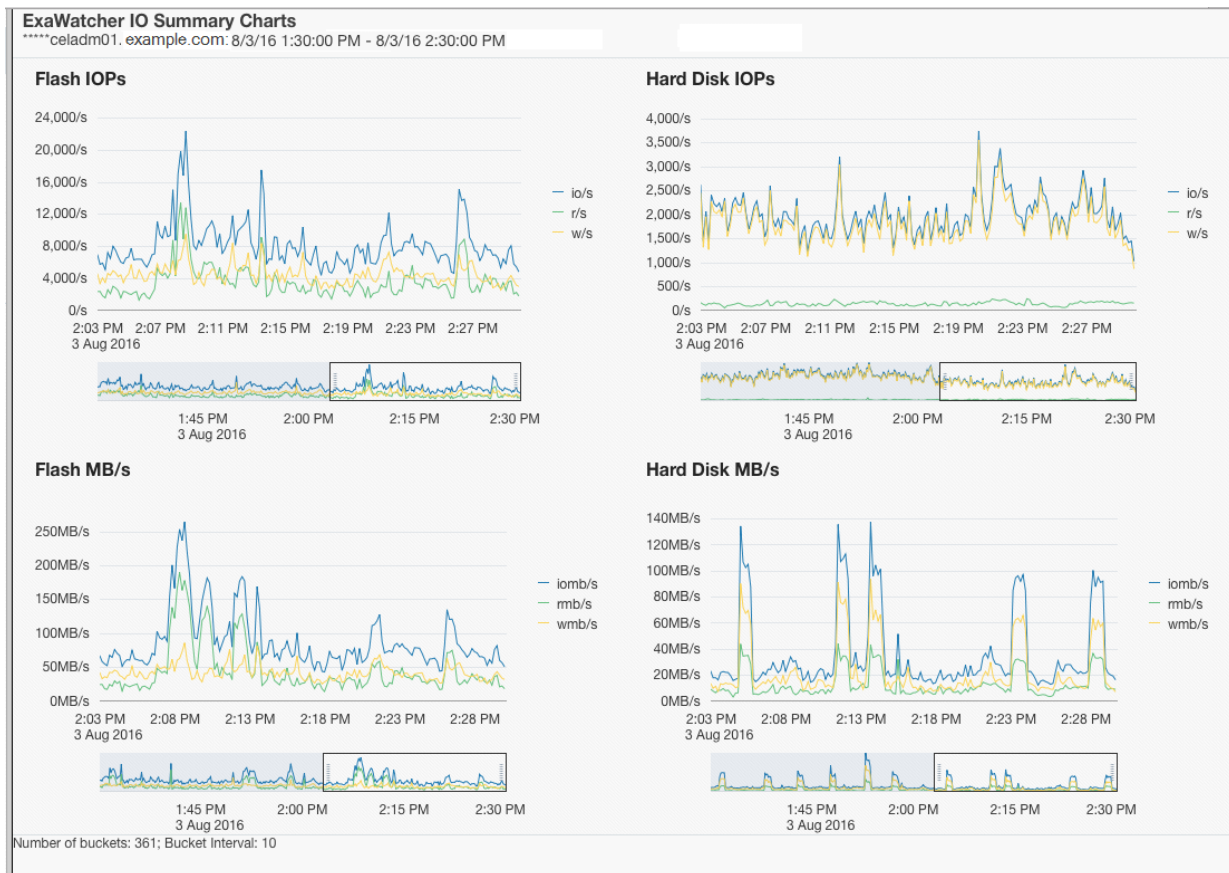
Below each chart, there is a range selector that you can use to drill down to a specific time within the chart. Moving the range selector on any chart affects all charts on the page.



**Note:**

The range selector is not accessible to screen readers. Also, not all values presented in the chart are accessible to a screen reader. Only the first value of each chart data point is.

Figure 1-3 IO Summary Charts Showing Range Selector



When you use the range selector, the displayed chart changes to show only the data for the time range specified by the range selector.

### 1.14.3.2 I/O Stat Detail

I/O Stat Detail shows performance for each disk on the storage server. The following charts are shown in this page:

**Table 1-5 Statistics for IO Stat Detail**

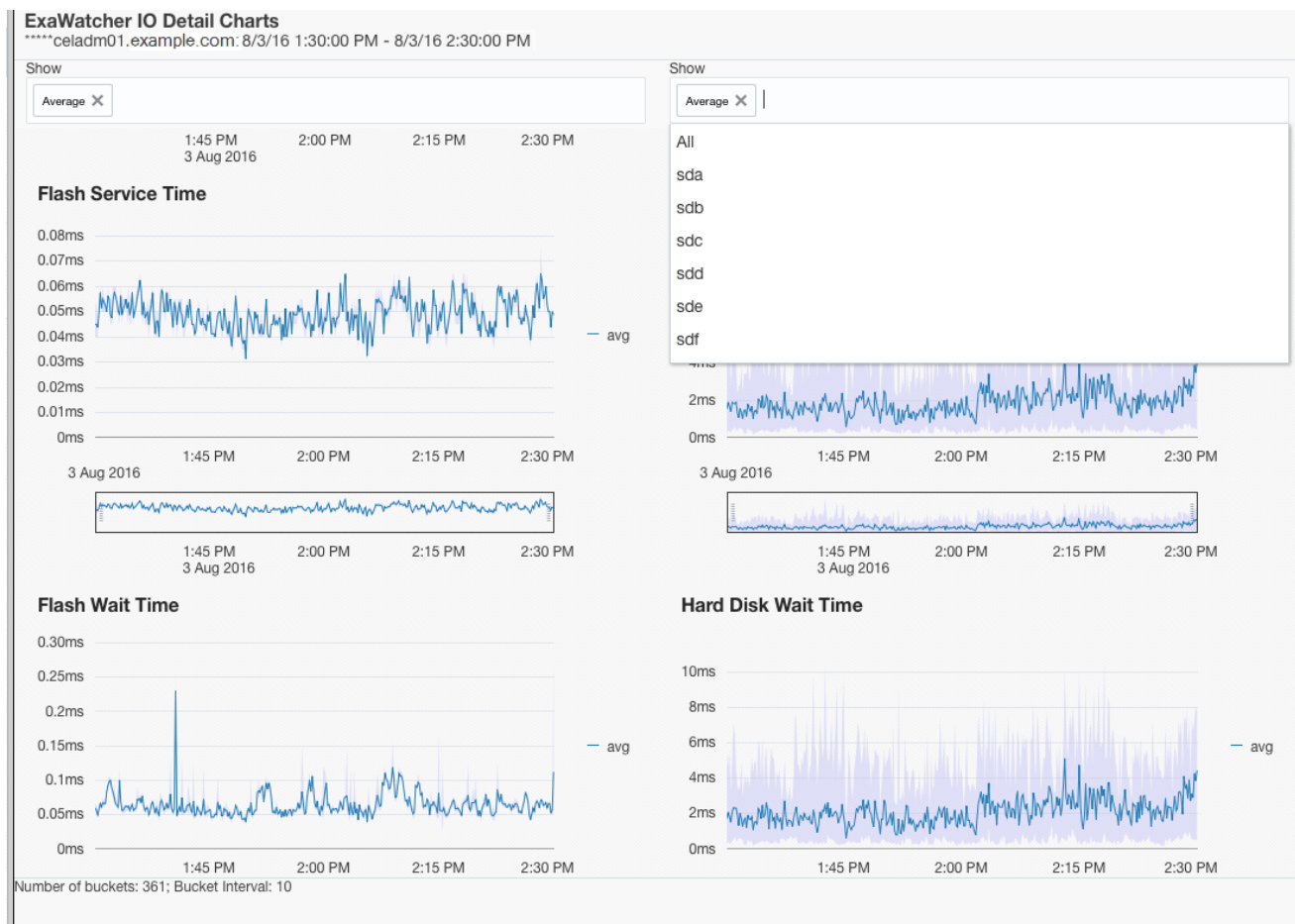
| Statistic                                    | Description                                                               |
|----------------------------------------------|---------------------------------------------------------------------------|
| Flash Service Time<br>Hard Disk Service Time | Average service time per disk contrasted against the range of wait times. |
| Flash Wait Time<br>Hard Disk Wait Time       | Average wait time per disk                                                |

By default, the charts include a line that depicts the average across all disks on the server. The shaded, background image indicates the minimum and maximum range for the statistic. You can choose to display individual disks by using the drop down selector.

If the background image has a wide range, then this can indicate possible differences in disk performance. You can use this metric to look more closely at each individual disk on the storage server to see if there is an imbalance. If the background image has a narrow range, then that indicates the disks are performing similarly.

The individual disk IOPs and MB/s for a storage server can also be compared to the data sheet numbers to see if the disks are potentially hitting their maximum capacity.

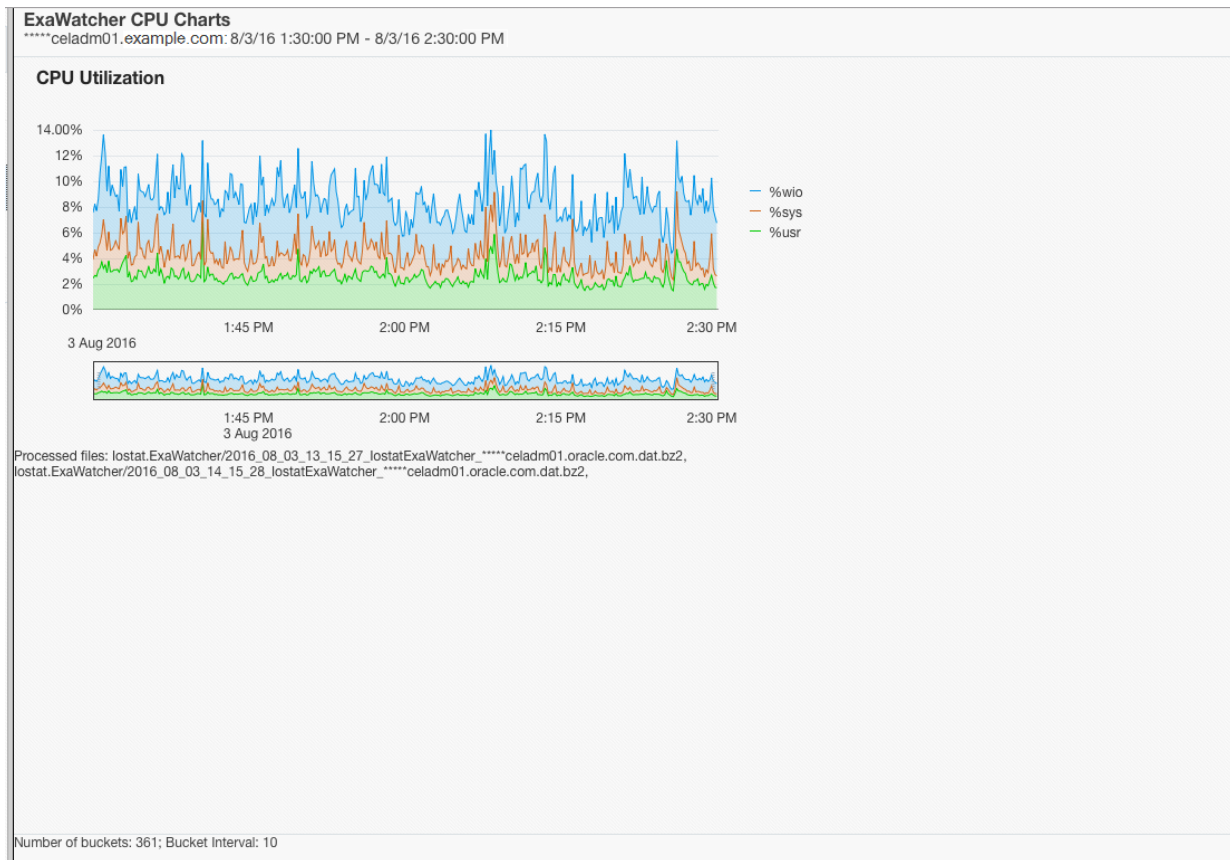
Figure 1-4 IO Detail Charts



### 1.14.4 CPU Charts

The CPU charts show CPU utilization for the server. These statistics are from `iostat` (avg-cpu: %user, %system, %iowait).

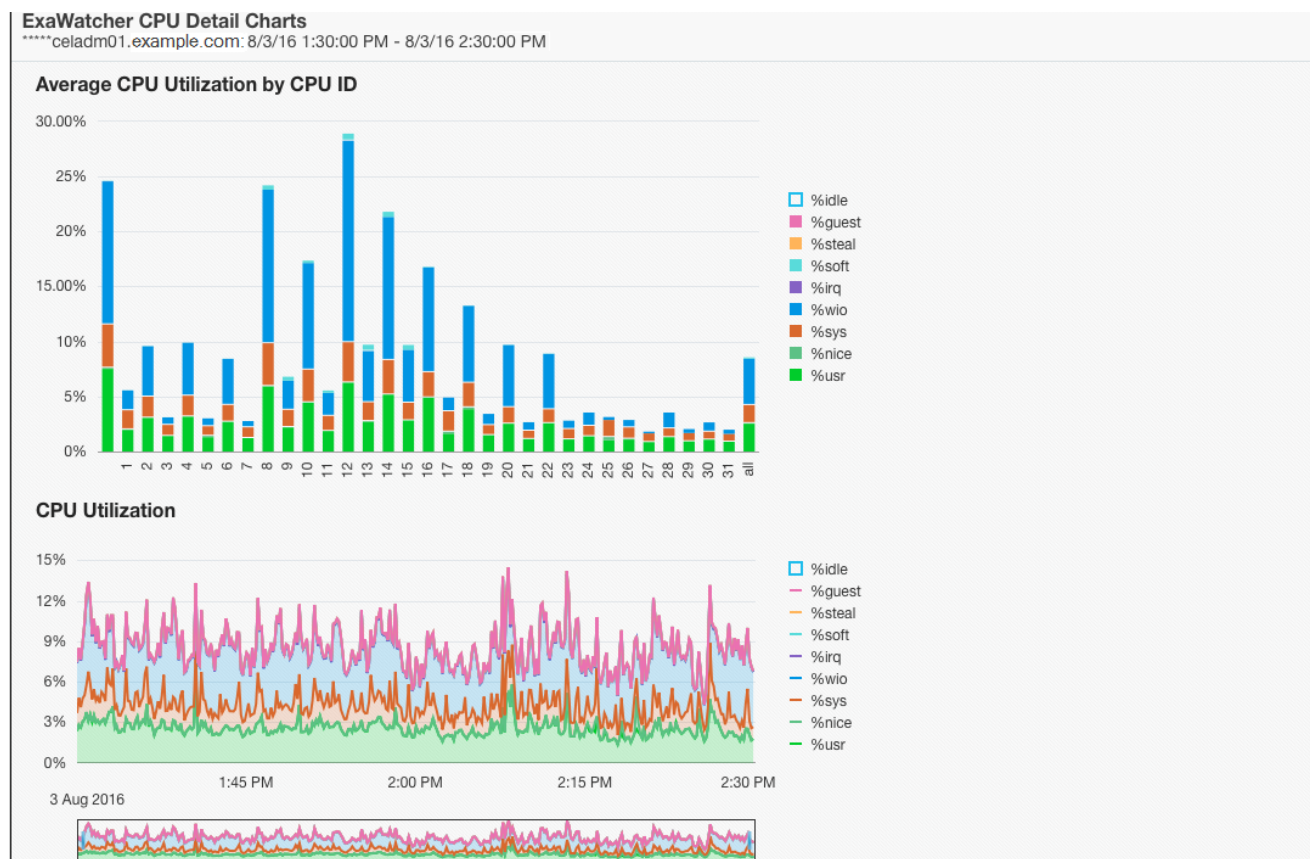
Figure 1-5 CPU Charts



### 1.14.5 CPU Detail

The CPU detail charts show detailed information for CPU usage, including the average CPU utilization per CPU ID. These statistics are from `mpstat`.

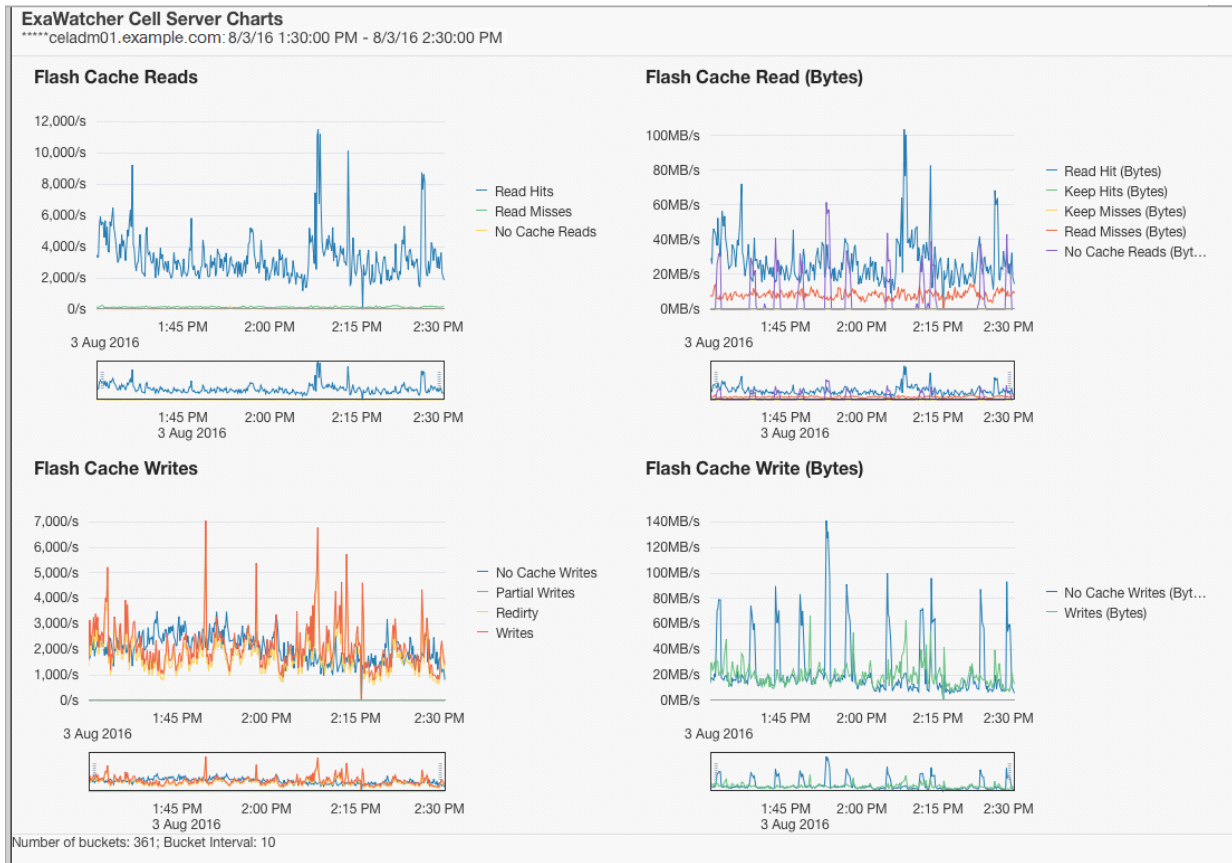
Figure 1-6 CPU Detail Charts



### 1.14.6 Cell Server Charts

Cell server statistics are useful for tracking features that are specific to Exadata storage servers. This page displays statistics related to Smart Flash Cache and Smart IOs.

Figure 1-7 Cell Server Charts

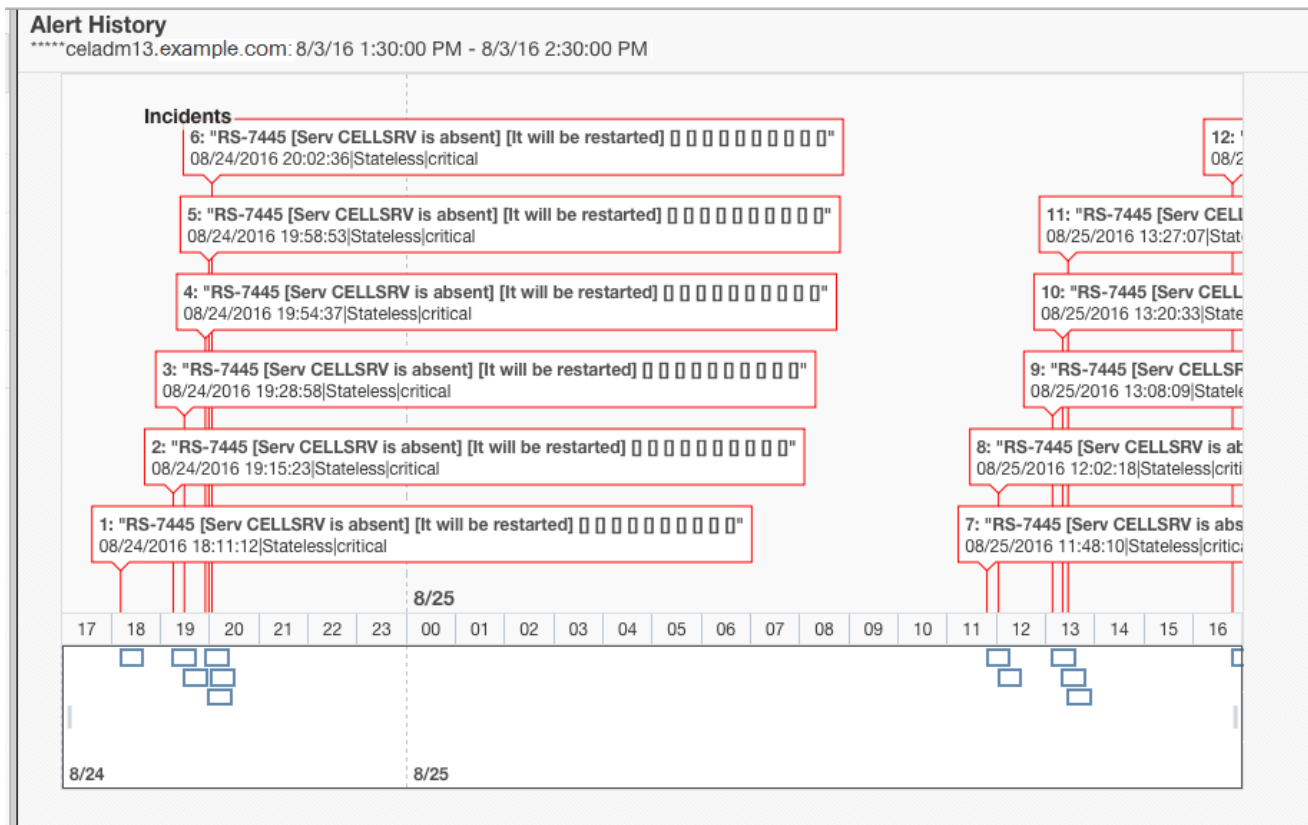


### 1.14.7 Alert History

This page displays alerts that were present during the specified time frame. Alerts may be raised from errors or issues, which may result in IO performance issues on the servers.



Figure 1-8 Alert History



# 2

## Maintaining Exadata Database Servers



### Note:

For ease of reading, the name "Oracle Exadata Rack" is used when information refers to both Exadata Database Machine and Oracle Exadata Storage Expansion Rack.

- [Management Server on Database Servers](#)
- [Maintaining the Local Storage on Exadata Database Servers](#)  
Repair of the local drives does not require an Oracle Exadata Database Machine database server to be shut down.
- [Maintaining Flash Disks on Exadata Database Servers](#)  
Flash disks should be monitored and replaced when necessary.
- [Adding the Disk Expansion Kit to Database Servers](#)  
You can add local storage space to an Oracle Exadata Database Server by using a disk expansion kit.
- [Adding Memory Expansion Kit to Database Servers](#)
- [Verifying and Modifying the Link Speed on the Client Network Ports for X7 and Later Systems](#)  
You can configure 10 GbE connections or 25 GbE connections on the client network on Exadata Database Machine X7 and later database servers.
- [Adding and Configuring an Extra Network Card on Exadata Database Machine X6-2 and Later](#)  
You can add an additional network card on Exadata Database Machine X6-2 and later systems.
- [Increasing the Number of Active Cores on Database Servers](#)  
You can increase the number of active cores on Exadata Database Machine using capacity-on-demand.
- [Extending LVM Partitions](#)  
Logical Volume Manager (LVM) provides flexibility to reorganize the partitions in the database servers.
- [Creating a Snapshot-Based Backup of Oracle Linux Database Server](#)
- [Recovering Oracle Linux Database Servers Using a Snapshot-Based Backup](#)  
You can recover a database server file systems running Oracle Linux using a snapshot-based backup after severe disaster conditions happen for the database server, or when the server hardware is replaced to such an extent that it amounts to new hardware.
- [Re-Imaging the Oracle Exadata Database Server](#)  
The re-image procedure is necessary when a database server needs to be brought to an initial state for various reasons.



- [Changing Existing Elastic Configurations for Database Servers](#)  
Elastic configurations provide a flexible and efficient mechanism to change the server configuration of your Exadata Database Machine.
- [Managing Quorum Disks for High Redundancy Disk Groups](#)  
Quorum disks allow for the Oracle RAC voting files to be stored in a high redundancy disk group on an Oracle Exadata Rack with fewer than five storage servers due to the presence of two extra failure groups.
- [Using vmetrics](#)  
The vmetrics package enables you to display system statistics gathered by the vmetrics service.
- [Using FIPS mode](#)  
On database servers running Oracle Linux 7, you can enable the kernel to run in FIPS mode.
- [LED Status Descriptions](#)  
The LEDs on the Oracle Exadata Rack components help you identify the component that needs servicing.
- [Exadata Database Server Images](#)  
The Exadata database server models have different external layouts and physical appearance.

## 2.1 Management Server on Database Servers

Management Server (MS) running on database servers provides monitoring, alerting, and other administrative capabilities. It also provides the DBMCLI command-line administration tool.



### See Also:

- "Management Server on Database Servers" in the *Oracle Exadata Database Machine System Overview* guide
- [Using the DBMCLI Utility](#)

## 2.2 Maintaining the Local Storage on Exadata Database Servers

Repair of the local drives does not require an Oracle Exadata Database Machine database server to be shut down.

No downtime of the rack is required, however the individual server may require downtime, temporarily taking it out of the cluster.

- [Verifying the Database Server Configuration](#)  
Oracle recommends verifying the status of the database server RAID devices to avoid possible performance impact, or an outage.
- [Monitoring a Database Server RAID Set Rebuilding](#)

- [Reclaiming a Hot Spare Drive After Upgrading to Oracle Exadata System Software Release 12.1.2.1.0 or Later](#)
- [Understanding Automated File Deletion Policy](#)

## 2.2.1 Verifying the Database Server Configuration

Oracle recommends verifying the status of the database server RAID devices to avoid possible performance impact, or an outage.

The impact of validating the RAID devices is minimal. The impact of corrective actions will vary depending on the specific issue uncovered, and may range from simple reconfiguration to an outage.

- [About the RAID Storage Configuration](#)  
The local storage drives are configured in a RAID configuration.
- [Verifying Disk Controller Configuration on Oracle Exadata Database Machine X7-8 or Later Systems](#)
- [Verifying Disk Controller Configuration on Oracle Exadata Database Machine X6-8 and Earlier](#)
- [Verifying Virtual Drive Configuration](#)
- [Verifying Physical Drive Configuration](#)  
Check your system for critical, degraded, or failed disks.

### 2.2.1.1 About the RAID Storage Configuration

The local storage drives are configured in a RAID configuration.

**Table 2-1 Disk Configurations for Exadata Database Machine Two-Socket Systems**

| Server Type                           | RAID Controller       | Disk Configuration                                              |
|---------------------------------------|-----------------------|-----------------------------------------------------------------|
| Oracle Exadata Database Machine X9M-2 | N/A                   | Two mirrored (RAID-1) NVMe flash drives in each database server |
| Oracle Exadata Database Machine X8M-2 | MegaRAID SAS 9361-16i | 4 disk drives in a RAID-5 configuration on each database server |
| Oracle Exadata Database Machine X8-2  | MegaRAID SAS 9361-16i | 4 disk drives in a RAID-5 configuration on each database server |
| Oracle Exadata Database Machine X7-2  | MegaRAID SAS 9361-16i | 4 disk drives in a RAID-5 configuration on each database server |
| Oracle Exadata Database Machine X6-2  | MegaRAID SAS 9361-8i  | 4 disk drives in a RAID-5 configuration on each database server |
| Oracle Exadata Database Machine X5-2  | MegaRAID SAS 9361-8i  | 4 disk drives in a RAID-5 configuration on each database server |
| Oracle Exadata Database Machine X4-2  | MegaRAID SAS 9261-8i  | 4 disk drives in a RAID-5 configuration on each database server |

**Table 2-1 (Cont.) Disk Configurations for Exadata Database Machine Two-Socket Systems**

| Server Type                          | RAID Controller      | Disk Configuration                                              |
|--------------------------------------|----------------------|-----------------------------------------------------------------|
| Oracle Exadata Database Machine X3-2 | MegaRAID SAS 9261-8i | 4 disk drives in a RAID-5 configuration on each database server |
| Oracle Exadata Database Machine X2-2 | MegaRAID SAS 9261-8i | 4 disk drives in a RAID-5 configuration on each database server |

**Table 2-2 Disk Configurations for Exadata Database Machine Eight-Socket Systems**

| Server Type                           | RAID Controller      | Disk Configuration                                                                                               |
|---------------------------------------|----------------------|------------------------------------------------------------------------------------------------------------------|
| Oracle Exadata Database Machine X9M-8 | N/A                  | Two mirrored (RAID-1) NVMe flash drives in each database server                                                  |
| Oracle Exadata Database Machine X8M-8 | N/A                  | Two mirrored (RAID-1) NVMe flash drives in each database server                                                  |
| Oracle Exadata Database Machine X8-8  | N/A                  | Two mirrored (RAID-1) NVMe flash drives in each database server                                                  |
| Oracle Exadata Database Machine X7-8  | N/A                  | Two mirrored (RAID-1) NVMe flash drives in each database server                                                  |
| Oracle Exadata Database Machine X5-8  | MegaRAID SAS 9361-8i | 8 disk drives in each database server with one virtual drive created across the RAID-5 set                       |
| Oracle Exadata Database Machine X4-8  | MegaRAID SAS 9261-8i | 7 disk drives in each database server configured as one 6-disk RAID-5 with one global hot spare drive by default |
| Oracle Exadata Database Machine X3-8  | MegaRAID SAS 9261-8i | 8 disk drives in each database server with one virtual drive created across the RAID-5 set                       |

### 2.2.1.2 Verifying Disk Controller Configuration on Oracle Exadata Database Machine X7-8 or Later Systems

- Query `mdstat` to view the database server disk controller configuration on Oracle Exadata Database Machine X7-8 or later systems.

```
[root@dbnode01adm01 ~]# cat /proc/mdstat
Personalities : [raid1]
md34 : active raid1 nvme3n1[1] nvme1n1[0]
 3125613568 blocks super external:/md126/0 [2/2] [UU]

md24 : active raid1 nvme2n1[1] nvme0n1[0]
```

```

262144000 blocks super external:/md127/0 [2/2] [UU]

md25 : active raid1 nvme2n1[1] nvme0n1[0]
 2863467520 blocks super external:/md127/1 [2/2] [UU]

md126 : inactive nvme3n1[1](S) nvme1n1[0](S)
 6306 blocks super external:imsm

md127 : inactive nvme2n1[1](S) nvme0n1[0](S)
 6306 blocks super external:imsm

unused devices: <none>

```

If the output you see is different, then investigate and correct the problem. Degraded virtual drives usually indicate absent or failed physical disks. Disks that show [1/2] and [U\_] or [\_U] for the state indicate that one of the NVME disks is down. Failed disks should be replaced quickly.

### 2.2.1.3 Verifying Disk Controller Configuration on Oracle Exadata Database Machine X6-8 and Earlier

For Oracle Exadata Database Machine X4-2, Oracle Exadata Database Machine X3-2, and Oracle Exadata Database Machine X2-2, the expected output is one virtual drive, none degraded or offline, five physical devices (one controller and four disks), four disks, and no critical or failed disks.

For Oracle Exadata Database Machine X3-8 Full Rack and Oracle Exadata Database Machine X2-8 Full Rack, the expected output is one virtual drive, none degraded or offline, 11 physical devices (one controller, two SAS2 expansion ports, and eight disks), eight disks, and no critical or failed disks.

If your output is different, then investigate and correct the problem. Degraded virtual drives usually indicate absent or failed physical disks. Critical disks should be replaced immediately to avoid the risk of data loss if the number of failed disks in the node exceed the count needed to sustain the operations of the system. Failed disks should also be replaced quickly.

#### Note:

If additional virtual drives or a hot spare are present, then it may be that the procedure to reclaim disks was not performed at deployment time or that a bare metal restore procedure was performed without using the `dualboot=no` qualifier. Contact Oracle Support Services and reference My Oracle Support note 1323309.1 for additional information and corrective steps.

When upgrading a database server that has a hot spare to Oracle Exadata System Software release 11.2.3.2.0 or later, the hot spare is removed, and added as an active drive to the RAID configuration. The database servers have the same availability in terms of RAID5 redundancy, and can survive the loss of one drive. When a drive failure happens, Oracle Auto Service Request (ASR) sends out a notification to replace the faulty drive at the earliest opportunity.

- Use the following command to verify the database server disk controller configuration on all systems prior to Oracle Exadata Database Machine X7-8:

```
if [[-d /proc/xen && ! -f /proc/xen/capabilities]]
then
 echo -e "\nThis check will not run in a user domain of a
virtualized environment. Execute this check in the management
domain.\n"
else
 if [-x /opt/MegaRAID/storcli/storcli64]
 then
 export CMD=/opt/MegaRAID/storcli/storcli64
 else
 export CMD=/opt/MegaRAID/MegaCli/MegaCli64
 fi
 RAW_OUTPUT=$($CMD AdpAllInfo -aALL -nolog | grep "Device Present"
-A 8);
 echo -e "The database server disk controller configuration found
is:\n\n$RAW_OUTPUT";
fi;
```

 **Note:**

This check is not applicable to Oracle Exadata Database Machine X7-8 or later database servers because they do not have any conventional disk drives.

**Example 2-1 Checking the disk controller configuration for Oracle Exadata Database Machine 2-socket system (X2-2 or later) without the disk expansion kit**

The following is an example of the output from the command for Oracle Exadata Database Machine 2-socket system (X2-2 or later) without the disk expansion kit.

```

 Device Present
 =====
Virtual Drives : 1
 Degraded : 0
 Offline : 0
Physical Devices : 5
 Disks : 4
 Critical Disks : 0
 Failed Disks : 0
```

**Example 2-2 Checking the disk controller configuration for Oracle Exadata Database Machine X4-8 Full Rack**

The following is an example of the output from the command for Oracle Exadata Database Machine X4-8 Full Rack:

```

 Device Present
 =====
Virtual Drives : 1
 Degraded : 0
```

```

Offline : 0
Physical Devices : 8
Disks : 7
Critical Disks : 0
Failed Disks : 0

```

### Example 2-3 Checking the disk controller configuration for Oracle Exadata Database Machine X5-8 or X6-8 Full Rack

The following is an example of the output from the command for Oracle Exadata Database Machine X5-8 or X6-8 Full Rack:

```

 Device Present
 =====
Virtual Drives : 1
 Degraded : 0
 Offline : 0
Physical Devices : 9
 Disks : 8
 Critical Disks : 0
 Failed Disks : 0

```

## 2.2.1.4 Verifying Virtual Drive Configuration

To verify the virtual drive configuration, use the following command to verify the virtual drive configuration:



### Note:

If you are running Oracle Exadata System Software 19.1.0 or later, substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands:

```

/opt/MegaRAID/MegaCli/MegaCli64 CfgDsply -aALL | grep "Virtual Drive:"; \
/opt/MegaRAID/MegaCli/MegaCli64 CfgDsply -aALL | grep "Number Of Drives"; \
/opt/MegaRAID/MegaCli/MegaCli64 CfgDsply -aALL | grep "^State"

```

The following is an example of the output for Oracle Exadata Database Machine X4-2, Oracle Exadata Database Machine X3-2 and Oracle Exadata Database Machine X2-2. The virtual device 0 should have four drives, and the state is `Optimal`.

```

Virtual Drive : 0 (Target Id: 0)
Number Of Drives : 4
State : Optimal

```

The expected output for Oracle Exadata Database Machine X3-8 Full Rack and Oracle Exadata Database Machine X2-8 Full Rack displays the virtual device has eight drives and a state of `optimal`.

 **Note:**

If a disk replacement was performed on a database server without using the `dualboot=no` option, then the database server may have three virtual devices. Contact Oracle Support and reference My Oracle Support note 1323309.1 for additional information and corrective steps.

## 2.2.1.5 Verifying Physical Drive Configuration

Check your system for critical, degraded, or failed disks.

To verify physical drive configuration, use the following command to verify the database server physical drive configuration:

 **Note:**

If you are running Oracle Exadata System Software 19.1.0 or later, substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands:

```
/opt/MegaRAID/MegaCli/MegaCli64 -PDList -aALL | grep "Firmware state"
```

The following is an example of the output for Oracle Exadata Database Machine X4-2, Oracle Exadata Database Machine X3-2, and Oracle Exadata Database Machine X2-2:

```
Firmware state: Online, Spun Up
Firmware state: Online, Spun Up
Firmware state: Online, Spun Up
Firmware state: Online, Spun Up
```

The drives should show a state of `Online, Spun Up`. The order of the output is not important. The output for Oracle Exadata Database Machine X3-8 Full Rack or Oracle Exadata Database Machine X2-8 Full Rack should be eight lines of output showing a state of `Online, Spun Up`.

If your output is different, then investigate and correct the problem.

Degraded virtual drives usually indicate absent or failed physical disks. Critical disks should be replaced immediately to avoid the risk of data loss if the number of failed disks in the node exceed the count needed to sustain the operations of the system. Failed disks should be replaced quickly.

## 2.2.2 Monitoring a Database Server RAID Set Rebuilding

If a drive in a database server RAID set is replaced, then the progress of the RAID set rebuild should be monitored.

Use the following command on the database server that has the replaced disk. The command is run as the `root` user.

 **Note:**

If you are running Oracle Exadata System Software 19.1.0 or later, substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands:

```
/opt/MegaRAID/MegaCli/MegaCli64 -pdrbld -showprog -physdrv \
[disk_enclosure:slot_number] -a0
```

In the preceding command, `disk_enclosure` and `slot_number` indicate the replacement disk identified by the `MegaCli64 -PDList` command. The following is an example of the output from the command:

```
Rebuild Progress on Device at Enclosure 252, Slot 2 Completed 41% in 13 Minutes.
```

## 2.2.3 Reclaiming a Hot Spare Drive After Upgrading to Oracle Exadata System Software Release 12.1.2.1.0 or Later

Oracle Exadata Database Machines upgraded to Oracle Exadata System Software release 12.1.2.1.0 or later that have a hot spare drive cannot use the `reclaimdisks.sh` script to reclaim the drive. The following procedure describes how to manually reclaim the drive:

 **Note:**

During the procedure, the database server is restarted twice. The steps in the procedure assume that the Oracle Grid Infrastructure restart is disabled after the server restart.

The sample output shows Oracle Exadata Database Machine X2-2 database server with four disks. The enclosure identifier, slot number, and such may be different for your system.

 **Note:**

If you are running Oracle Exadata System Software 19.1.0 or later, substitute the string `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands:

1. Identify the hot spare drive.

```
/opt/MegaRAID/MegaCli/MegaCli64 -PDList -aALL
```



The following is an example of the output from the command for the hot spare drive:

```
...
Enclosure Device ID: 252
Slot Number: 3
Enclosure position: N/A
Device Id: 8
WWN: 5000CCA00A9FAA5F
Sequence Number: 2
Media Error Count: 0
Other Error Count: 0
Predictive Failure Count: 0
Last Predictive Failure Event Seq Number: 0
PD Type: SAS
Hotspare Information:
Type: Global, with enclosure affinity, is revertible

Raw Size: 279.396 GB [0x22ecb25c Sectors]
Non Coerced Size: 278.896 GB [0x22dcb25c Sectors]
Coerced Size: 278.464 GB [0x22cee000 Sectors]
Sector Size: 0
Logical Sector Size: 0
Physical Sector Size: 0
Firmware state: Hotspare, Spun down
Device Firmware Level: A2A8
Shield Counter: 0
Successful diagnostics completion on : N/A
...
```

The command identified the hot spare drive on enclosure identifier 252, slot 3.

**2. Obtain the virtual drive information.**

```
/opt/MegaRAID/MegaCli/MegaCli64 -LDInfo -Lall -Aall
```

The following is an example of the output from the command:

```
Adapter 0 -- Virtual Drive Information:
Virtual Drive: 0 (Target Id: 0)
Name :DBSYS
RAID Level : Primary-5, Secondary-0, RAID Level Qualifier-3
Size : 556.929 GB
Sector Size : 512
Is VD emulated : No
Parity Size : 278.464 GB
State : Optimal
Strip Size : 1.0 MB
Number Of Drives : 3
Span Depth : 1
Default Cache Policy: WriteBack, ReadAheadNone, Direct, No Write
Cache if Bad BBU
Current Cache Policy: WriteBack, ReadAheadNone, Direct, No Write
Cache if Bad BBU
```

```
Default Access Policy: Read/Write
Current Access Policy: Read/Write
Disk Cache Policy : Disabled
Encryption Type : None
Is VD Cached: No
```

The command identified a RAID 5 configuration for virtual drive 0 on adapter 0.

**3. Remove the hot spare drive.**

```
/opt/MegaRAID/MegaCli/MegaCli64 -PDHSP -Rmv -PhysDrv[252:3] -a0
```

**4. Add the drive as an active RAID 5 drive.**

```
/opt/MegaRAID/MegaCli/MegaCli64 -LDRecon -Start -r5 \
 -Add -PhysDrv[252:3] -L0 -a0
```

```
Start Reconstruction of Virtual Drive Success.
Exit Code: 0x00
```

 **Note:**

If the message Failed to Start Reconstruction of Virtual Drive is displayed, then follow the instructions in My Oracle Support note 1505157.1.

**5. Monitor the progress of the RAID reconstruction.**

```
/opt/MegaRAID/MegaCli/MegaCli64 -LDRecon -ShowProg -L0 -a0
```

```
Reconstruction on VD #0 (target id #0) Completed 1% in 2 Minutes.
```

The following output shows the output of the command after the hot spare drive is added to the RAID 5 configuration, and the reconstruction is finished:

```
Reconstruction on VD #0 is not in Progress.
```

**6. Verify the number of drives.**

```
/opt/MegaRAID/MegaCli/MegaCli64 -LDInfo -Lall -Aall
```

The following is an example of the output from the command:

```
Adapter 0 -- Virtual Drive Information:
Virtual Drive: 0 (Target Id: 0)
Name :DBSYS
RAID Level : Primary-5, Secondary-0, RAID Level Qualifier-3
Size : 835.394 GB
Sector Size : 512
Is VD emulated : No
Parity Size : 278.464 GB
State : Optimal
```

```
Strip Size : 1.0 MB
Number Of Drives : 4
Span Depth : 1
Default Cache Policy: WriteBack, ReadAheadNone, Direct, No Write
Cache if Bad BBU
Current Cache Policy: WriteBack, ReadAheadNone, Direct, No Write
Cache if Bad BBU
Default Access Policy: Read/Write
Current Access Policy: Read/Write
Disk Cache Policy : Disabled
Encryption Type : None
Is VD Cached: No
```

**7. Check the size of the RAID.**

```
parted /dev/sda print

Model: LSI MR9261-8i (scsi)
Disk /dev/sda: 598GB
Sector size (logical/physical): 512B/4096B
Partition Table: msdos

Number Start End Size Type File system Flags
1 32.3kB 132MB 132MB primary ext3 boot
2 132MB 598GB 598GB primary lvm
```

**8. Restart the server in order for the changes to take effect.**

**9. Check the size of the RAID again.**

```
parted /dev/sda print

Model: LSI MR9261-8i (scsi)
Disk /dev/sda: 897GB
Sector size (logical/physical): 512B/4096B
Partition Table: msdos

Number Start End Size Type File system Flags
1 32.3kB 132MB 132MB primary ext3 boot
2 132MB 598GB 598GB primary lvm
```

The increased RAID size allows for extending the volume group. To extend the volume group, you must add an additional partition to the drive.

**10. Obtain the new size, in sectors.**

```
parted /dev/sda

GNU Parted 2.1
Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) unit s
(parted) print
Model: LSI MR9261-8i (scsi)
Disk /dev/sda: 1751949312s
```

```
Sector size (logical/physical): 512B/4096B
Partition Table: msdos
```

```
Number Start End Size Type File system Flags
1 63s 257039s 256977s primary ext3 boot
2 257040s 1167957629s 1167700590s primary lvm
```

In the preceding example, a third partition can be created starting at sector 1167957630, and ending at the end of the disk at sector 1751949311.

#### 11. Create an additional partition on the drive.

```
parted /dev/sda

GNU Parted 2.1
Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) unit s

(parted) mkpart

Partition type? primary/extended? primary
File system type? [ext2]? ext2
Start? 1167957630
End? 1751949311
Warning: The resulting partition is not properly aligned for best
performance.
Ignore/Cancel? Ignore
Warning: WARNING: the kernel failed to re-read the partition table
on /dev/sda (Device or resource busy). As a
result, it may not reflect all of your changes until after reboot.
(parted)

(parted) print
Model: LSI MR9261-8i (scsi)
Disk /dev/sda: 1751949312s
Sector size (logical/physical): 512B/4096B
Partition Table: msdos

Number Start End Size Type File system Flags
1 63s 257039s 256977s primary ext3 boot
2 257040s 1167957629s 1167700590s primary lvm
3 1167957630s 1751949311s 583991682s primary

(parted) set 3 lvm on

Warning: WARNING: the kernel failed to re-read the partition table
on /dev/sda (Device or resource busy). As a
result, it may not reflect all of your changes until after reboot.
(parted) print
Model: LSI MR9261-8i (scsi)
Disk /dev/sda: 1751949312s
Sector size (logical/physical): 512B/4096B
Partition Table: msdos
```

```

Number Start End Size Type File system Flags
1 63s 257039s 256977s primary ext3 boot
2 257040s 1167957629s 1167700590s primary lvm
3 1167957630s 1751949311s 583991682s primary lvm

```

12. Restart the database server.

13. Create the physical volume.

```
pvcreate /dev/partition_name
```

14. Add the physical volume to the existing volume group.

In the following example, substitute the actual names for the *volume\_group*, *partition\_name*, and *volume\_name*.

```
vgextend volume_group /dev/partition_name
```

```
Volume group "volume_name" successfully extended
```

15. Resize the logical volume and file systems as described in "[Extending LVM Partitions](#)."

#### Related Topics

- [Failed to Start Reconstruction of Virtual Drive - Adding a hot-spare manually \(My Oracle Support Doc ID 1505157.1\)](#)

## 2.2.4 Understanding Automated File Deletion Policy

Management Server (MS) includes a file deletion policy for the / (root) directory on the database servers that is triggered when file system utilization is high. Deletion of files is triggered when file utilization is 80 percent, and an alert is sent before the deletion begins. The alert includes the name of the directory, and space usage for the subdirectories. In particular, the deletion policy is as follows:

Files in the following directories are deleted using a policy based on the file modification time stamp.

- /opt/oracle/dbserver/log
- /opt/oracle/dbserver/dbms/deploy/config/metrics
- /opt/oracle/dbserver/dbms/deploy/log

Files that are older than the number of days set by the `metricHistoryDays` attribute are deleted first, then successive deletions occur for earlier files, down to files with modification time stamps older than or equal to 10 minutes, or until file system utilization is less than 75 percent. The `metricHistoryDays` attribute applies to files in /opt/oracle/dbserver/dbms/deploy/config/metrics. For the other log and trace files, use the `diagHistoryDays` attribute.

Starting with Oracle Exadata System Software release 12.1.2.2.0, the maximum amount of space for `ms-odl.trc` and `ms-odl.log` files is 100 MB (twenty 5 MB files) for `*.trc` files and 100 MB (twenty 5 MB files) for `*.log` files. Previously, it was 50 MB (ten 5 MB files) for both `*.trc` and `*.log` files.

`ms-odl` generation files are renamed when they reach 5 MB, and the oldest are deleted when the files use up 100 MB of space.

## 2.3 Maintaining Flash Disks on Exadata Database Servers

Flash disks should be monitored and replaced when necessary.

Starting with Exadata Database Machine X7-8, the database servers contain flash devices instead of hard disks. These flash devices can be replaced without shutting down the server.

- [Monitoring the Status of Flash Disks](#)  
You can monitor the status of a flash disk on the Exadata Database Machine by checking its attributes with the `DBMCLI LIST PHYSICALDISK` command.
- [Performing a Hot-Pluggable Replacement of a Flash Disk](#)  
For Exadata Database Machine X7-8 and X8-8 models, the database server uses hot-pluggable flash disks instead of hard disk drives.

### 2.3.1 Monitoring the Status of Flash Disks

You can monitor the status of a flash disk on the Exadata Database Machine by checking its attributes with the `DBMCLI LIST PHYSICALDISK` command.

For example, a flash disk status equal to `failed` is probably having problems and needs to be replaced.

- Use the `DBMCLI` command `LIST PHYSICALDISK` to determine the status of a flash disk:

```
DBMCLI> LIST PHYSICALDISK WHERE disktype=flashdisk AND status!=normal
DETAIL
name: FLASH_1_1
deviceName: /dev/nvme0n1
diskType: FlashDisk
luns: 1_1
makeModel: "Oracle Flash Accelerator F640 PCIe Card"
physicalFirmware: QDV1RD09
physicalInsertTime: 2017-08-11T12:25:00-07:00
physicalSerial: PHLE6514003R6P4BGN-1
physicalSize: 2.910957656800747T
slotNumber: "PCI Slot: 1; FDOM: 1"
status: failed - dropped for replacement
```

The Exadata Database Server flash disk statuses are as follows:

- normal
- normal - dropped for replacement
- failed
- failed - dropped for replacement
- failed - rejected due to incorrect disk model
- failed - rejected due to incorrect disk model - dropped for replacement
- failed - rejected due to wrong slot
- failed - rejected due to wrong slot - dropped for replacement

- warning - peer failure
- warning - predictive failure, write-through caching
- warning - predictive failure
- warning - predictive failure - dropped for replacement
- warning - write-through caching

## 2.3.2 Performing a Hot-Pluggable Replacement of a Flash Disk

For Exadata Database Machine X7-8 and X8-8 models, the database server uses hot-pluggable flash disks instead of hard disk drives.

1. Determine if the flash disk is ready to be replaced.

When performing a hot-pluggable replacement of a flash device on Exadata Database Machine X7-8 and X8-8 database servers, the disk status should be `Dropped for replacement`, which indicates the flash disk is ready for online replacement.

```
DBMCLI> LIST PHYSICALDISK WHERE DISKTYPE=flashdisk AND STATUS LIKE
'.*dropped
for replacement.*' DETAIL
```

```

name: FLASH_1_1
deviceName: /dev/nvme0n1
diskType: FlashDisk
luns: 1_1
makeModel: "Oracle Flash Accelerator F640 PCIe
Card"
physicalFirmware: QDV1RD09
physicalInsertTime: 2017-08-11T12:25:00-07:00
physicalSerial: PHLE6514003R6P4BGN-1
physicalSize: 2.910957656800747T
slotNumber: "PCI Slot: 1; FDOM: 1"
status: failed - dropped for replacement
```

2. Locate the failed flash disk based on the PCI number and FDOM number.

A white Locator LED is lit to help locate the affected database server. An amber Fault-Service Required LED is lit to identify the affected flash card.

3. Make sure the DPCC OK LED is off on the card.

### Caution:

Removing a card with the DPCC OK LED on could result in a system crash. If a failed disk has a status of `Failed - dropped for replacement` but the DPCC OK LED is still on, contact Oracle Support.

4. Remove and replace the failed flash disk.
  - a. Slide out the DPCC and replace the flash card inside.
  - b. Slide the DPCC carrier back to the slot.

5. Use a stylus to press both ATTN buttons on the front of the DPCC.
  - If only a single PCIe card is present, press only the corresponding ATTN button.
  - If you are not performing a hot-pluggable replacement, then this step is not necessary.

The buttons alert the system to a request to bring the devices online. When the system acknowledges the request, it lights the DPCC OK LED indicators on the DPCC. If you do not press the ATTN buttons, then the flash disk will not be detected by the operating system.

## 2.4 Adding the Disk Expansion Kit to Database Servers

You can add local storage space to an Oracle Exadata Database Server by using a disk expansion kit.

- [Adding the Disk Expansion Kit to Database Servers: X9M-2](#)
- [Adding the Disk Expansion Kit to Database Servers: X8M-2 and Prior](#)

### 2.4.1 Adding the Disk Expansion Kit to Database Servers: X9M-2

This procedure describes adding the disk expansion kit to Oracle Exadata Database Machine X9M-2 database servers.

Before starting, the server must be powered on so that the disk controller can sense the addition of the new drives.

To add the disk expansion kit:

1. Remove the plastic filler panels that cover the vacant drive bays and insert the drives contained in the disk expansion kit.

The drives may be installed in any order. Both drives must be installed at the same time (within 30 minutes) so that the disk controller can sense the new drives before any of them enter a power-saving mode.

When the disk controller senses the new drives, the following occurs automatically:

- The new drives are configured into a new mirrored (RAID1) physical volume (PV) at `/dev/md26`.
- The newly created PV is added to the volume group (VG) named `VGExaDb`.

These automatic configuration steps normally completed within one minute after the disks are inserted.

2. Confirm the addition of the disk expansion kit.

Run the following commands as the system administrator (`root`):

- a. Run `pvs` and confirm the presence of the new PV at `/dev/md26`.

For example:

```
pvs
PV VG Fmt Attr PSize PFree
/dev/md25 VGExaDb lvm2 a-- 3.48t 39.75g
/dev/md26 VGExaDb lvm2 a-- 3.49t 3.49t
```



- b. Run `vgs` and confirm the new size of `VGExaDb` along with the amount of free space.

For example:

```
vgs
VG #PV #LV #SN Attr VSize VFree
VGExaDb 2 11 0 wz--n- <6.98t 3.53t
```

You can now use the additional space to create or extend LVM partitions and their associated file systems.

### Related Topics

- [Extending LVM Partitions](#)  
Logical Volume Manager (LVM) provides flexibility to reorganize the partitions in the database servers.
- [Expanding /EXAVMIMAGES on the KVM host](#)  
Use this procedure to expand `/EXAVMIMAGES` using available space on the KVM host.

## 2.4.2 Adding the Disk Expansion Kit to Database Servers: X8M-2 and Prior

Note the following restrictions and requirements:

- The disk expansion kit is supported only on 2-socket systems starting with Oracle Exadata Database Machine X5-2.
- Oracle Exadata System Software release 12.1.2.3.0 or later is required.
- For systems running Oracle Linux 6 (OL6), a reboot is required for the Linux kernel to recognize a newly added disk partition.
- If you are adding the disk expansion kit to an Oracle Exadata Database Machine X7-2 system, and you are using an Oracle Exadata System Software release before 18.1.11, then ensure that the following symbolic link is present on the database server before proceeding:

```
ls -l /opt/MegaRAID/MegaCli/MegaCli64
lrwxrwxrwx 1 root root 31 Jun 4 03:40 /opt/MegaRAID/MegaCli/
MegaCli64 -> /opt/MegaRAID/storcli/storcli64
```

If the symbolic link is not present, then use the following commands to create it:

```
mkdir -p /opt/MegaRAID/MegaCli
ln -s /opt/MegaRAID/storcli/storcli64 /opt/MegaRAID/MegaCli/
MegaCli64
```

To add the disk expansion kit to an Oracle Exadata Database Server:

1. Remove the plastic filler panels that cover the vacant drive bays and insert the four drives that are contained in the disk expansion kit.

The server should be powered on so that the disk controller can sense the new drives.

The drives may be installed in any order. All four drives must be installed at the same time (within 30 minutes) so that the disk controller can sense the new drives before any of them enter a power-saving mode.

When the disk controller senses the new drives, the RAID reconstruction process automatically begins.

2. Monitor the server alert history. Ensure that the RAID reconstruction process completes successfully before proceeding.

The RAID reconstruction process may take several hours to complete (7 hours in the following example). Look out for the clear message (message 1\_2 below), which indicates that the RAID reconstruction process is completed.

```
dbmcli -e list alerthistory

 1_1 2016-02-15T14:01:00-08:00 warning "A disk
expansion kit was installed. The additional physical drives were
automatically
added to the existing RAID5 configuration, and reconstruction of the
corresponding virtual drive was automatically started."

 1_2 2016-02-15T21:01:01-08:00 clear "Virtual
drive
reconstruction due to disk expansion was completed."
```

At the end of the RAID reconstruction process, the virtual drive at `/dev/sda` includes the additional storage space from the disk expansion kit.

3. If you are adding the disk expansion kit as part of deploying a new system, then proceed with this step. Otherwise, skip to the next step.

This section uses `reclaimdisks.sh` to extend the `VGExaDb` volume group so that it consumes the additional storage space provided by the disk expansion kit.

`reclaimdisks.sh` works only during initial deployment, before installation of the database software.

- a. Run `/opt/oracle.SupportTools/reclaimdisks.sh -extend-vgexadb` to extend the `VGExaDb` volume group.

If prompted to fix the GUID Partition Table (GPT) or to continue with the current settings, enter `F` to fix the GPT.

For example:

```
/opt/oracle.SupportTools/reclaimdisks.sh -extend-vgexadb
Model is ORACLE SERVER X6-2
Number of LSI controllers: 1
Physical disks found: 8 (252:0 252:1 252:2 252:3 252:4 252:5 252:6
252:7)
Logical drives found: 1
Linux logical drive: 0
RAID Level for the Linux logical drive: 5
Physical disks in the Linux logical drive: 8 (252:0 252:1 252:2 252:3
252:4 252:5 252:6 252:7)
Dedicated Hot Spares for the Linux logical drive: 0
Global Hot Spares: 0
```

Valid. Disks configuration: RAID5 from 8 disks with no global and dedicated hot spare disks.

Valid. Booted: Linux. Layout: Linux + DOM0.

```
[INFO] Size of system block device /dev/sda: 4193GB
[INFO] Last partition on /dev/sda ends on: 1797GB
[INFO] Unused space detected on the system block
device: /dev/sda
[INFO] Label of partition table on /dev/sda: gpt
[INFO] Adjust the partition table to use all of the space
on /dev/sda
[INFO] Respond to the following prompt by typing 'F'
Warning: Not all of the space available to /dev/sda appears to
be used, you can fix the GPT to use all of the space (an extra
4679680000 blocks) or
continue with the current setting?
Fix/Ignore? F
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 4193GB
Sector size (logical/physical): 512B/512B
Partition Table: gpt
```

| Number | Start  | End    | Size   | File system | Name    | Flags |
|--------|--------|--------|--------|-------------|---------|-------|
| 1      | 32.8kB | 537MB  | 537MB  | ext4        | primary | boot  |
| 2      | 537MB  | 123GB  | 122GB  |             | primary | lvm   |
| 3      | 123GB  | 1690GB | 1567GB |             | primary |       |
| 4      | 1690GB | 1797GB | 107GB  |             | primary | lvm   |

```
[INFO] Check for Linux with inactive DOM0 system disk
[INFO] Valid Linux with inactive DOM0 system disk is
detected
[INFO] Number of partitions on the system device /dev/sda: 4
[INFO] Higher partition number on the system device /dev/
sda: 4
[INFO] Last sector on the system device /dev/sda: 8189440000
[INFO] End sector of the last partition on the system
device /dev/sda: 3509759000
[INFO] Unmount /u01 from /dev/mapper/VGExaDbOra-LVDbOra1
[INFO] Remove inactive system logical volume /dev/VGExaDb/
LVDbSys3
[INFO] Remove xen files from /boot
[INFO] Remove logical volume /dev/VGExaDbOra/LVDbOra1
[INFO] Remove volume group VGExaDbOra
[INFO] Remove physical volume /dev/sda4
[INFO] Remove partition /dev/sda4
[INFO] Remove device /dev/sda4
[INFO] Remove partition /dev/sda3
[INFO] Remove device /dev/sda3
[INFO] Create primary partition 3 using 240132160 8189439966
[INFO] Set lvm flag for the primary partition 3 on
device /dev/sda
[INFO] Add device /dev/sda3
[INFO] Primary LVM partition /dev/sda3 has size 7949307807
sectors
[INFO] Create physical volume on partition /dev/sda3
[INFO] LVM Physical Volume /dev/sda3 has size 3654340511
```

```
sectors
[INFO] Size of LVM physical volume less than size of device /dev/
sda3
[INFO] Remove LVM physical volume /dev/sda3
[INFO] Reboot is required to apply the changes in the partition
table
```

- b. Examine the end of the output from the previous command. If a reboot is not required, then skip to the next substep. If a reboot is required, then reboot the server and re-run `/opt/oracle.SupportTools/reclaimdisks.sh -extend-vgexadb`.

For example:

```
shutdown -r now
```

Then, after system reboot:

```
/opt/oracle.SupportTools/reclaimdisks.sh -extend-vgexadb
```

- c. Run `/opt/oracle.SupportTools/reclaimdisks.sh` with no arguments. In the output, confirm that there are no errors and that the output references the additional disks from the disk expansion kit.

```
/opt/oracle.SupportTools/reclaimdisks.sh
Model is ORACLE SERVER X6-2
Number of LSI controllers: 1
Physical disks found: 8 (252:0 252:1 252:2 252:3 252:4 252:5 252:6
252:7)
Logical drives found: 1
Linux logical drive: 0
RAID Level for the Linux logical drive: 5
Physical disks in the Linux logical drive: 8 (252:0 252:1 252:2 252:3
252:4 252:5 252:6 252:7)
Dedicated Hot Spares for the Linux logical drive: 0
Global Hot Spares: 0
Valid. Disks configuration: RAID5 from 8 disks with no global and
dedicated hot spare disks.
Valid. Booted: Linux. Layout: Linux.
```

You can now continue with deploying the system and use the additional storage space provided by the disk expansion kit. Do not perform the next step.

4. If you are adding the disk expansion kit to a previously deployed system, then proceed with this step.

This step uses Operating System commands to consume the additional storage space provided by the disk expansion kit.

- a. Run `parted` to view the sector information for `/dev/sda`.

If you see a request to fix the GPT, respond with `F`.

```
parted /dev/sda
GNU Parted 2.1Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) unit s
```

```
(parted) print
Warning: Not all of the space available to /dev/sda appears to
be used, you can
fix the GPT to use all of the space (an extra 4679680000 blocks)
or continue
with the current setting? Fix/Ignore? F
```

```
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
```

| Number | Start    | End        | Size       | File system |
|--------|----------|------------|------------|-------------|
| 1      | 64s      | 1046591s   | 1046528s   | ext3        |
| 4      | 1046592s | 1048639s   | 2048s      |             |
| 2      | 1048640s | 240132159s | 239083520s |             |

```
(parted) q
```

Examine the output and note the disk size. Note also the largest end sector value, which should be the end sector of the last partition. In the preceding example, the disk size is 8189440000 sectors, and the largest end sector value is 240132159. You will use these values in the next step.

- b.** Create a new partition in `/dev/sda`.

The command requires a start sector and an end sector, which you must derive from the values that you noted previously.

For the start sector, add 1 to largest end sector value from the previous step.

For example:  $240132159 + 1 = 240132160$ .

For the end sector, subtract 34 from the disk size value. For example:

$8189440000 - 34 = 8189439966$ .

```
parted -s /dev/sda mkpart primary 240132160s 8189439966s
```

This command produces no output.

- c.** Review the updated partition table and take note of the partition number for the new partition.

In this example, the new partition number is 3. You will use this value in the following commands.

```
parted -s /dev/sda unit s print
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Number Start End Size File system
Name Flags
```

```

1 64s 1046591s 1046528s ext4 primary
boot
4 1046592s 1048639s 2048s primary
bios_grub
2 1048640s 240132159s 239083520s primary
lvm
3 240132160s 8189439966s 7949307807s primary

```

- d. Set the LVM flag for the new partition.

In this example, the new partition number is 3. Use the partition number that you observed in the previous step.

```
parted -s /dev/sda set 3 lvm on
```

- e. Examine the output from the previous step, and reboot the server if required.

For systems running Oracle Linux 6 (OL6), a server reboot is required for the Linux kernel to recognize the newly added disk partition.

Reboot the server if the output from the previous step contains a message similar to the following:

```
Warning: WARNING: the kernel failed to re-read the partition table
on /dev/sda (Device or
resource busy). As a result, it may not reflect all of your changes
until after reboot.
```

- f. Create an LVM physical volume (PV) on the newly created partition.

In this example, the new partition number is 3, so the physical volume is `/dev/sda3`. Adjust the command based on the partition number that you observed previously.

```
lvm pvcreate --force /dev/sda3
Physical volume "/dev/sda3" successfully created
```

- g. Extend the LVM volume group `VGExaDb` to use the newly created physical volume.

```
lvm vgextend VGExaDb /dev/sda3
Volume group "VGExaDb" successfully extended
```

You can now use the additional storage space provided by the disk expansion kit to extend various storage volumes and file systems on the server.

### Related Topics

- [Extending LVM Partitions](#)  
Logical Volume Manager (LVM) provides flexibility to reorganize the partitions in the database servers.
- [Expanding /EXAVMIMAGES on the KVM host](#)  
Use this procedure to expand `/EXAVMIMAGES` using available space on the KVM host.

## 2.5 Adding Memory Expansion Kit to Database Servers

Additional memory can be added to database servers. The following procedure describes how to add the memory:

1. Power down the database server.
2. Replace the plastic fillers with the DIMMs.
  - X9-2: See "Servicing the DIMMs" in *Oracle Exadata X9-2 Database Server Service Manual* at <https://docs.oracle.com/en/servers/x86/x9-2/exa-database/servicing-dimms.html>
  - X8-2: See "Servicing the DIMMs (CRU)" in *Oracle Exadata X8-2 Database Server Service Manual* at <https://docs.oracle.com/en/servers/x86/x8-2/exadata-db-service-manual/gqvkr.html>
  - X8-8 and X9-8: See "Servicing DIMMs (CRU)" in *Oracle Exadata Database Server X8-8 Service Manual* at [https://docs.oracle.com/cd/E93360\\_01/html/E96097/gsumt.html#scrolltoc](https://docs.oracle.com/cd/E93360_01/html/E96097/gsumt.html#scrolltoc)
  - X7-2: See "Servicing the DIMMs (CRU)" in *Oracle Server X7-2 Service Manual* at [http://docs.oracle.com/cd/E72435\\_01/html/E72445/gqvkr.html#scrolltoc](http://docs.oracle.com/cd/E72435_01/html/E72445/gqvkr.html#scrolltoc)
  - X7-8: See "Servicing DIMMs" in *Oracle Server X7-8 Service Manual* at [https://docs.oracle.com/cd/E71925\\_01/html/E71936/gsumt.html#scrolltoc](https://docs.oracle.com/cd/E71925_01/html/E71936/gsumt.html#scrolltoc)
  - X6-2: See "Servicing the DIMMs (CRU)" in *Oracle Server X6-2 Service Manual* at [http://docs.oracle.com/cd/E62159\\_01/html/E62171/z40003f01425075.html#scrolltoc](http://docs.oracle.com/cd/E62159_01/html/E62171/z40003f01425075.html#scrolltoc)
  - X5-2: See "Servicing the DIMMs (CRU)" in *Oracle Server X5-2 Service Manual* at [http://docs.oracle.com/cd/E41059\\_01/html/E48320/z40003f01425075.html](http://docs.oracle.com/cd/E41059_01/html/E48320/z40003f01425075.html)
  - X5-8: See "Replace a Failed DIMM" in *Oracle Server X5-8 Service Manual* at [https://docs.oracle.com/cd/E56301\\_01/html/E56311/gnsdu.html#scrolltoc](https://docs.oracle.com/cd/E56301_01/html/E56311/gnsdu.html#scrolltoc)
  - X4-8: See "Replace a Failed DIMM" in *Sun Server X4-8 Service Manual* at [http://docs.oracle.com/cd/E40591\\_01/html/E40317/gnsdu.html#scrolltoc](http://docs.oracle.com/cd/E40591_01/html/E40317/gnsdu.html#scrolltoc)
  - X4-2: See "Servicing the DIMMs (CRU)" in *Sun Server X4-2 Service Manual* at [http://docs.oracle.com/cd/E36975\\_01/html/E38045/z40003f01425075.html#scrolltoc](http://docs.oracle.com/cd/E36975_01/html/E38045/z40003f01425075.html#scrolltoc)
  - X3-2: See "Servicing the DIMMs (CRU)" in *Sun Server X3-2 (formerly Sun Fire X4170 M3) Service Manual* at [http://docs.oracle.com/cd/E22368\\_01/html/E27242/z40003f01425075.html#scrolltoc](http://docs.oracle.com/cd/E22368_01/html/E27242/z40003f01425075.html#scrolltoc)
  - X2-8: See "Removing and Installing DIMMs (CRU)" in *Sun Server X2-8 (formerly Sun Fire X4800 M2) Service Manual* at [http://docs.oracle.com/cd/E20815\\_01/html/E20819/gjfvw.html#scrolltoc](http://docs.oracle.com/cd/E20815_01/html/E20819/gjfvw.html#scrolltoc)
  - Sun Fire X4170 M2: See "Servicing DDR3 DIMMs" in *Sun Fire X4170 M2 Server Service Manual* at [http://docs.oracle.com/cd/E19762-01/E22369-02/motherboard\\_x4170.html#50581523\\_71526](http://docs.oracle.com/cd/E19762-01/E22369-02/motherboard_x4170.html#50581523_71526)
3. Power on the database server.
4. Add the database server back to the cluster.

Additional notes:

- Memory for Sun Server X4-2 Oracle Database Servers and Sun Server X3-2 Oracle Database Servers can be expanded to a maximum of 512 GB with the memory expansion kit.
- Memory for Sun Fire X4170 Oracle Database Servers can be expanded to a maximum of 144 GB by removing the existing memory, and replacing it with three X2-2 Memory Expansion Kits.

- Sun Fire X4170 M2 Oracle Database Servers ship from the factory with 96 GB of memory, with 12 of the 18 DIMM slots populated with 8 GB DIMMs. The optional X2-2 Memory Expansion Kit can be used to populate the remaining 6 empty slots with 16 GB DIMMs to bring the total memory to 192 GB (12 x 8 GB and 6 x 16GB).

The memory expansion kit is primarily for consolidation workloads where many databases are run on each database server. In this scenario, the CPU usage is often low while the memory usage is very high.

However, there is a downside to populating all the memory slots as the frequency of the memory DIMMs drop to 800 MHz from 1333 MHz. The performance effect of the slower memory appears as increased CPU utilization. The average measured increase in CPU utilization is typically between 5% and 10%. The increase varies greatly by workload. In test workloads, several workloads had almost zero increase, while one workload had as high as a 20% increase.

- When adding memory to Oracle Exadata Database Machines running Oracle Linux, Oracle recommends updating the `/etc/security/limits.conf` file with the following:

```
oracle soft memlock 75%
oracle hard memlock 75%
```

## 2.6 Verifying and Modifying the Link Speed on the Client Network Ports for X7 and Later Systems

You can configure 10 GbE connections or 25 GbE connections on the client network on Exadata Database Machine X7 and later database servers.



### Note:

You should configure the client network ports using Oracle Exadata Deployment Assistant (OEDA) during system deployment. See Using Oracle Exadata Deployment Assistant.

The following steps may be necessary to configure a client access port if the OEDA deployment was not performed or was performed incorrectly. You can also use these steps to change the client network from 10 GbE to 25 GbE, or from 25 GbE to 10 GbE.

1. For each network interface (designated by `x`) that does not have the link detected, run the following commands:

- For 10GbE network interfaces:

```
ifdown ethx
ethtool -s ethx 10000 duplex full autoneg off
ifup ethx
ethtool ethx
```

For 10 Gb/s, you must use SFP+ transceivers; SFP28 transceivers do not support 10 Gb/s traffic.



- For 25GbE network interfaces:

```
ifdown ethx
ethtool -s ethx 25000 duplex full autoneg off
ifup ethx
ethtool ethx
```

2. Confirm that the output from the `ethtool` command shows `yes` for Link detected.

**Link detected: yes**

3. Edit the appropriate files in `/etc/sysconfig/network-scripts`, where `x` is the number associated with the network interface.

- a. Locate the `/etc/sysconfig/network-scripts/ifcfg-ethx` file. Add the following lines, if they are not already present in the file:

- For 10 GbE network interfaces:

```
ONBOOT=YES
ETHTOOL_OPTS="speed 10000 duplex full autoneg off"
```

- For 25 GbE network interfaces:

```
ONBOOT=YES
ETHTOOL_OPTS="speed 25000 duplex full autoneg off"
```

- b. Repeat the previous step for all network interfaces that do not have the `ETHTOOL_OPTS` setting in the associated `ifcfg-ethx` file and are connected to 10 GbE or 25 GbE switches.

The network interface should now show the link as detected. These changes are persistent, and do not need to be repeated after a server reboot.

4. Check the ILOM on each compute node to validate the LAN on Motherboard is properly configured to detect the 25 GbE transceiver.

```
show /HOST/network
/HOST/network
Targets:

Properties:
 active_media = none
 auto_media_detection = enabled
 current_active_media = (none)

Commands:
 cd
 set
 show
```

If the NIC is not working, change the `active_media` and `current_active_media` to the proper values:

- For 25 GbE transceivers (Fiber or Copper) these parameters should be set to `SPF28`

- For 10 GbE network using RJ-45 ended CAT6 cables, these parameters should be set to `RJ45`

## 2.7 Adding and Configuring an Extra Network Card on Exadata Database Machine X6-2 and Later

You can add an additional network card on Exadata Database Machine X6-2 and later systems.

### Prerequisites

Ensure you are using the correct link speed for Exadata Database Machine X7-2 and later compute nodes. Complete the steps in [Verifying and Modifying the Link Speed on the Client Network Ports for X7 and Later Systems](#).

### Exadata Database Machine X6-2

Exadata Database Machine X6-2 database server offers highly available copper 10G network on the motherboard, and an optical 10G network via a PCI card on slot 2. Oracle offers an additional Ethernet card for customers that require additional connectivity. The additional card provides either dual port 10GE copper connectivity (part number 7100488) or dual port 10GE optical connectivity (part number X1109A-Z). You install this card in PCIe slot 1 on the Exadata Database Machine X6-2 database server.

After you install the card and connect it to the network, the Oracle Exadata System Software automatically recognizes the new card and configures the two ports as `eth6` and `eth7` interfaces on the X6-2 database server. You can use these additional ports to provide an additional client network, or to create a separate backup or data recovery network. On a database server that runs virtual machines, you could use this to isolate traffic from two virtual machines.

### Exadata Database Machine X7-2

Exadata Database Machine X7-2 and later database servers offer 2 copper (RJ45) or 2 optical (SFP28) network connections on the motherboard plus 2 optical (SFP28) network connections in PCIe card slot 1. Oracle offers an additional 4 copper (RJ45) 10G network connections for customers that require additional connectivity. The additional card is the Oracle Quad Port 10 GBase-T card (part number 7111181). You install this card in PCIe slot 3 on the database server.

After you install the card and connect it to the network, the Oracle Exadata System Software automatically recognizes the new card and configures the four ports as `eth5` to `eth8` interfaces on the database server. You can use these additional ports to provide an additional client network, or to create a separate backup or data recovery networks. On a database server that runs virtual machines, you could use this additional client network to isolate traffic from two virtual machines.

After you add a card to a database server, you need to configure the card. See the following topics for instructions:

- [Viewing the Network Interfaces](#)
- [Configuring the Additional Network Card for a Non-Oracle VM Environment](#)
- [Configuring the Additional Network Card for an Oracle VM Environment](#)

### Exadata Database Machine X8-2 and X8M-2

Exadata Database Machine X8-2 and X8M-2 database servers offer 2 copper (RJ45) or 2 copper/optical (SFP28) network connections on the motherboard plus 2 optical (SFP28) network connections in PCIe card slot 1. Oracle offers an additional 4 copper 1/10G (RJ45) or 2 optical 10/25G (SFP28) network connections for customers that require additional connectivity. The additional cards are:

- Quad-Port 10 GBase-T card (part number 7111181)
- Dual-Port 25 Gb Ethernet adapter (part number 7118016)

The additional card is installed in PCIe slot 3 on the database server.

After you install the card and connect it to the network, the Oracle Exadata System Software automatically recognizes the new card and configures either the four ports as `eth5` to `eth8` interfaces for the quad port card, or `eth5` and `eth6` for the dual port card on the database server. You can use these additional ports to provide an additional client network, or to create a separate backup or data recovery networks. On a database server that runs virtual machines, you could use this additional client network to isolate traffic from two virtual machines.

After you add a card to a database server, you need to configure the card. See the following topics for instructions:

- [Viewing the Network Interfaces](#)
- [Configuring the Additional Network Card for a Non-Oracle VM Environment](#)
- [Configuring the Additional Network Card for an Oracle VM Environment](#)

### Exadata Database Machine X9M-2

Exadata Database Machine X9M-2 database servers offer a variety of flexible network configurations using one, two, or three network interface cards. Each card can be one of the following:

- Quad-Port 10 GBase-T card (RJ45) (part number 7111181)
- Dual-Port 25 Gb Ethernet adapter (SFP28) (part number 7118016)

After initial deployment, you can add network interface cards up to a maximum of three network interface cards on each database server.

On non-Eighth Rack systems only, you can optionally install one Dual-Port 100 Gb Ethernet adapter (QSFP28) (part number 7603661) in PCIe slot 2.

After you install the card and connect it to the network, the Oracle Exadata System Software automatically recognizes the new card and configures the physical ports as follows:

- Quad-Port 10 GBase-T card in PCIe slot 1: `eth1`, `eth2`, `eth3`, and `eth4`
- Dual-Port 25 Gb Ethernet adapter in PCIe slot 1: `eth1` and `eth2`
- Quad-Port 10 GBase-T card in PCIe slot 2: `eth5`, `eth6`, `eth7`, and `eth8`
- Dual-Port 25 Gb Ethernet adapter in PCIe slot 2: `eth5` and `eth6`
- Dual-Port 100 Gb Ethernet adapter in PCIe slot 2: `eth5` and `eth6`
- Quad-Port 10 GBase-T card in PCIe slot 3: `eth9`, `eth10`, `eth11`, and `eth12`

- Dual-Port 25 Gb Ethernet adapter in PCIe slot 3: eth9 and eth10

You can use the network ports to provide multiple client networks, or to create separate dedicated networks for backup/recovery and bulk data transfer. On a database server that runs virtual machines (VMs), you can use multiple client networks to isolate traffic for each VM cluster.

After you add a card to a database server, you need to configure the card. See the following topics for instructions:

- [Viewing the Network Interfaces](#)  
To view the network interfaces, you can run the `ipconf.pl` command.
- [Configuring the Additional Network Card for a Non-Oracle VM Environment](#)  
You can configure the additional network card on an Exadata Database Machine X6-2 or later database server for a non-Oracle VM environment.
- [Configuring the Additional Network Card for an Oracle VM Environment](#)  
You can configure the additional network card on an Exadata Database Machine X6-2 and later database server for an Oracle VM environment.

## 2.7.1 Viewing the Network Interfaces

To view the network interfaces, you can run the `ipconf.pl` command.

### Example 2-4 Viewing the default network interfaces for an Exadata Database Machine X8M-2 database server

The following example shows the output for an Exadata Database Machine X8M-2 database server without the additional network card. In addition to the RDMA Network Fabric interfaces, the output shows the interfaces for three network cards:

- A single port 1/10Gb card, eth0
- A dual port 10 or 25Gb card, on eth1 and eth2
- A dual port 10 or 25Gb card, on eth3 and eth4

```
root@scaz23adm01 ibdiagtools]# /opt/oracle.cellos/ipconf.pl
[Info]: ipconf command line: /opt/oracle.cellos/ipconf.pl
Logging started to /var/log/cellos/ipconf.log
Interface re0 is Linked. hca: mlx5_0
Interface re1 is Linked. hca: mlx5_0
Interface eth0 is Linked. driver/mac: igb/00:10:e0:c3:b7:9c
Interface eth1 is Unlinked. driver/mac: bnxt_en/00:10:e0:c3:b7:9d
(slave of bondeth0)
Interface eth2 is Linked. driver/mac: bnxt_en/00:10:e0:c3:b7:9d
(slave of bondeth0)
Interface eth3 is Unlinked. driver/mac: bnxt_en/00:0a:f7:c3:28:30
Interface eth4 is Unlinked. driver/mac: bnxt_en/00:0a:f7:c3:28:38
```

### Example 2-5 Viewing the default network interfaces for an Exadata Database Machine X7-2 or X8-2 database server

The following example shows the output for an Exadata Database Machine X7-2 or X8-2 database server without the additional network card. In addition to the RDMA Network Fabric interfaces, the output shows the interfaces for three network cards:

- A single port 10Gb card, on eth0
- A dual port 10 or 25Gb card, on eth1 and eth2
- A dual port 25Gb card, on eth3 and eth4

```
/opt/oracle.cellos/ipconf.pl
Logging started to /var/log/cellos/ipconf.log
Interface ib0 is Linked. hca: mlx4_0
Interface ib1 is Linked. hca: mlx4_0
Interface eth0 is Linked. driver/mac: igb/00:
10:e0:c3:ba:72
Interface eth1 is Linked. driver/mac: bnxt_en
/00:10:e0:c3:ba:73
Interface eth2 is Linked. driver/mac: bnxt_en
/00:10:e0:c3:ba:74
Interface eth3 is Linked. driver/mac: bnxt_en
/00:0a:f7:c3:14:a0 (slave of bondeth0)
Interface eth4 is Linked. driver/mac: bnxt_en
/00:0a:f7:c3:14:a0 (slave of bondeth0)
```

### Example 2-6 Viewing the default network interfaces for an Exadata Database Machine X6-2 database server

The following example shows the output for an Exadata Database Machine X6-2 database server without the additional network card. In addition to the RDMA Network Fabric interfaces, the output shows the interfaces for two network cards:

- A quad port 10Gb card, on eth0 to eth3
- A dual port 10Gb card, on eth4 and eth5

```
cd /opt/oracle.cellos/

./ipconf.pl
Logging started to /var/log/cellos/ipconf.log
Interface ib0 is Linked. hca: mlx4_0
Interface ib1 is Linked. hca: mlx4_0
Interface eth0 is Linked. driver/mac: ixgbe/
00:10:e0:8b:24:b6
Interface eth1 is Linked. driver/mac: ixgbe/
00:10:e0:8b:24:b7
Interface eth2 is Linked. driver/mac: ixgbe/
00:10:e0:8b:24:b8
Interface eth3 is Linked. driver/mac: ixgbe/
00:10:e0:8b:24:b9
Interface eth4 is Linked. driver/mac: ixgbe/
90:e2:ba:ac:20:ec (slave of bondeth0)
Interface eth5 is Linked. driver/mac: ixgbe/
90:e2:ba:ac:20:ec (slave of bondeth0)
```

## 2.7.2 Configuring the Additional Network Card for a Non-Oracle VM Environment

You can configure the additional network card on an Exadata Database Machine X6-2 or later database server for a non-Oracle VM environment.

This procedure assumes that you have already installed the network card in the Exadata Database Machine database server but have not yet completed the configuration with Oracle Exadata Deployment Assistant (OEDA).

### ⚠ WARNING:

If you have already installed Oracle Grid Infrastructure on Exadata Database Machine, then refer to the Oracle Clusterware documentation. Use caution when changing the network interfaces for the cluster.

1. Ensure you have the following information for the new network card.

You will need to input this information when you run `ipconf.pl`.

- IP address
- Netmask
- Gateway

2. Run the `ipconf.pl` script to configure the card.

The following example shows a sample `ipconf.pl` session. The output shows three network cards:

- A quad port 10Gb card, on eth0 to eth3
- A dual port 10Gb card, on eth4 and eth5, with only one port cabled
- A dual port 10Gb card, on eth6 and eth7, with only one port cabled. This is the new network card.

For sample output for Exadata Database Machine X7-2, see [Viewing the Network Interfaces](#).

```
cd /opt/oracle.cellos/
./ipconf.pl

Logging started to /var/log/cellos/ipconf.log
Interface ib0 is Linked. hca: mlx4_0
Interface ib1 is Linked. hca: mlx4_0
Interface eth0 is Linked. driver/mac:
ixgbe/00:10:e0:8b:22:e8 (slave of vmeth0)
Interface eth1 is Linked. driver/mac:
ixgbe/00:10:e0:8b:22:e9 (slave of bondeth0)
Interface eth2 is Linked. driver/mac:
ixgbe/00:10:e0:8b:22:e9 (slave of bondeth0)
Interface eth3 is Linked. driver/mac:
ixgbe/00:10:e0:8b:22:eb
Interface eth4 is Linked. driver/mac:
```

```

ixgbe/90:e2:ba:ac:1d:e4
Interface eth5 is Unlinked. driver/mac:
ixgbe/90:e2:ba:ac:1d:e5
Interface eth6 is ... Linked. driver/mac:
ixgbe/90:e2:ba:78:d0:10
Interface eth7 is Unlinked. driver/mac:
ixgbe/90:e2:ba:78:d0:11

bondeth0 eth1,eth2 UP vmbondeth0 10.128.1.169 255.255.240.0
10.128.0.1 SCAN test08client02.example.com
bondeth1 None UNCONF
bondeth2 None UNCONF
bondeth3 None UNCONF
Select interface name to configure or press Enter to continue: eth6
Selected interface. eth6
IP address or up or none: 10.129.19.34
Netmask: 255.255.248.0
Gateway (IP address or none) or none: 10.129.16.0

Select network type for interface from the list below
1: Management
2: SCAN
3: Other
Network type: 3

Fully qualified hostname or none: test08adm02-bkup.example.com
Continue configuring or re-configuring interfaces? (y/n) [y]: n
...
Do you want to configure basic ILOM settings (y/n) [y]: n
[Info]: Custom changes have been detected in /etc/sysconfig/network-
script
s/ifcfg-eth6
[Info]: Original file /etc/sysconfig/network-scripts/ifcfg-eth6
will be
saved in /opt/oracle.cellos/conf/network-scripts/
backup_by_Exadata_ipconf
[Info]: Original file /etc/ssh/sshd_config will be saved
in /etc/ssh/sshd_
config.backupbyExadata
[Info]: Generate /etc/ssh/sshd_config with ListenAddress(es)
10.128.18.106,
10.129.19.34, 10.128.1.169, 192.168.18.44, 192.168.18.45
Stopping sshd: [OK]
Starting sshd: [OK]
[Info]: Save /etc/sysctl.conf in /etc/sysctl.conf.backupbyExadata
[Info]: Adjust settings for IB interfaces in /etc/sysctl.conf
Re-login using new IP address 10.128.18.106 if you were
disconnected after
following commands
ip addr show vmbondeth0
ip addr show bondeth0
ip addr show vmeth0
ip addr show eth0
ifup eth6
sleep 1

```

```

ifup vmeth6
sleep 1
ip addr show vmeth6
ip addr show eth6
sleep 4
service sshd condrestart

```

**3. If you need to set up the network card with VLAN, perform these steps:**

**a. Add the VLAN ID to the `/opt/oracle.cellos/cell.conf` file.**

- Locate the Ethernet interface in the file. For example:

```

<Interfaces>
 <Gateway>10.129.16.0</Gateway>
 <Hostname>test08adm02-bkup.example.com</Hostname>
 <IP_address>10.129.19.34</IP_address>
 <IP_enabled>yes</IP_enabled>
 <IP_ssh_listen>enabled</IP_ssh_listen>
 <Inet_protocol>IPv4</Inet_protocol>
 <Name>eth6</Name>
 <Net_type>Other</Net_type>
 <Netmask>255.255.248.0</Netmask>
 <State>1</State>
 <Status>UP</Status>
 <Vlan_id>0</Vlan_id>
</Interfaces>

```

- Add the VLAN ID to the `<Vlan_id>` element. The following example shows the interface configured with VLAN ID of 2122.

```

<Interfaces>
 <Gateway>10.129.16.0</Gateway>
 <Hostname>test08adm02-bkup.example.com</Hostname>
 <IP_address>10.129.19.34</IP_address>
 <IP_enabled>yes</IP_enabled>
 <IP_ssh_listen>enabled</IP_ssh_listen>
 <Inet_protocol>IPv4</Inet_protocol>
 <Name>eth6</Name>
 <Net_type>Other</Net_type>
 <Netmask>255.255.248.0</Netmask>
 <State>1</State>
 <Status>UP</Status>
 <Vlan_id>2122</Vlan_id>
</Interfaces>

```

**b. Run the following command to configure the network interface using the modified `cell.conf` file:**

```
/opt/oracle.cellos/ipconf.pl -init -force
```



- c. Validate the interface has the VLAN configured by checking that the `/etc/sysconfig/network-scripts` directory contains files with the VLAN ID in the filename. For example, if the VLAN ID is 2122, you should see the following files:

```
ls -ltr /etc/sysconfig/network-scripts/*2122*

-rw-r----- 1 root root 250 Sep 7 14:39 /etc/sysconfig/network-
scripts/ifcfg-eth6.2122
-rw-r----- 1 root root 85 Sep 7 14:39 /etc/sysconfig/network-
scripts/route-eth6.2122
-rw-r----- 1 root root 56 Sep 7 14:39 /etc/sysconfig/network-
scripts/rule-eth6.2122
```

4. Reboot the database server for the changes to take effect.

```
shutdown -r now
```

5. Check that the network is working by pinging the gateway. For example:

```
ping 10.129.16.0
```

## 2.7.3 Configuring the Additional Network Card for an Oracle VM Environment

You can configure the additional network card on an Exadata Database Machine X6-2 and later database server for an Oracle VM environment.

This procedure assumes that you have already installed the network card in the Exadata Database Machine database server but have not yet completed the configuration with Oracle Exadata Deployment Assistant (OEDA).

### Caution:

Do not attempt this procedure if you have already installed Oracle Grid Infrastructure on Exadata Database Machine.

1. Add a section for the new network in the `/EXAVMIMAGES/conf/virtual_machine_config_file` configuration file in `dom0`.

The following example assumes the bridge is called `vmeth6`, and the interface is called `eth1`. The virtual machine configuration file name is `/EXAVMIMAGES/conf/test08adm01vm01.example.com-vm.xml`.

```
<Interfaces>
 <Bridge>vmeth6</Bridge>
 <Gateway>10.129.16.0</Gateway>
 <Hostname>test08adm02-bkup.example.com</Hostname>
 <IP_address>10.129.19.34</IP_address>
 <Name>eth1</Name>
 <IP_enabled>yes</IP_enabled>
```

```

<IP_ssh_listen>disabled</IP_ssh_listen>
<Net_type>Other</Net_type>
<Netmask>255.255.248.0</Netmask>
<Vlan_id>0</Vlan_id>
<State>1</State>
<Status>UP</Status>
</Interfaces>

```

If you are using VLANs, enter the appropriate VLAN ID [1-4095] in the `<Vlan_id>` element.

## 2. Create the bridge.

- a. To create an unbonded bridge named `vmeth6`:

```
/opt/exadata_ovm/exadata.img.domu_maker add-single-bridge-dom0
vmeth6
```

- b. To create an bonded bridge, use a command similar to the following:

```
/opt/exadata_ovm/exadata.img.domu_maker add-bonded-bridge-dom0
bridge_name slave1 slave2 [vlan]
```

`slave1` and `slave2` are the names of the bonded interfaces.

For example:

```
/opt/exadata_ovm/exadata.img.domu_maker add-bonded-bridge-dom0
vmbondeth1 eth6 eth7
```

## 3. (X2 to X8 servers only) Allocate the GUIDs for the InfiniBand Network Fabric:

```
/opt/exadata_ovm/exadata.img.domu_maker allocate-guids
virtual_machine_config_file virtual_machine_config_file_final
```

The virtual machine configuration files are located in the `/EXAVMIMAGES/conf` directory. For example:

```
/opt/exadata_ovm/exadata.img.domu_maker allocate-guids /EXAVMIMAGES/
conf/
test08adm01vm01.example.com-vm.xml /EXAVMIMAGES/conf/final-test08adm01vm01
.example.com-vm.xml
```

## 4. Shut down the guest and restart it.

```
/opt/exadata_ovm/exadata.img.domu_maker remove-domain /EXAVMIMAGES/conf
/final-test08adm01vm01.example.com-vm.xml
```

```
/opt/exadata_ovm/exadata.img.domu_maker start-domain /EXAVMIMAGES/conf
/final-test08adm01vm01.example.com-vm.xml
```

## 5. Once the guest is running, use the `ip addr` command to verify the interface is valid.

The following example checks the `eth1` interface.

```
ip addr show eth1
eth1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP
qlen 1000
 link/ether 00:16:3e:53:56:00 brd ff:ff:ff:ff:ff:ff
 inet 10.129.19.34/21 brd 10.129.23.255 scope global eth1
 valid_lft forever preferred_lft forever
```

## 2.8 Increasing the Number of Active Cores on Database Servers

You can increase the number of active cores on Exadata Database Machine using capacity-on-demand.

The number of active cores on the database servers on Oracle Exadata Database Machine X4-2 and newer systems can be reduced during installation. The number of active cores can be increased when additional capacity is needed. This is known as *capacity-on-demand*.

Additional cores are increased in increments of 2 cores on 2-socket systems, and in increments of 8 cores on 8-socket systems. The following table lists the capacity-on-demand core processor configurations.

**Table 2-3 Capacity-on-Demand Core Processor Configurations**

Exadata Database Machine	Eligible Systems	Minimum Cores per Server	Maximum Cores per Server	Core Increments
Exadata Database Machine X9M-2	Any configuration except Eighth Rack	14	64	From 14 to 64, in increments of 2: 14, 16, 18, ..., 62, 64
Exadata Database Machine X9M-2	Eighth rack	8	32	From 8 to 32, in increments of 2: 8, 10, 12, ..., 30, 32
Exadata Database Machine X7-2, X8-2, and X8M-2	Any configuration except Eighth Rack	14	48	From 14 to 48, in increments of 2: 14, 16, 18, ..., 46, 48
Exadata Database Machine X7-2, X8-2, and X8M-2	Eighth rack	8	24	From 8 to 24, in increments of 2: 8, 10, 12, ..., 22, 24
Oracle Exadata Database Machine X6-2	Any configuration except Eighth Rack	14	44	From 14 to 44, in increments of 2: 14, 16, 18, ..., 42, 44

**Table 2-3 (Cont.) Capacity-on-Demand Core Processor Configurations**

Exadata Database Machine	Eligible Systems	Minimum Cores per Server	Maximum Cores per Server	Core Increments
Oracle Exadata Database Machine X6-2	Eighth rack	8	22	From 8 to 22, in increments of 2: 8, 10, 12, ..., 20, 22
Oracle Exadata Database Machine X5-2	Any configuration except Eighth Rack	14	36	From 14 to 36, in increments of 2: 14, 16, 18, ..., 34, 36
Oracle Exadata Database Machine X5-2	Eighth rack	8	18	From 8 to 18, in increments of 2: 8, 10, 12, 14, 16, 18
Oracle Exadata Database Machine X4-2	Full rack Half rack Quarter rack	12	24	From 12 to 24, in increments of 2: 12, 14, 16, 18, 20, 22, 24
Exadata Database Machine X7-8, X8-8, X8M-8, and X9M-8	Any configuration	56	192	From 56 to 192, in increments of 8: 56, 64, 72, ..., 184, 192
Exadata Database Machine X6-8 and X5-8	Any configuration	56	144	From 56 to 144, in increments of 8: 56, 64, 72, ..., 136, 144
Oracle Exadata Database Machine X4-8	Full rack	48	120	From 48 to 120, in increments of 8: 48, 56, 64, ..., 112, 120

**Note:**

Oracle recommends licensing the same number of cores on each server, in case of failover.

Database servers can be added one at a time, and capacity-on-demand can be applied to the individual database servers. This option includes Oracle Exadata Database Machine X5-2 Eighth Racks.

The database server must be restarted after enabling additional cores. If the database servers are part of a cluster, then they can be enabled in a rolling fashion.

1. Verify the number of active physical cores using the following command:

```
DBMCLI> LIST DBSERVER attributes coreCount
```

2. Use the following command to increase the number of active physical cores:

```
DBMCLI> ALTER DBSERVER pendingCoreCount =
new_number_of_active_physical_cores
```

3. Verify the pending number of active physical cores using the following command:

```
DBMCLI> LIST DBSERVER attributes pendingCoreCount
```

4. Restart the server.

5. Verify the number of active physical cores using the following command:

```
DBMCLI> LIST DBSERVER attributes coreCount
```

### Related Topics

- *Oracle Exadata Database Machine Licensing Information User's Guide*

## 2.9 Extending LVM Partitions

Logical Volume Manager (LVM) provides flexibility to reorganize the partitions in the database servers.

### Note:

- Keep at least 1 GB of free space in the `VGExaDb` volume group. This space is used for the LVM snapshot created by the `dbnodeupdate.sh` utility during software maintenance.
- If you make snapshot-based backups of the `/` (root) and `/u01` directories by following the steps in [Creating a Snapshot-Based Backup of Oracle Linux Database Server](#), then keep at least 6 GB of free space in the `VGExaDb` volume group.

This section contains the following topics:

- [Extending the root LVM Partition](#)  
The procedure for extending the root LVM partition depends on your Oracle Exadata System Software release.
- [Resizing a Non-root LVM Partition](#)  
The procedure for resizing a non-root LVM partition depends on your Oracle Exadata System Software release.
- [Extending the Swap Partition](#)  
This procedure describes how to extend the size of the swap (`/swap`) partition.

## 2.9.1 Extending the root LVM Partition

The procedure for extending the root LVM partition depends on your Oracle Exadata System Software release.

- [Extending the root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later](#)
- [Extending the root LVM Partition on Systems Running Oracle Exadata System Software Earlier than Release 11.2.3.2.1](#)

You can extend the size of the root (/) partition on systems running Oracle Exadata System Software earlier than release 11.2.3.2.1 using this procedure.

### 2.9.1.1 Extending the root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later

The following procedure describes how to extend the size of the root (/) partition on systems running Oracle Exadata System Software release 11.2.3.2.1 or later:

 **Note:**

- This procedure does not require an outage on the server.
- For management domain systems, the active and inactive Sys LVM's are `LVDbSys2` and `LVDbSys3` instead of `LVDbSys1` and `LVDbSys2`.
- Make sure that `LVDbSys1` and `LVDbSys2` are sized the same.

1. Collect information about the current environment.
  - a. Use the `df` command to identify the amount of free and used space in the root partition (/).

```
df -h /
```

The following is an example of the output from the command:

```
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbSys1 30G 22G 6.2G 79% /
```

 **Note:**

The active root partition may be either `LVDbSys1` or `LVDbSys2`, depending on previous maintenance activities.

- b. Use the `lvs` command to display the current volume configuration.

```
lvs -o lv_name,lv_path,vg_name,lv_size
```

The following is an example of the output from the command:

LV	Path	VG	LSize
LVDboRa1	/dev/VGExaDb/LVDboRa1	VGExaDb	100.00g
LVDboSwap1	/dev/VGExaDb/LVDboSwap1	VGExaDb	24.00g
LVDboSys1	/dev/VGExaDb/LVDboSys1	VGExaDb	30.00g
LVDboSys2	/dev/VGExaDb/LVDboSys2	VGExaDb	30.00g
LVDboNotRemoveOrUse	/dev/VGExaDb/LVDboNotRemoveOrUse	VGExaDb	1.00g

2. Use the `df` command to identify the file system type that is used in the root partition (`/`).

```
df -hT /
```

The following is an example of the output from the command:

```
Filesystem Type Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDboSys1 ext3 30G 22G 6.2G 79% /
```

In this example, the file system type is `ext3`.

3. If the file system type is not `xf`s, use the following `tune2fs` command to check the online resize option. If the file system type is `xf`s, then you can skip this step.

```
tune2fs -l /dev/mapper/vg_name-lv_name | grep resize_inode
```

For example:

```
tune2fs -l /dev/mapper/VGExaDb-LVDboSys1 | grep resize_inode
```

The `resize_inode` option should be listed in the output from the command. If the option is not listed, then the file system must be unmounted before resizing the partition. Refer to [Extending the root LVM Partition on Systems Running Oracle Exadata System Software Earlier than Release 11.2.3.2.1](#) to resize the partition.

4. Verify there is available space in the volume group `VGExaDb` using the `vgdisplay` command.

```
vgdisplay -s
```

The following is an example of the output from the command:

```
"VGExaDb" 834.89 GB [184.00 GB used / 650.89 GB free]
```

The volume group must contain enough free space to increase the size of both system partitions, and maintain at least 1 GB of free space for the LVM snapshot created by the `dbnodeupdate.sh` utility during upgrade.

If there is not enough free space, then verify that the `reclaimdisks.sh` utility has been run. If the utility has not been run, then use the following command to reclaim disk space:

```
/opt/oracle.SupportTools/reclaimdisks.sh -free -reclaim
```

If the utility has been run and there is not enough free space, then the LVM cannot be resized.

 **Note:**

`reclaimdisks.sh` cannot run at the same time as a RAID rebuild (that is, a disk replacement or expansion). Wait until the RAID rebuild is complete, then run `reclaimdisks.sh`.

5. Resize both `LVDbSys1` and `LVDbSys2` logical volumes using the `lvextend` command.

In the following example, `XG` is the amount of space in GB that the logical volume will be extended. The amount of space added to each system partition must be the same.

```
lvextend -L +XG --verbose /dev/VGExaDb/LVDbSys1
lvextend -L +XG --verbose /dev/VGExaDb/LVDbSys2
```

The following example extends the logical volumes by 10 GB:

```
lvextend -L +10G /dev/VGExaDb/LVDbSys1
lvextend -L +10G /dev/VGExaDb/LVDbSys2
```

6. Resize the file system within the logical volume.

- For `ext3` and `ext4` file system types, use the `resize2fs` command:

```
resize2fs /dev/VGExaDb/LVDbSys1
resize2fs /dev/VGExaDb/LVDbSys2
```

- For the `xf`s file system type:

- a. Use the `xf`s\_`growfs` command to resize the active file system:

```
xfs_growfs /
```

- b. Determine the inactive root partition.

The inactive root partition is `LVDbSys1` or `LVDbSys2`, whichever is not currently mounted.

Examine the output from the `df` command to confirm the active partition. For example:

```
df -hT /
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbSys1 30G 22G 6.2G 79% /
```



The example shows `LVDbSys1` as the active partition. Therefore, the inactive partition is `LVDbSys2`.

- c. Mount the inactive root partition to a temporary location.

For example:

```
mkdir -p /tmp/mnt/root
mount -t xfs /dev/VGExaDb/LVDbSys2 /tmp/mnt/root
```

- d. Use the `xfs_growfs` command to resize the inactive file system:

```
xfs_growfs /tmp/mnt/root
```

- e. Unmount the inactive root partition.

For example:

```
umount /tmp/mnt/root
```

7. Verify the space was extended for the active system partition using the `df` command.

```
df -h /
```

## 2.9.1.2 Extending the root LVM Partition on Systems Running Oracle Exadata System Software Earlier than Release 11.2.3.2.1

You can extend the size of the root (`/`) partition on systems running Oracle Exadata System Software earlier than release 11.2.3.2.1 using this procedure.

### Note:

- This procedure requires the system to be offline and restarted.
- Keep at least 1 GB of free space in the `VGExaDb` volume group to be used for the LVM snapshot created by the `dbnodeupdate.sh` utility during software maintenance. If you make snapshot-based backups of the `/` (root) and `/u01` directories by following the steps in [Creating a Snapshot-Based Backup of Oracle Linux Database Server](#), then keep at least 6 GB of free space in the `VGExaDb` volume group.
- For management domain systems, active and inactive Sys LVM's are `LVDbSys2` and `LVDbSys3` instead of `LVDbSys1` and `LVDbSys2`.
- Make sure `LVDbSys1` and `LVDbSys2` are sized the same.

1. Collect information about the current environment.
  - a. Use the `df` command to identify the mount points for the root partition (`/`) and the non-root partition (`/u01`), and their respective LVMs.

The following is an example of the output from the command:

```
df
Filesystem 1K-blocks Used Available Use% /
Mounted on
/dev/mapper/VGExaDb-LVDbSys1 30963708 21867152 7523692 75% /
/dev/sda1 126427 16355 103648 14% /
boot
/dev/mapper/VGExaDb-LVDbOra1 103212320 67404336 30565104
69% /u01
tmpfs 84132864 3294608 80838256
4% /dev/shm
```

The file system name in the `df` command output is in the following format:

```
/dev/mapper/VolumeGroup-LogicalVolume
```

The full logical volume name of the root file system in the preceding example is `/dev/VGExaDb/LVDbSys1`.

- b.** Use the `lvscan` command to display logical volumes.

```
#lvm lvscan

ACTIVE '/dev/VGExaDb/LVDbSys1' [30.00 GB] inherit
ACTIVE '/dev/VGExaDb/LVDbSwap1' [24.00 GB] inherit
ACTIVE '/dev/VGExaDb/LVDbOra1' [100.00 GB] inherit
```

- c.** Use the `lvdisplay` command to display the current logical volume and the volume group configuration.

```
#lvm lvdisplay /dev/VGExaDb/LVDbSys1

--- Logical volume ---
LV Name /dev/VGExaDb/LVDbSys1
VG Name VGExaDb
LV UUID GScpD7-1Ka8-gLg9-oBo2-uWaM-ZZ4W-Keazih
LV Write Access read/write
LV Status available
open 1
LV Size 30.00 GB
Current LE 7680
Segments 1
Allocation inherit
Read ahead sectors auto
 - currently set to 256
Block device 253:0
```

- d.** Verify there is available space in the volume group `VGExaDb` so the logical volume can be extended.

```
lvm vgdisplay VGExaDb -s
"VGExaDb" 556.80 GB [154.00 GB used / 402.80 GB free]
```

If the command shows there is zero free space, then neither the logical volume or the file system can be extended.

2. Restart the server using the `diagnostics.iso` file.

See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.

3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. Unmount the `root` file system.

```
cd /
umount /mnt/cell
```

5. Verify the logical volume name.

```
lvm lvscan
ACTIVE '/dev/VGExaDb/LVDbSys1' [30.00 GB] inherit
ACTIVE '/dev/VGExaDb/LVDbSwap1' [24.00 GB] inherit
ACTIVE '/dev/VGExaDb/LVDbOra1' [100.00 GB] inherit
```

6. Resize the `LVDbSys1` and `LVDbSys2` holding the current and backup root file system.

In the following commands, `XG` is the amount of space in GB that the logical volume will be extended.

```
lvm lvextend -L+XG --verbose /dev/VGExaDb/LVDbSys1
lvm lvextend -L+XG --verbose /dev/VGExaDb/LVDbSys2
```

For example, if the logical volume is expanded 5 GB, then the commands would be:

```
lvm lvextend -L+5G --verbose /dev/VGExaDb/LVDbSys1
lvm lvextend -L+5G --verbose /dev/VGExaDb/LVDbSys2
```

7. Verify the file system is valid using `e2fsck`.

```
e2fsck -f /dev/VGExaDb/LVDbSys1
e2fsck -f /dev/VGExaDb/LVDbSys2
```

**8.** Resize the file system.

```
resize2fs -p /dev/VGExaDb/LVDbSys1
resize2fs -p /dev/VGExaDb/LVDbSys2
```

**9.** Restart the system in normal mode.

```
shutdown -r now
```

**10.** Log in to the system.**11.** Verify the root file system mount mounts without issues with the new size.

## 2.9.2 Resizing a Non-root LVM Partition

The procedure for resizing a non-root LVM partition depends on your Oracle Exadata System Software release.

- [Extending a Non-root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later](#)  
This procedure describes how to extend the size of a non-root (`/u01`) partition on systems running Oracle Exadata System Software release 11.2.3.2.1 or later.
- [Extending a Non-root LVM Partition on Systems Running Oracle Exadata System Software Earlier than Release 11.2.3.2.1](#)  
This procedure describes how to extend the size of a non-root (`/u01`) partition on systems running Oracle Exadata System Software earlier than release 11.2.3.2.1.
- [Reducing a Non-root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later](#)  
You can reduce the size of a non-root (`/u01`) partition on systems running Oracle Exadata System Software release 11.2.3.2.1 or later.

### 2.9.2.1 Extending a Non-root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later

This procedure describes how to extend the size of a non-root (`/u01`) partition on systems running Oracle Exadata System Software release 11.2.3.2.1 or later.

This procedure does not require an outage on the server.

- 1.** Collect information about the current environment.
  - a.** Use the `df` command to identify the amount of free and used space in the `/u01` partition.

```
df -h /u01
```

The following is an example of the output from the command:

```
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbOra1 99G 25G 70G 26% /u01
```

- b. Use the `lvs` command to display the current logical volume configuration used by the `/u01` file system.

```
lvs -o lv_name,lv_path,vg_name,lv_size
```

The following is an example of the output from the command:

LV	Path	VG	LSize
LVDbOra1	/dev/VGExaDb/LVDbOra1	VGExaDb	100.00G
LVDbSwap1	/dev/VGExaDb/LVDbSwap1	VGExaDb	24.00G
LVDbSys1	/dev/VGExaDb/LVDbSys1	VGExaDb	30.00G
LVDbSys2	/dev/VGExaDb/LVDbSys2	VGExaDb	30.00G

2. Use the `df` command to identify the file system type that is used in the `/u01` partition.

```
df -hT /u01
```

The following is an example of the output from the command:

Filesystem	Type	Size	Used	Avail	Use%	Mounted on
/dev/mapper/VGExaDb-LVDbOra1	xfs	99G	25G	70G	26%	/u01

In this example, the file system type is `xfs`.

3. If the file system type is not `xfs`, use the following `tune2fs` command to check the online resize option. If the file system type is `xfs`, then you can skip this step.

```
tune2fs -l /dev/mapper/vg_name | grep resize_inode
```

The `resize_inode` option should be listed in the output from the command. If the option is not listed, then the file system must be unmounted before resizing the partition. Refer to "[Extending a Non-root LVM Partition on Systems Running Oracle Exadata System Software Earlier than Release 11.2.3.2.1](#)" when resizing the partition.

4. Verify there is available space in the volume group `VGExaDb` using the `vgdisplay` command.

```
vgdisplay -s
```

The following is an example of the output from the command:

```
"VGExaDb" 834.89 GB [184.00 GB used / 650.89 GB free]
```

If the output shows there is less than 1 GB of free space, then neither the logical volume nor file system should be extended. Maintain at least 1 GB of free space in the `VGExaDb` volume group for the LVM snapshot created by the `dbnodeupdate.sh` utility during an upgrade.

If there is not enough free space, then verify that the `reclaimdisks.sh` utility has been run. If the utility has not been run, then use the following command to reclaim disk space:

```
/opt/oracle.SupportTools/reclaimdisks.sh -free -reclaim
```

If the utility has been run and there is not enough free space, then the LVM cannot be resized.

 **Note:**

- `reclaimdisks.sh` cannot run at the same time as a RAID rebuild (that is, a disk replacement or expansion). Wait until the RAID rebuild is complete, then run `reclaimdisks.sh`.

5. Resize the logical volume using the `lvextend` command.

```
lvextend -L +sizeG /dev/VGExaDb/LVDbOra1
```

In the preceding command, *size* is the amount of space to be added to the logical volume.

The following example extends the logical volume by 10 GB:

```
lvextend -L +10G /dev/VGExaDb/LVDbOra1
```

6. Resize the file system within the logical volume.

- For `ext3` and `ext4` file system types, use the `resize2fs` command:

```
resize2fs /dev/VGExaDb/LVDbOra1
```

- For the `xfs` file system type, use the `xfs_growfs` command:

```
xfs_growfs /u01
```

7. Verify the space was extended using the `df` command

```
df -h /u01
```

## 2.9.2.2 Extending a Non-root LVM Partition on Systems Running Oracle Exadata System Software Earlier than Release 11.2.3.2.1

This procedure describes how to extend the size of a non-root (`/u01`) partition on systems running Oracle Exadata System Software earlier than release 11.2.3.2.1.

In this procedure, `/dev/VGExaDb/LVDbOra1` is mounted at `/u01`.

 **Note:**

- Keep at least 1 GB of free space in the `VGExaDb` volume group. This space is used for the LVM snapshot created by the `dbnodeupdate.sh` utility during software maintenance.
- If you make snapshot-based backups of the `/` (root) and `/u01` directories by following the steps in [Creating a Snapshot-Based Backup of Oracle Linux Database Server](#), then keep at least 6 GB of free space in the `VGExaDb` volume group.

**1. Collect information about the current environment.**

- Use the `df` command to identify the mount points for the root partition (`/`) and the non-root partition (`/u01`), and their respective LVMs.

```
df
Filesystem 1K-blocks Used Available Use%
Mounted on
/dev/mapper/VGExaDb-LVDbSys1 30963708 21867152 7523692
75% /
/dev/sda1 126427 16355 103648
14% /boot
/dev/mapper/VGExaDb-LVDbOra1 103212320 67404336 30565104
69% /u01
tmpfs 84132864 3294608 80838256
4% /dev/shm
```

- Use the `lvm lvscan` command to display logical volumes.

```
ACTIVE '/dev/VGExaDb/LVDbSys1' [30.00 GB] inherit
ACTIVE '/dev/VGExaDb/LVDbSwap1' [24.00 GB] inherit
ACTIVE '/dev/VGExaDb/LVDbOra1' [100.00 GB] inherit
```

- Use the `lvdisplay` command to display the current volume group configuration.

```
lvdisplay /dev/VGExaDb/LVDbOra1

--- Logical volume ---
LV Name /dev/VGExaDb/LVDbOra1
VG Name VGExaDb
LV UUID vzoIE6-uZrX-10Du-UD78-314Y-WXmz-f7SXY
LV Write Access read/write
LV Status available
open 1
LV Size 100.00 GB
Current LE 25600
Segments 1
Allocation inherit
Read ahead sectors auto
```

```
- currently set to 256
Block device 253:2
```

- d. Verify there is available space in the volume group `VGExaDb` so the logical drive can be extended.

If the command shows there is zero free space, then neither the logical volume or file system can be extended.

```
lvm vgdisplay VGExaDb -s

"VGExaDb" 556.80 GB [154.00 GB used / 402.80 GB free]
```

2. Shut down any software that uses `/u01`.

The following software typically uses `/u01`:

- Oracle Clusterware, Oracle ASM, and Oracle Database

```
Grid_home/bin/crsctl stop crs
```

- Trace File Analyzer

```
Grid_home/bin/tfactl stop
```

- OS Watcher

```
/opt/oracle.oswatcher/osw/stopOSW.sh
```

- Oracle Enterprise Manager agent

```
(oracle)$ agent_home/bin/emctl stop agent
```

3. Unmount the partition as the `root` user.

```
umount /u01
```



 **Note:**

If the `umount` command reports that the file system is busy, then use the `fuser(1)` command to identify processes still accessing the file system that must be stopped before the `umount` command will succeed.

```
umount /u01
umount: /u01: device is busy
umount: /u01: device is busy

fuser -mv /u01

 USER PID ACCESS COMMAND
/u01: root 6788 ..c.. ssh
 root 8422 ..c.. bash
 root 11444 ..c.. su
 oracle 11445 ..c.. bash
 oracle 11816m mgr
 root 16451 ..c.. bash
```

**4. Verify the file system.**

```
e2fsck -f /dev/VGExaDb/LVDbOra1
```

**5. Extend the partition.**

In this example, the logical volume is expanded to 80% of the physical volume size. At the same time, the file system is resized with the command.

```
lvextend -L+XG --verbose /dev/VGExaDb/LVDbOra1
```

In the preceding command, *XG* is the amount of GB the logical volume will be extended. The following example shows how to extend the logical volume by an additional 200 GB:

```
lvextend -L+200G --verbose /dev/VGExaDb/LVDbOra1
```

 **Caution:**

Use extreme caution when reducing the size. The new size must be large enough to hold all the original content of the partition. To reduce the size, use a command similar to the following:

```
lvreduce -L60G --resizefs --verbose /dev/VGExaDb/LVDbOra1
```

In the preceding command, the size of `/u01` is reduced to 60 GB.

6. Check the `/u01` file system using the `e2fsck` command.

```
e2fsck -f /dev/VGExaDb/LVDbOra1
```

7. Resize the `/u01` file system.

```
resize2fs -p /dev/VGExaDb/LVDbOra1
```

8. Mount the partition.

```
mount -t ext3 /dev/VGExaDb/LVDbOra1 /u01
```

9. Verify the space was extended.

```
$ df -h /u01
```

10. Restart any software that was stopped in step 2.

- Oracle Clusterware, Oracle ASM, and Oracle Database

```
Grid_home/bin/crsctl start crs
```

- Trace File Analyzer

```
Grid_home/bin/tfactl start
```

- OS Watcher

```
/opt/oracle.cellos/vldrun -script oswatcher
```

- Oracle Enterprise Manager agent

```
(oracle)$ agent_home/bin/emctl start agent
```

### Related Topics

- [Creating a Snapshot-Based Backup of Oracle Linux Database Server with Customized Partitions](#)

## 2.9.2.3 Reducing a Non-root LVM Partition on Systems Running Oracle Exadata System Software Release 11.2.3.2.1 or Later

You can reduce the size of a non-root (`/u01`) partition on systems running Oracle Exadata System Software release 11.2.3.2.1 or later.

### Note:

- This procedure does not require an outage on the server.
- It is recommended that you back up your file system before performing this procedure.

1. Use the `df` command to determine the amount of free and used space in the `/u01` partition:

```
df -h /u01
```

The following is an example of the output:

```
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbOral 193G 25G 159G 14% /u01
```

2. Use the `lvm` command to display the current logical volume configuration used by the `/u01` file system.

In this example, the size of the `LVDbOral` partition needs to be reduced so that `LVDbSys2` (30.00 GB in size) can be created by the `dbserver_backup.sh` script.

```
lvm vgdisplay VGExaDb -s
"VGExaDb" 271.82 GB [250.04 GB used / 21.79 GB free]

lvm lvscan
ACTIVE '/dev/VGExaDb/LVDbSys1' [30.00 GB] inherit
ACTIVE '/dev/VGExaDb/LVDbSwap1' [24.00 GB] inherit
ACTIVE '/dev/VGExaDb/LVDbOral' [196.04 GB] inherit
```

3. Shut down any software that uses `/u01`.

The following software typically uses `/u01`:

- Oracle Clusterware, Oracle ASM, and Oracle Database

```
Grid_home/bin/crsctl stop crs
```

- Trace File Analyzer

```
Grid_home/bin/tfactl stop
```

- OS Watcher (releases earlier than 11.2.3.3.0)

```
/opt/oracle.oswatcher/osw/stopOSW.sh
```

- ExaWatcher (release 11.2.3.3.0 and later)

```
/opt/oracle.ExaWatcher/ExaWatcher.sh --stop
```

- Oracle Enterprise Manager agent

```
(oracle)$ agent_home/bin/emctl stop agent
```

4. Unmount the partition as the root user.

```
umount /u01
```

 **Note:**

If the `umount` command reports that the file system is busy, then use the `fuser(1)` command to identify the processes still accessing the file system that must be stopped before the `umount` command will succeed.

```
umount /u01
umount: /u01: device is busy
umount: /u01: device is busy

fuser -mv /u01

 USER PID ACCESS COMMAND
/u01: root 6788 ..c.. ssh
 root 8422 ..c.. bash
 root 11444 ..c.. su
 oracle 11445 ..c.. bash
 oracle 11816m mgr
 root 16451 ..c.. bash
```

**5. Verify the file system.**

```
e2fsck -f /dev/VGExaDb/LVDbOral

fsck 1.39 (29-May-2006)
e2fsck 1.39 (29-May-2006)
Pass 1: Checking inodes, blocks, and sizes
Pass 2: Checking directory structure
Pass 3: Checking directory connectivity
Pass 4: Checking reference counts
Pass 5: Checking group summary information
DBORA: 72831/25706496 files (2.1% non-contiguous), 7152946/51389440 blocks
```

**6. Resize the file system to the required size (120G in the example below).**

```
resize2fs /dev/VGExaDb/LVDbOral 120G
resize2fs 1.39 (29-May-2017)
Resizing the filesystem on /dev/VGExaDb/LVDbOral to 26214400 (4k) blocks.
The filesystem on /dev/VGExaDb/LVDbOral is now 26214400 blocks long.
```

**7. Resize the LVM to the desired size.**

```
lvm lvreduce -L 120G --verbose /dev/VGExaDb/LVDbOral
 Finding volume group VGExaDb
 WARNING: Reducing active logical volume to 120.00 GB
 THIS MAY DESTROY YOUR DATA (filesystem etc.)
Do you really want to reduce LVDbOral? [y/n]: y
 Archiving volume group "VGExaDb" metadata (seqno 8).
 Reducing logical volume LVDbOral to 120.00 GB
 Found volume group "VGExaDb"
 Found volume group "VGExaDb"
 Loading VGExaDb-LVDbOral table (253:2)
```

```
Suspending VGExaDb-LVDbOral (253:2) with device flush
Found volume group "VGExaDb"
Resuming VGExaDb-LVDbOral (253:2)
Creating volume group backup "/etc/lvm/backup/VGExaDb" (seqno
9).
Logical volume LVDbOral successfully resized
```

#### 8. Mount the partition.

```
mount -t ext3 /dev/VGExaDb/LVDbOral /u01
```

#### 9. Verify the space was reduced.

```
df -h /u01
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbOral 119G 25G 88G 22% /u01

lvm vgdisplay -s
"VGExaDb" 271.82 GB [174.00 GB used / 97.82 GB free]
```

#### 10. Restart any software that was stopped in step 3.

- Oracle Clusterware, Oracle ASM, and Oracle Database

```
Grid_home/bin/crsctl start crs
```

- Trace File Analyzer

```
Grid_home/bin/tfactl start
```

- OS Watcher (releases earlier than 11.2.3.3.0)

```
/opt/oracle.cellos/vldrun -script oswatcher
```

- ExaWatcher (release 11.2.3.3.0 to release 18.1.x)

```
/opt/oracle.cellos/vldrun -script oswatcher
```

- ExaWatcher (release 19.0.0.0 and later)

```
systemctl start ExaWatcher
```

- Oracle Enterprise Manager agent

```
(oracle)$ agent_home/bin/emctl start agent
```

## 2.9.3 Extending the Swap Partition

This procedure describes how to extend the size of the swap (`/swap`) partition.

### Note:

This procedure requires the system to be offline and restarted.

Keep at least 1 GB of free space in the `VGExaDb` volume group to be used for the Logical Volume Manager (LVM) snapshot created by the `dbnodeupdate.sh` utility during software maintenance. If you make snapshot-based backups of the `/` (root) and `/u01` directories by following the steps in "[Creating a Snapshot-Based Backup of Oracle Linux Database Server](#)", then keep at least 6 GB of free space in the `VGExaDb` volume group.

1. Collect information about the current environment.
  - a. Use the `swapon` command to identify the swap partition.

```
swapon -s
Filename Type Size Used Priority
/dev/dm-2 partition 25165816 0 -1
```

- b. Use the `lvm lvscan` command to display the logical volumes.

```
lvm lvscan
ACTIVE '/dev/VGExaDb/LVDbSys1' [30.00 GiB] inherit
ACTIVE '/dev/VGExaDb/LVDbSys2' [30.00 GiB] inherit
ACTIVE '/dev/VGExaDb/LVDbSwap1' [24.00 GiB] inherit
ACTIVE '/dev/VGExaDb/LVDbOral' [103.00 GiB] inherit
ACTIVE '/dev/VGExaDb/LVDoNotRemoveOrUse' [1.00 GiB] inherit
```

- c. Use the `vgdisplay` command to display the current volume group configuration.

```
vgdisplay
--- Volume group ---
VG Name VGExaDb
System ID
Format lvm2
Metadata Areas 1
Metadata Sequence No 4
VG Access read/write
VG Status resizable
MAX LV 0
Cur LV 3
Open LV 3
Max PV 0
Cur PV 1
Act PV 1
VG Size 556.80 GB
PE Size 4.00 MB
```

```

Total PE 142541
Alloc PE / Size 39424 / 154.00 GB
Free PE / Size 103117 / 402.80 GB
VG UUID po3xVH-9prk-ftEI-vijh-giTy-5chm-Av0fBu

```

- d. Use the `pvdisplay` command to display the name of the physical device created by LVM and used with the operating system.

```

pvdisplay
--- Physical volume ---
PV Name /dev/sda2
VG Name VGExaDb
PV Size 556.80 GB / not usable 2.30 MB
Allocatable yes
PE Size (KByte) 4096
Total PE 142541
Free PE 103117
Allocated PE 39424
PV UUID Eq0e7e-plfS-FyGN-zrvj-00qd-oUSb-55x2TX

```

2. Restart the server using the `diagnostics.iso` file.

See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.

3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```

Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.

```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. Verify the file system is valid.

Use the following command:

```
#fsck -f /dev/VGExaDb/LVDbSwap1
```

5. Extend the partition.

In this example, the logical volume is expanded to 80% of the physical volume size. At the same time, the file system is resized with this command. In the following command, the value for *LogicalVolumePath* is obtained by the `lvm`

lvscan command, and the value for *PhysicalVolumePath* is obtained by the `pvdisplay` command.

```
#lvextend -l+80%PVS --resizefs --verbose LogicalVolumePath
PhysicalVolumePath
```

6. Restart the system in normal mode.

## 2.10 Creating a Snapshot-Based Backup of Oracle Linux Database Server

A backup should be made before and after every significant change to the software on the database server. For example, a backup should be made before and after the following procedures:

- Application of operating system patches
- Application of Oracle patches
- Reconfiguration of significant operating parameters
- Installation or reconfiguration of significant non Oracle software

Starting with Oracle Exadata System Software release 19.1.0, the SSHD `ClientAliveInterval` defaults to 600 seconds. If the time needed for completing backups exceeds 10 minutes, then you can specify a larger value for `ClientAliveInterval` in the `/etc/ssh/sshd_config` file. You must restart the SSH service for changes to take effect. After the long running operation completes, remove the modification to the `ClientAliveInterval` parameter and restart the SSH service.

This section contains the following topics:

- [Creating a Snapshot-Based Backup of Exadata Database Servers X8M and Later with Uncustomized Partitions](#)  
This procedure describes how to take a snapshot-based backup of an Exadata Database Machine X8M and later database server with uncustomized storage partitions.
- [Creating a Snapshot-Based Backup of Exadata X8 or Earlier Database Servers with Uncustomized Partitions](#)  
This procedure describes how to take a snapshot-based backup. The values shown in the procedure are examples.
- [Creating a Snapshot-Based Backup of Oracle Linux Database Server with Customized Partitions](#)

### 2.10.1 Creating a Snapshot-Based Backup of Exadata Database Servers X8M and Later with Uncustomized Partitions

This procedure describes how to take a snapshot-based backup of an Exadata Database Machine X8M and later database server with uncustomized storage partitions.

Starting with Exadata Database Machine X8M and Oracle Exadata System Software release 19.3, the database servers use the following storage partitions:



File System Mount Point	Logical Volume Name
/ (root)	LVDbSys1 or LVDbSys2, whichever is active.
/u01	LVDbOral
/home	LVDbHome
/var	LVDbVar1 or LVDbVar2, whichever is active.
/var/log	LVDbVarLog
/var/log/audit	LVDbVarLogAudit
/tmp	LVDbTmp

 **Note:**

- This procedure relies on the exact storage partitions that are originally shipped on the database server. If you modified the storage partitions in any way, then you cannot use this procedure and the associated recovery procedure without modification. Modifications include changing partition sizes, renaming partitions, adding partitions, or removing partitions.
- All steps must be performed as the `root` user.

**1. Prepare a destination to hold the backup.**

The destination should reside outside of the local machine, such as a writable NFS location, and be large enough to hold the backup `tar` files. For non-customized partitions, the space needed for holding the backup is approximately 145 GB.

You can use the following commands to prepare a backup destination using NFS.

```
mkdir -p /root/remote_FS
mount -t nfs -o rw,intr,soft,proto=tcp,nolock ip_address:/
nfs_location/ /root/remote_FS
```

In the `mount` command, `ip_address` is the IP address of the NFS server, and `nfs_location` is the NFS location holding the backups.

**2. Remove the `LVDotNotRemoveOrUse` logical volume.**

The logical volume `/dev/VGExaDb/LVDotNotRemoveOrUse` is a placeholder to make sure there is always free space available to create a snapshot.

Use the following script to check for the existence of the `LVDotNotRemoveOrUse` logical volume and remove it if present.

```
lvm lvdisplay --ignorelockingfailure /dev/VGExaDb/LVDotNotRemoveOrUse
if [$? -eq 0]; then
 # LVDotNotRemoveOrUse logical volume exists.
 lvm lvremove -f /dev/VGExaDb/LVDotNotRemoveOrUse
 if [$? -ne 0]; then
 echo "Unable to remove logical volume: LVDotNotRemoveOrUse. Do
```

```
not proceed with backup."
 fi
fi
```

If the `LVDNotRemoveOrUse` logical volume does not exist, then do not proceed with the remaining steps and determine the reason.

### 3. Determine the active system volume.

You can use the `imageinfo` command and examine the device hosting the active system partition.

```
imageinfo
```

```
Kernel version: 4.14.35-1902.5.1.4.el7uek.x86_64 #2 SMP Wed Oct 9
19:29:16 PDT 2019 x86_64
Image kernel version: 4.14.35-1902.5.1.4.el7uek
Image version: 19.3.1.0.0.191018
Image activated: 2019-11-04 19:18:32 -0800
Image status: success
Node type: KVMHOST
System partition on device: /dev/mapper/VGExaDb-LVDbSys1
```

In the `imageinfo` output, the system partition device ends with the name of the logical volume supports the active root (`/`) file system. Depending on the system image that is in use, the logical volume name is `LVDbSys1` or `LVDbSys2`. Likewise, the logical volume for the `/var` file system is either `LVDbVar1` or `LVDbVar2`.

You can also confirm the active devices by using the `df -hT` command and examining the output associated with the root (`/`) and `/var` file systems. For example:

```
df -hT
```

Filesystem	Type	Size	Used	Avail	Use%	Mounted on
devtmpfs	devtmpfs	378G	0	378G	0%	/dev
tmpfs	tmpfs	755G	1.0G	754G		
1% /dev/shm						
tmpfs	tmpfs	378G	4.8M	378G	1%	/run
tmpfs	tmpfs	378G	0	378G		
0% /sys/fs/cgroup						
/dev/mapper/VGExaDb-LVDbSys1	xfs	15G	7.7G	7.4G	52%	/
/dev/sda1	xfs	510M	112M	398M	22%	/boot
/dev/sda2	vfat	254M	8.5M	246M	4%	/
boot/efi						
/dev/mapper/VGExaDb-LVDbHome	xfs	4.0G	33M	4.0G	1%	/home
/dev/mapper/VGExaDb-LVDbVar1	xfs	2.0G	139M	1.9G	7%	/var
/dev/mapper/VGExaDb-LVDbVarLog	xfs	18G	403M	18G		
3% /var/log						
/dev/mapper/VGExaDb-LVDbVarLogAudit	xfs	1014M	143M	872M		
15% /var/log/audit						
/dev/mapper/VGExaDb-LVDbTmp	xfs	3.0G	148M	2.9G	5%	/tmp
/dev/mapper/VGExaDb-LVDbOral	xfs	100G	32G	69G	32%	/u01
tmpfs	tmpfs	76G	0	76G	0%	/run/
user/0						

The remaining examples in the procedure use `LVDbSys1` and `LVDbVar1`, which is consistent with the above `imageinfo` and `df` output. However, if the active image is using `LVDbSys2`, then modify the examples in the following steps to use `LVDbSys2` instead of `LVDbSys1`, and `LVDbVar2` instead of `LVDbVar1`.

4. Take snapshots of the logical volumes on the server.

Depending on the active system partition identified in the previous step, remember to use either `LVDbSys1` or `LVDbSys2` to identify the logical volume for the root (`/`) file system, and likewise use either `LVDbVar1` or `LVDbVar2` to identify the logical volume for the `/var` file system.

```
lvcreate -L1G -s -c 32K -n root_snap /dev/VGExaDb/LVDbSys1
lvcreate -L5G -s -c 32K -n u01_snap /dev/VGExaDb/LVDbOra1
lvcreate -L1G -s -c 32K -n home_snap /dev/VGExaDb/LVDbHome
lvcreate -L1G -s -c 32K -n var_snap /dev/VGExaDb/LVDbVar1
lvcreate -L1G -s -c 32K -n varlog_snap /dev/VGExaDb/LVDbVarLog
lvcreate -L1G -s -c 32K -n audit_snap /dev/VGExaDb/LVDbVarLogAudit
lvcreate -L1G -s -c 32K -n tmp_snap /dev/VGExaDb/LVDbTmp
```

5. Label the snapshots.

```
xfs_admin -L DBSYS_SNAP /dev/VGExaDb/root_snap
xfs_admin -L DBORA_SNAP /dev/VGExaDb/u01_snap
xfs_admin -L HOME_SNAP /dev/VGExaDb/home_snap
xfs_admin -L VAR_SNAP /dev/VGExaDb/var_snap
xfs_admin -L VARLOG_SNAP /dev/VGExaDb/varlog_snap
xfs_admin -L AUDIT_SNAP /dev/VGExaDb/audit_snap
xfs_admin -L TMP_SNAP /dev/VGExaDb/tmp_snap
```

6. Mount the snapshots.

Mount all of the snapshots under a common directory location; for example, `/root/mnt`.

```
mkdir -p /root/mnt
mount -t xfs -o nouuid /dev/VGExaDb/root_snap /root/mnt
mkdir -p /root/mnt/u01
mount -t xfs -o nouuid /dev/VGExaDb/u01_snap /root/mnt/u01
mkdir -p /root/mnt/home
mount -t xfs -o nouuid /dev/VGExaDb/home_snap /root/mnt/home
mkdir -p /root/mnt/var
mount -t xfs -o nouuid /dev/VGExaDb/var_snap /root/mnt/var
mkdir -p /root/mnt/var/log
mount -t xfs -o nouuid /dev/VGExaDb/varlog_snap /root/mnt/var/log
mkdir -p /root/mnt/var/log/audit
mount -t xfs -o nouuid /dev/VGExaDb/audit_snap /root/mnt/var/log/audit
mkdir -p /root/mnt/tmp
mount -t xfs -o nouuid /dev/VGExaDb/tmp_snap /root/mnt/tmp
```

7. Back up the snapshots.

Use the following commands to write a backup file as a compressed tar file to your prepared NFS backup destination.

```
cd /root/mnt
tar --acls --xattrs --xattrs-include=* --format=pax -pjcvf /root/
remote_FS/mybackup.tar.bz2 * /boot > /tmp/backup_tar.stdout 2> /tmp/
backup_tar.stderr
```

8. Check the `/tmp/backup_tar.stderr` file for any significant errors. Errors about failing to tar open sockets, and other similar errors, can be ignored.
9. Unmount and remove all of the snapshots.

```
cd /
umount /root/mnt/tmp
umount /root/mnt/var/log/audit
umount /root/mnt/var/log
umount /root/mnt/var
umount /root/mnt/home
umount /root/mnt/u01
umount /root/mnt
lvremove /dev/VGExaDb/tmp_snap
lvremove /dev/VGExaDb/audit_snap
lvremove /dev/VGExaDb/varlog_snap
lvremove /dev/VGExaDb/var_snap
lvremove /dev/VGExaDb/home_snap
lvremove /dev/VGExaDb/u01_snap
lvremove /dev/VGExaDb/root_snap
```

10. Unmount the NFS backup destination.

```
umount /root/remote_FS
```

11. Remove the mount point directories that you created during this procedure.

```
rm -r /root/mnt
rmdir /root/remote_FS
```

12. Recreate the `/dev/VGExaDb/LVDoNotRemoveOrUse` logical volume.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L2G VGExaDb -y
```

## 2.10.2 Creating a Snapshot-Based Backup of Exadata X8 or Earlier Database Servers with Uncustomized Partitions

This procedure describes how to take a snapshot-based backup. The values shown in the procedure are examples.

If you have not customized the database server partitions from their original shipped configuration, then use the procedures in this section to take a backup and use the backup to restore the database server using the backup.

 **Note:**

- The recovery procedure restores the exact partitions, including the name and sizes, as they were originally shipped. If you modified the partitions in any way, then you cannot use this procedure. Modifications include changing sizes, renaming, addition or removal of partitions.
- All steps must be performed as the `root` user.

**1. Prepare a destination to hold the backup.**

The destination can be a large, writable NFS location. The NFS location should be large enough to hold the backup tar files. For uncustomized partitions, 145 GB should be adequate.

**a. Create a mount point for the NFS share.**

```
mkdir -p /root/tar
```

**b. Mount the NFS location.**

In the following command, *ip\_address* is the IP address of the NFS server, and *nfs\_location* is the NFS location.

```
mount -t nfs -o rw,intr,soft,proto=tcp,nolock
ip_address:/nfs_location/ /root/tar
```

**2. Take a snapshot-based backup of the / (root) and /u01 directories.****a. Create a snapshot named `root_snap` for the root directory.**

`LVDbSys1` is used in the example below, but you should use the value based on the output of `imageinfo`. If the active image is on `LVDbSys2`, then the command would be: `lvcreate -L1G -s -c 32K -n root_snap /dev/VGExaDb/LVDbSys2`.

```
lvcreate -L1G -s -c 32K -n root_snap /dev/VGExaDb/LVDbSys1
```

**b. Label the snapshot.**

```
e2label /dev/VGExaDb/root_snap DBSYS_SNAP
```

**c. Determine the file system type of the / (root) and /u01 directories.**

- InfiniBand Network Fabric-based servers running Oracle Exadata System Software release 12.1.2.1.0 or later use the `ext4` file system type.
- InfiniBand Network Fabric-based servers running a release of Oracle Exadata System Software earlier than 12.1.2.1.0 use the `ext3` file system type.

- Exadata X5 servers or earlier server models that were updated to Oracle Exadata System Software release 12.1.2.1.0 or later also use `ext3` file system type.

```
mount -l
/dev/mapper/VGExaDb-LVDbSys1 on / type ext4 (rw) [DBSYS]
...
```

**d. Mount the snapshot.**

In the `mount` command below, *filesystem\_type\_of\_root\_directory* is a placeholder for the file system type as determined in the previous step.

```
mkdir /root/mnt
mount /dev/VGExaDb/root_snap /root/mnt -t
filesystem_type_of_root_directory
```

**e. Create a snapshot named `u01_snap` for the `/u01` directory.**

```
lvcreate -L5G -s -c 32K -n u01_snap /dev/VGExaDb/LVDbOra1
```

**f. Label the snapshot.**

```
e2label /dev/VGExaDb/u01_snap DBORA_SNAP
```

**g. Mount the snapshot.**

In the `mount` command below, *filesystem\_type\_of\_u01\_directory* is a placeholder for the file system type as determined in step 2.c above.

```
mkdir -p /root/mnt/u01
mount /dev/VGExaDb/u01_snap /root/mnt/u01 -t
filesystem_type_of_u01_directory
```

**h. Change to the directory for the backup.**

```
cd /root/mnt
```

**i. Create the backup file using one of the following commands:**

- System does not have NFS mount points:

```
tar -pjcvf /root/tar/mybackup.tar.bz2 * /boot --exclude \
tar/mybackup.tar.bz2 > /tmp/backup_tar.stdout 2> /tmp/
backup_tar.stderr
```

- System has NFS mount points:

In the following command, *nfs\_mount\_points* are the NFS mount points. Excluding the mount points prevents the generation of large files and long backup times.

```
tar -pjcvf /root/tar/mybackup.tar.bz2 * /boot --exclude \
tar/mybackup.tar.bz2 --exclude nfs_mount_points > \
/tmp/backup_tar.stdout 2> /tmp/backup_tar.stderr
```

- j. Check the `/tmp/backup_tar.stderr` file for any significant errors.

Errors about failing to tar open sockets, and other similar errors, can be ignored.

3. Unmount the snapshots and remove the snapshots for the `/` (root) and `/u01` directories.

```
cd /
umount /root/mnt/u01
umount /root/mnt
/bin/rm -rf /root/mnt
lvremove /dev/VGExaDb/u01_snap
lvremove /dev/VGExaDb/root_snap
```

4. Unmount the NFS share.

```
umount /root/tar
```

## 2.10.3 Creating a Snapshot-Based Backup of Oracle Linux Database Server with Customized Partitions

When you have customized the partitions, the backup procedure is generally the same as the procedure used for non-customized database servers, with the following alterations:

- You must add the commands to back up any additional partitions. Throughout the procedure, use the command relating to the `/u01` partition as a template, and modify the arguments to suit.
- If any partitions are altered, then use the modified attributes in your commands. For example, if `/u01` is renamed to `/myown_u01`, then use `/myown_u01` in the commands.

### Related Topics

- [Creating a Snapshot-Based Backup of Exadata Database Servers X8M and Later with Uncustomized Partitions](#)  
This procedure describes how to take a snapshot-based backup of an Exadata Database Machine X8M and later database server with uncustomized storage partitions.
- [Creating a Snapshot-Based Backup of Exadata X8 or Earlier Database Servers with Uncustomized Partitions](#)  
This procedure describes how to take a snapshot-based backup. The values shown in the procedure are examples.

## 2.11 Recovering Oracle Linux Database Servers Using a Snapshot-Based Backup

You can recover a database server file systems running Oracle Linux using a snapshot-based backup after severe disaster conditions happen for the database server, or when the server hardware is replaced to such an extent that it amounts to new hardware.

For example, replacing all hard disks leaves no trace of original software on the system. This is similar to replacing the complete system as far as the software is concerned. In addition, it provides a method for disaster recovery of the database servers using an LVM snapshot-based backup taken when the database server was healthy before the disaster condition.

The recovery procedures described in this section do not include backup or recovery of storage servers or the data within the Oracle databases. Oracle recommends testing the backup and recovery procedures on a regular basis.

- [Overview of Snapshot-Based Recovery of Database Servers](#)  
The recovery process consists of a series of tasks.
- [Recovering Oracle Linux Database Server with Uncustomized Partitions](#)  
You can recover the Oracle Linux database server from a snapshot-based backup when using uncustomized partitions.
- [Recovering Exadata X9M Database Servers with Customized Partitions](#)  
This procedure describes how to recover an Exadata Database Machine X9M database server with RoCE Network Fabric from a snapshot-based backup when using customized partitions.
- [Recovering Exadata X8M Database Servers with Customized Partitions](#)  
This procedure describes how to recover an Exadata Database Machine X8M database server with RoCE Network Fabric from a snapshot-based backup when using customized partitions.
- [Recovering Exadata Database Servers X7 or X8 with Customized Partitions](#)  
This procedure describes how to recover an Exadata Database Machine X7 or X8 Oracle Linux database server with InfiniBand Network Fabric from a snapshot-based backup when using customized partitions.
- [Recovering Exadata X6 or Earlier Database Servers with Customized Partitions](#)  
This procedure describes how to recover Oracle Exadata Database Servers for Exadata Database Machine X6-2 or earlier running Oracle Linux from a snapshot-based backup when using customized partitions.
- [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#)

## 2.11.1 Overview of Snapshot-Based Recovery of Database Servers

The recovery process consists of a series of tasks.

The recovery procedures use the `diagnostics.iso` image as a virtual CD-ROM to restart the database server in rescue mode using the ILOM.



### Note:

Restoring files from tape may require additional drives to be loaded, and is not covered in this chapter. Oracle recommends backing up files to an NFS location, and using existing tape options to back up and recover from the NFS host.

The general work flow includes the following tasks:

1. Recreate the following:
  - Boot partitions



- Physical volumes
  - Volume groups
  - Logical volumes
  - File system
  - Swap partition
2. Activate the swap partition.
  3. Ensure the `/boot` partition is the active boot partition.
  4. Restore the data.
  5. Reconfigure GRUB.
  6. Restart the server.

If you use quorum disks, then after recovering the database servers from backup, you must manually reconfigure the quorum disk for the recovered server. See [Reconfigure Quorum Disk After Restoring a Database Server](#) for more information.

## 2.11.2 Recovering Oracle Linux Database Server with Uncustomized Partitions

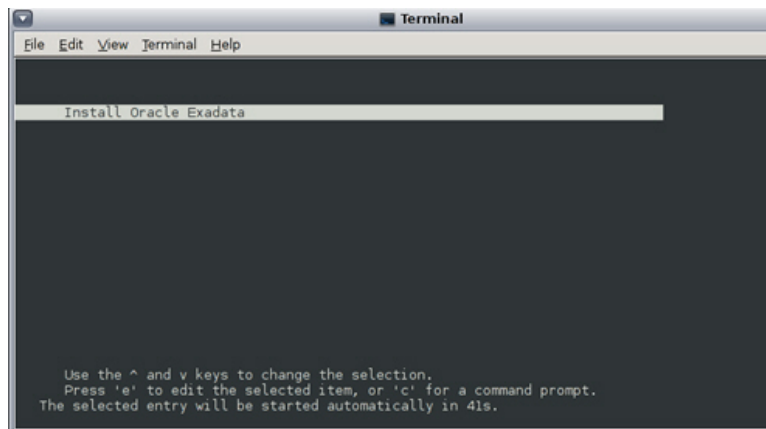
You can recover the Oracle Linux database server from a snapshot-based backup when using uncustomized partitions.

This procedure is applicable when the layout of the partitions, logical volumes, file systems, and their sizes are equal to the layout when the database server was initially deployed.

### **Caution:**

All existing data on the disks is lost during the procedure.

1. Prepare an NFS server to host the backup archive file (`mybackup.tar.bz2`).  
The NFS server must be accessible by IP address.  
For example, on an NFS server with the IP address `nfs_ip`, where the directory `/export` is exported as an NFS mount, put the backup file (`mybackup.tar.bz2`) in the `/export` directory.
2. Restart the recovery target system using the `diagnostics.iso` file.  
See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.
3. Answer as indicated in these examples when prompted by the system. The responses are shown in bold.  
Note that for Oracle Exadata System Software release 12.1.2.2.0 or later, DHCP is used and you do not have to manually set up the network.
  - If you are using Oracle Exadata System Software release 18.1 or later, running on Oracle Exadata Database Machine X7 or later, then the prompt looks like the following:



- If you are using Oracle Exadata System Software release 18.1 or later and restoring through one of the 10GbE Ethernet SFP+ ports on Oracle Exadata Database Machine X3-2 or later, then the prompt looks like the following:

```


 Choose from the following by typing letter in '()':
 (e)nter interactive diagnostics shell.
 Use diagnostics shell password to login as root user
 (reboot or power cycle to exit the shell),
 (r)estore system from NFS backup archive,
Select: r
Continue (y/n) [n]: y
Rescue password:
[INFO] Enter path to the backup file on the NFS server in
format:
 Enter path to the backup file on the NFS server in format:
 <ip_address_of_the_NFS_share>:/<path>/<archive_file>
 For example, 10.124.1.15:/export/nfs/share/
backup.2010.04.08.tar.bz2
NFS line: <nfs_ip>:/export/mybackup.tar.bz2
[INFO] The backup file could be created either from LVM or non-
LVM
based COMPUTE node
[INFO] Versions below 11.2.1.3.0 do not support LVM based
partitioning
Use LVM based scheme. (y/n) [y]: y
Configure network settings on host via DHCP. (y/n) [y]: n
Configure bonded network interface. (y/n) [y]: y
IP Address of bondeth0 on this host: <IP address of the DB host>

Netmask of bondeth0 on this host: <netmask for the above IP address>
Bonding mode:active-backup or 802.3ad [802.3ad]: active-backup
Slave interface1 for bondeth0 (ethX) [eth4]: eth4
Slave interface2 for bondeth0 (ethX) [eth5]: eth5
...
[354.619610] bondeth0: first active interface up!
[354.661427] ixgbe 0000:13:00.1 eth5: NIC Link is Up 10 Gbps, Flow
Control: RX/TX

```

```
[354.724414] bondeth0: link status definitely up for
interface eth5, 10000 Mbps full duplex
Default gateway: <Gateway for the above IP address>
```

- If you are using Oracle Exadata System Software release 12.1.x or 12.2.x, then the prompts look like the following:

```


Use diagnostics shell password to login as root user
(reboot or power cycle to exit the shell),
(r)estore system from NFS backup archive.
Select: r
Continue (y/n) [n]: y
Rescue password:
[INFO:] Enter path to the backup file on the NFS server in
format:
Enter path to the backup file on the NFS server in format:
<ip_address_of_the_NFS_share>:/<path>/<archive_file>
For example, 10.124.1.15:/export/nfs/share/
backup.2010.04.08.tar.bz2
NFS line: <nfs_ip>:/export/mybackup.tar.bz2
[INFO:] The backup file could be created either from LVM or non-
LVM based COMPUTE node
[INFO:] Versions below 11.2.1.3.0 do not support LVM based
partitioning
Use LVM based scheme. (y/n) [y]: y
```

- If you are using Oracle Exadata System Software release earlier than 12.1.2.2.0, then the prompts look like the following

```


Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle
support to login (reboot or power cycle to exit the shell),
(r)estore system from NFS backup archive,
Select: r
Are you sure (y/n) [n]: y

The backup file could be created either from LVM or non-LVM
based compute node
versions below 11.2.1.3.1 and 11.2.2.1.0 or higher do not
support LVM based partitioning
use LVM based scheme(y/n) : y

Enter path to the backup file on the NFS server in format:
ip_address_of_the_NFS_share:/path/archive_file
For example, 10.10.10.10:/export/operating_system.tar.bz2
NFS line: <nfs_ip>:/export/mybackup.tar.bz2
```

```

IP Address of this host:IP address of the DB host
Netmask of this host:netmask for the above IP address
Default gateway:Gateway for the above IP address. If there is no
default gateway in your network, enter 0.0.0.0.


```

When the recovery completes, the log in screen appears.

4. Log in as the `root` user.

If you do not have the password for the `root` user, then contact Oracle Support Services.

5. Restart the system.

```
shutdown -r now
```

The restoration process is complete.

6. Verify that all Oracle software can start and function by logging in to the database server.

The `/usr/local/bin/imagehistory` command indicates that the database server was reconstructed.

The following is an example of the output:

```

imagehistory

Version : 11.2.2.1.0
Image activation date : 2010-10-13 13:42:58 -0700
Imaging mode : fresh
Imaging status : success

Version : 11.2.2.1.0
Image activation date : 2010-10-30 19:41:18 -0700
Imaging mode : restore from nfs backup
Imaging status : success

```

7. On systems with InfiniBand Network Fabric only, run `reclaimdisks.sh` on the restored database server.

```
/opt/oracle.SupportTools/reclaimdisks.sh -free -reclaim
```

 **Note:**

This command is not required on RoCE-based Exadata database servers.

8. If the recovery was on Oracle Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

## 2.11.3 Recovering Exadata X9M Database Servers with Customized Partitions

This procedure describes how to recover an Exadata Database Machine X9M database server with RoCE Network Fabric from a snapshot-based backup when using customized partitions.

1. Prepare an NFS server to host the backup archive file (`mybackup.tar.bz2`).

The NFS server must be accessible by IP address.

For example, on an NFS server with the IP address `nfs_ip`, where the directory `/export` is exported as an NFS mount, put the backup file (`mybackup.tar.bz2`) in the `/export` directory.

2. Restart the recovery target system using the `diagnostics.iso` file.

See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.

3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. If it is mounted, unmount `/mnt/cell`

```
umount /mnt/cell
```

5. Confirm the `md` devices on the server.

Confirm that the server contains the devices listed in the following example. Do not proceed and contact Oracle Support if your server differs substantially.

```
ls -al /dev/md*
brw-rw---- 1 root disk 9, 126 Jul 15 06:59 /dev/md126
brw-rw---- 1 root disk 259, 4 Jul 15 06:59 /dev/md126p1
brw-rw---- 1 root disk 259, 5 Jul 15 06:59 /dev/md126p2
brw-rw---- 1 root disk 9, 127 Jul 15 06:28 /dev/md127
brw-rw---- 1 root disk 9, 25 Jul 15 06:28 /dev/md25

/dev/md:
total 0
drwxr-xr-x 2 root root 140 Jul 15 06:59 .
drwxr-xr-x 18 root root 3400 Jul 15 06:59 ..
```

```
lrwxrwxrwx 1 root root 8 Jul 15 06:59 24_0 -> ../md126
lrwxrwxrwx 1 root root 10 Jul 15 06:59 24_0p1 -> ../md126p1
lrwxrwxrwx 1 root root 10 Jul 15 06:59 24_0p2 -> ../md126p2
lrwxrwxrwx 1 root root 7 Jul 15 06:28 25 -> ../md25
lrwxrwxrwx 1 root root 8 Jul 15 06:28 imsm0 -> ../md127
```

6. Create the boot partition.

- a. Start an interactive session using the `parted` command.

```
parted /dev/md126
```

- b. Assign a disk label.

```
(parted) mklabel gpt
```

- c. Set the unit size as sector.

```
(parted) unit s
```

- d. Check the partition table by displaying the existing partitions.

```
(parted) print
```

- e. Remove the partitions listed in the previous step.

```
(parted) rm part#
```

- f. Create a new first partition.

```
(parted) mkpart primary 64s 15114206s
```

- g. Specify this is a bootable partition.

```
(parted) set 1 boot on
```

7. Create the second primary (boot) partition.

- a. Create a second primary partition as a UEFI boot partition with `fat32`.

```
(parted) mkpart primary fat32 15114207s 15638494s
```

```
(parted) set 2 boot on
```

- b. Write the information to disk, then quit.

```
(parted) quit
```

8. Create the physical volume and volume group.

```
lvm pvcreate /dev/md25
```

```
lvm vgcreate VGExaDb /dev/md25
```

If the physical volume or volume group already exists, then remove and then re-create them as follows:

```
lvm vgremove VGExaDb
lvm pvremove /dev/md25
lvm pvcreate /dev/md25
lvm vgcreate VGExaDb /dev/md25
```

9. Re-create the customized LVM partitions, then create and mount the file systems.

 **Note:**

Use the following information and examples as guidance for this step. You must make the necessary adjustments for customized LVM partitions and file systems. For example, you may need to adjust the names and sizes of various partitions to match your previous customizations, or you may need to create additional custom partitions.

- a. Create the logical volumes.

For example, the following commands re-create the logical volumes that exist by default on Exadata Database Machine X9M systems:

```
lvm lvcreate -n LVDbSys1 -L15G VGExaDb
lvm lvcreate -n LVDbSys2 -L15G VGExaDb
lvm lvcreate -n LVDbOra1 -L200G VGExaDb
lvm lvcreate -n LVDbHome -L4G VGExaDb
lvm lvcreate -n LVDbVar1 -L2G VGExaDb
lvm lvcreate -n LVDbVar2 -L2G VGExaDb
lvm lvcreate -n LVDbVarLog -L18G VGExaDb
lvm lvcreate -n LVDbVarLogAudit -L1G VGExaDb
lvm lvcreate -n LVDbTmp -L3G VGExaDb
```

- b. Create the file systems.

```
mkfs.xfs -f /dev/VGExaDb/LVDbSys1
mkfs.xfs -f /dev/VGExaDb/LVDbSys2
mkfs.xfs -f /dev/VGExaDb/LVDbOra1
mkfs.xfs -f /dev/VGExaDb/LVDbHome
mkfs.xfs -f /dev/VGExaDb/LVDbVar1
mkfs.xfs -f /dev/VGExaDb/LVDbVar2
mkfs.xfs -f /dev/VGExaDb/LVDbVarLog
mkfs.xfs -f /dev/VGExaDb/LVDbVarLogAudit
mkfs.xfs -f /dev/VGExaDb/LVDbTmp
mkfs.xfs -f /dev/md126p1
```

- c. Label the file systems.

```
xfs_admin -L DBSYS /dev/VGExaDb/LVDbSys1
xfs_admin -L DBORA /dev/VGExaDb/LVDbOra1
xfs_admin -L HOME /dev/VGExaDb/LVDbHome
xfs_admin -L VAR /dev/VGExaDb/LVDbVar1
xfs_admin -L DIAG /dev/VGExaDb/LVDbVarLog
```

```
xfs_admin -L AUDIT /dev/VGExaDb/LVDbVarLogAudit
xfs_admin -L TMP /dev/VGExaDb/LVDbTmp
xfs_admin -L BOOT /dev/md126p1
```

- d. Create mount points for all the partitions to mirror the original system, and mount the respective partitions.

For example, assuming that `/mnt` is used as the top level directory for the recovery operation, you could use the following commands to create the directories and mount the partitions:

```
mount -t xfs /dev/VGExaDb/LVDbSys1 /mnt
mkdir -p /mnt/u01
mount -t xfs /dev/VGExaDb/LVDbOra1 /mnt/u01
mkdir -p /mnt/home
mount -t xfs /dev/VGExaDb/LVDbHome /mnt/home
mkdir -p /mnt/var
mount -t xfs /dev/VGExaDb/LVDbVar1 /mnt/var
mkdir -p /mnt/var/log
mount -t xfs /dev/VGExaDb/LVDbVarLog /mnt/var/log
mkdir -p /mnt/var/log/audit
mount -t xfs /dev/VGExaDb/LVDbVarLogAudit /mnt/var/log/audit
mkdir -p /mnt/tmp
mount -t xfs /dev/VGExaDb/LVDbTmp /mnt/tmp
mkdir -p /mnt/boot
mount -t xfs /dev/md126p1 /mnt/boot
```

10. Create the system swap space.

For Exadata Database Machine X9M the default swap size is 16 GB.

For example:

```
lvcreate -n LVDbSwap1 -L16G VGExaDb
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

11. Create `/mnt/boot/efi`, label `/dev/md126p2`, and mount `/dev/md126p2` on `/mnt/boot/efi` with type `vfat`.

```
mkdir /mnt/boot/efi
dosfslabel /dev/md126p2 ESP
mount /dev/md126p2 /mnt/boot/efi -t vfat
```

12. Bring up the network.

```
ip address add ip_address_for_eth0/netmask_for_eth0 dev eth0
ip link set up eth0
ip route add default via gateway_address dev eth0
```

13. Mount the NFS server where you have the backup.



The following example assumes that the backup is located in the `/export` directory of the NFS server with IP address `nfs_ip`.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/export /
root/mnt
```

**14.** Restore from backup.

```
tar --acls --xattrs --xattrs-include=* --format=pax -pjxvf /
root/mnt/mybackup.tar.bz2 -C /mnt
```

**15.** Unmount the restored file systems.

For example:

```
umount /mnt/tmp
umount /mnt/var/log/audit
umount /mnt/var/log
umount /mnt/var
umount /mnt/home
umount /mnt/u01
umount /mnt/boot/efi
umount /mnt/boot
umount /mnt
```

**16.** Check the boot devices and set the boot order.

- a. Check the available boot devices, and identify the boot device that is associated with `\EFI\REDHAT\SHIMX64.EFI`.

For example:

```
efibootmgr -v
BootCurrent: 000C
Timeout: 1 seconds
BootOrder:
000C,0000,0001,0008,0009,0004,0005,0006,0007,0002,0003,000D
Boot0000* RedHat Boot Manager HD(2,GPT,9dd64f77-490d-4ece-8aed-
bc4d9e82d10f,0xe69fdf,0x80000)/File(\EFI\REDHAT\SHIMX64.EFI)
Boot0001* NET0:PXE IPv4 Intel(R) I210 Gigabit Network
Connection /Pci(0x1c,0x4)/Pci(0x0,0x0)/MAC(0010e0fc6e94,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot0002* PCIE5:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:22:38:0A /Pci(0x2,0x0)/Pci(0x0,0x0)/MAC(b8cef622380a,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot0003* PCIE5:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:22:38:0B /Pci(0x2,0x0)/Pci(0x0,0x1)/MAC(b8cef622380b,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot0004* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci(0x2,0x0)/Pci(0x0,0x0)/MAC(3cfdfe915070,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot0005* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci(0x2,0x0)/Pci(0x0,0x1)/MAC(3cfdfe915071,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot0006* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
```

```
Pci(0x2,0x0)/Pci(0x0,0x2)/MAC(3cfdfe915072,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot0007* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci(0x2,0x0)/Pci(0x0,0x3)/MAC(3cfdfe915073,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot0008* PCIE1:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:44:51:9C /Pci(0x2,0x0)/Pci(0x0,0x0)/MAC(b8cef644519c,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot0009* PCIE1:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:44:51:9D /Pci(0x2,0x0)/Pci(0x0,0x1)/MAC(b8cef644519d,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
Boot000C* USB:SP:SUN Remote ISO CDROM1.01 /Pci(0x14,0x0)/USB(7,0)/
USB(3,0)/CDROM(1,0x28,0x3100)..BO
Boot000D* Oracle Linux (grubx64.efi) HD(2,GPT,9dd64f77-490d-4ece-8aed-
bc4d9e82d10f,0xe69fdf,0x80000)/File(\EFI\REDHAT\GRUBX64.EFI)..BO
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false
```

- b.** Configure the device that is associated with `\EFI\REDHAT\SHIMX64.EFI` to be first in the boot order.

For example:

```
efibootmgr -o 0000
BootCurrent: 000C
Timeout: 1 seconds
BootOrder: 0000
Boot0000* RedHat Boot Manager
Boot0001* NET0:PXE IPv4 Intel(R) I210 Gigabit Network Connection
Boot0002* PCIE5:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:22:38:0A
Boot0003* PCIE5:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:22:38:0B
Boot0004* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0005* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0006* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0007* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE1:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:44:51:9C
Boot0009* PCIE1:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:44:51:9D
Boot000C* USB:SP:SUN Remote ISO CDROM1.01
Boot000D* Oracle Linux (grubx64.efi)
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false
```

- 17.** Restart the system.

```
reboot
```

- 18.** Disconnect the `diagnostics.iso` file.

See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.

- 19.** Log back into the server as the `root` user.

- 20.** Run the `imageinfo` command and verify that the image status is `success`.

For example:

```
imageinfo

Kernel version: 4.14.35-2047.502.5.el7uek.x86_64 #2 SMP Wed Apr 14
15:08:41
PDT 2021 x86_64
Uptrack kernel version: 4.14.35-2047.503.1.el7uek.x86_64 #2 SMP Fri
Apr 23
15:20:41 PDT 2021 x86_64
Image kernel version: 4.14.35-2047.502.5.el7uek
Image version: 21.2.1.0.0.210608
Image activated: 2021-07-12 14:58:03 +0900
Image status: success
Node type: COMPUTE
System partition on device: /dev/mapper/VGExaDb-LVDbSys1
```

21. If the recovery was on Oracle Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

## 2.11.4 Recovering Exadata X8M Database Servers with Customized Partitions

This procedure describes how to recover an Exadata Database Machine X8M database server with RoCE Network Fabric from a snapshot-based backup when using customized partitions.

1. Prepare an NFS server to host the backup archive file (`mybackup.tar.bz2`).

The NFS server must be accessible by IP address.

For example, on an NFS server with the IP address `nfs_ip`, where the directory `/export` is exported as an NFS mount, put the backup file (`mybackup.tar.bz2`) in the `/export` directory.

2. Restart the recovery target system using the `diagnostics.iso` file.

See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.

3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. If required, use `/opt/MegaRAID/storcli/storcli64` to configure the disk controller to set up the disks.

5. If it is mounted, unmount `/mnt/cell`

```
umount /mnt/cell
```

6. Create the boot partition.

- a. Start an interactive session using the `parted` command.

```
parted /dev/sda
```

- b. Assign a disk label.

```
(parted) mklabel gpt
```

- c. Set the unit size as sector.

```
(parted) unit s
```

- d. Check the partition table by displaying the existing partitions.

```
(parted) print
```

- e. Remove the partitions listed in the previous step.

```
(parted) rm part#
```

- f. Create a new first partition.

```
(parted) mkpart primary 64s 1048639s
```

- g. Specify this is a bootable partition.

```
(parted) set 1 boot on
```

7. Create second primary (boot) and third primary (LVM) partitions.

- a. Create a second primary partition as a UEFI boot partition with `fat32`.

```
(parted) mkpart primary fat32 1048640s 1572927s
(parted) set 2 boot on
```

- b. Create a new third partition.

```
(parted) mkpart primary 1572928s -1s
```

- c. Configure the third partition as a physical volume.

```
(parted) set 3 lvm on
```

- d. Write the information to disk, then quit.

```
(parted) quit
```

8. Create the physical volume and volume group.

```
lvm pvcreate /dev/sda3
lvm vgcreate VGExaDb /dev/sda3
```

If the physical volume or volume group already exists, then remove and then re-create them as follows:.

```
lvm vgremove VGExaDb
lvm pvremove /dev/sda3
lvm pvcreate /dev/sda3
lvm vgcreate VGExaDb /dev/sda3
```

9. Re-create the customized LVM partitions, then create and mount the file systems.

 **Note:**

Use the following information and examples as guidance for this step. You must make the necessary adjustments for customized LVM partitions and file systems. For example, you may need to adjust the names and sizes of various partitions to match your previous customizations, or you may need to create additional custom partitions.

- a. Create the logical volumes.

For example, the following commands re-create the logical volumes that exist by default on Exadata Database Machine X8M systems with Oracle Exadata System Software release 19.3 or later:

```
lvm lvcreate -n LVDbSys1 -L15G VGExaDb
lvm lvcreate -n LVDbSys2 -L15G VGExaDb
lvm lvcreate -n LVDbOra1 -L100G VGExaDb
lvm lvcreate -n LVDbHome -L4G VGExaDb
lvm lvcreate -n LVDbVar1 -L2G VGExaDb
lvm lvcreate -n LVDbVar2 -L2G VGExaDb
lvm lvcreate -n LVDbVarLog -L18G VGExaDb
lvm lvcreate -n LVDbVarLogAudit -L1G VGExaDb
lvm lvcreate -n LVDbTmp -L3G VGExaDb
```

- b. Create the file systems.

```
mkfs.xfs -f /dev/VGExaDb/LVDbSys1
mkfs.xfs -f /dev/VGExaDb/LVDbSys2
mkfs.xfs -f /dev/VGExaDb/LVDbOra1
mkfs.xfs -f /dev/VGExaDb/LVDbHome
mkfs.xfs -f /dev/VGExaDb/LVDbVar1
mkfs.xfs -f /dev/VGExaDb/LVDbVar2
mkfs.xfs -f /dev/VGExaDb/LVDbVarLog
```

```
mkfs.xfs -f /dev/VGExaDb/LVDbVarLogAudit
mkfs.xfs -f /dev/VGExaDb/LVDbTmp
mkfs.xfs -f /dev/sda1
```

**c.** Label the file systems.

```
xfs_admin -L DBSYS /dev/VGExaDb/LVDbSys1
xfs_admin -L DBORA /dev/VGExaDb/LVDbOra1
xfs_admin -L HOME /dev/VGExaDb/LVDbHome
xfs_admin -L VAR /dev/VGExaDb/LVDbVar1
xfs_admin -L DIAG /dev/VGExaDb/LVDbVarLog
xfs_admin -L AUDIT /dev/VGExaDb/LVDbVarLogAudit
xfs_admin -L TMP /dev/VGExaDb/LVDbTmp
xfs_admin -L BOOT /dev/sda1
```

**d.** Create mount points for all the partitions to mirror the original system, and mount the respective partitions.

For example, assuming that `/mnt` is used as the top level directory for the recovery operation, you could use the following commands to create the directories and mount the partitions:

```
mount -t xfs /dev/VGExaDb/LVDbSys1 /mnt
mkdir -p /mnt/u01
mount -t xfs /dev/VGExaDb/LVDbOra1 /mnt/u01
mkdir -p /mnt/home
mount -t xfs /dev/VGExaDb/LVDbHome /mnt/home
mkdir -p /mnt/var
mount -t xfs /dev/VGExaDb/LVDbVar1 /mnt/var
mkdir -p /mnt/var/log
mount -t xfs /dev/VGExaDb/LVDbVarLog /mnt/var/log
mkdir -p /mnt/var/log/audit
mount -t xfs /dev/VGExaDb/LVDbVarLogAudit /mnt/var/log/audit
mkdir -p /mnt/tmp
mount -t xfs /dev/VGExaDb/LVDbTmp /mnt/tmp
mkdir -p /mnt/boot
mount -t xfs /dev/sda1 /mnt/boot
```

**10.** Create the system swap space.

For Exadata Database Machine X8M, with Oracle Exadata System Software release 19.3 or later, the default swap size is 16 GB.

For example:

```
lvm lvcreate -n LVDbSwap1 -L16G VGExaDb
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

**11.** Create `/mnt/boot/efi`, label `/dev/sda2`, and mount `/dev/sda2` on `/mnt/boot/efi` with type `vfat`.

```
mkdir /mnt/boot/efi
dosfslabel /dev/sda2 ESP
mount /dev/sda2 /mnt/boot/efi -t vfat
```

**12. Bring up the network.**

```
ip address add ip_address_for_eth0/netmask_for_eth0 dev eth0
ip link set up eth0
ip route add default via gateway_address dev eth0
```

**13. Mount the NFS server where you have the backup.**

The following example assumes that the backup is located in the `/export` directory of the NFS server with IP address `nfs_ip`.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/export /
root/mnt
```

**14. Restore from backup.**

```
tar --acls --xattrs --xattrs-include=* --format=pax -pjxvf /
root/mnt/mybackup.tar.bz2 -C /mnt
```

**15. Unmount the restored file systems.**

For example:

```
umount /mnt/boot/efi
umount /mnt/boot
umount /mnt/tmp
umount /mnt/var/log/audit
umount /mnt/var/log
umount /mnt/var
umount /mnt/home
umount /mnt/u01
umount /mnt
```

**16. Detach the `diagnostics.iso` file.**

**17. Check the boot devices and boot order for the `ExadataLinux_1` device.**

**a. Check the available boot devices.**

```
efibootmgr
BootCurrent: 000C
Timeout: 1 seconds
BootOrder: 000C,0001,0002,0003,0004,0005,0007,0008,0009,000A,000B
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet
Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet
Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
```

```

Boot000B* Oracle Linux
Boot000C* USB:SUN

```

If the `Boot0000* ExadataLinux_1` device is not listed then create the device.

```

efibootmgr -c -d /dev/sda -p 2 -l '\EFI\REDHAT\SHIM.EFI' -L
'ExadataLinux_1'
BootCurrent: 000C
Timeout: 1 seconds
BootOrder: 0000,000C,0001,0002,0003,0004,0005,0007,0008,0009,000A,000B
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000B* Oracle Linux
Boot000C* USB:SUN
Boot0000* ExadataLinux_1

```

- b. Configure the `Boot0000* ExadataLinux_1` device to be first in the boot order.

```

efibootmgr -o 0000
BootCurrent: 000B
Timeout: 1 seconds
BootOrder: 0000
Boot0000* ExadataLinux_1
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000B* USB:SUN
Boot000C* UEFI OS

```

18. Restart the system and update the boot order in the BIOS.

```
reboot
```

Modify the boot order to set the `ExadataLinux_1` boot device as the first device.

- a. Press F2 when booting the system.
- b. Go to the Setup Utility.
- c. Select **BOOT**.
- d. Set `ExadataLinux_1` for Boot Option #1.



- e. Exit the Setup Utility.

This completes the restoration procedure for the server.

19. If the recovery was on Oracle Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

## 2.11.5 Recovering Exadata Database Servers X7 or X8 with Customized Partitions

This procedure describes how to recover an Exadata Database Machine X7 or X8 Oracle Linux database server with InfiniBand Network Fabric from a snapshot-based backup when using customized partitions.

### Note:

This task assumes you are running Oracle Exadata System Software release 18c (18.1.0) or greater.

1. Prepare an NFS server to host the backup archive file (`mybackup.tar.bz2`).  
The NFS server must be accessible by IP address.  
For example, on an NFS server with the IP address `nfs_ip`, where the directory `/export` is exported as an NFS mount, put the backup file (`mybackup.tar.bz2`) in the `/export` directory.
2. Restart the recovery target system using the `diagnostics.iso` file.  
See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.
3. Log in to the diagnostics shell as the `root` user.  
When prompted, enter the diagnostics shell.  
For example:  

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

  
If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.
4. If required, use `/opt/MegaRAID/storcli/storcli64` (or `/opt/MegaRAID/MegaCli/MegaCli64` for releases earlier than Oracle Exadata System Software 19c) to configure the disk controller to set up the disks.

5. If it is mounted, unmount `/mnt/cell`

```
umount /mnt/cell
```

6. Create the boot partition.

- a. Start an interactive session using the `partd` command.

```
parted /dev/sda
```

- b. Assign a disk label.

```
(parted) mklabel gpt
```

- c. Set the unit size as sector.

```
(parted) unit s
```

- d. Check the partition table by displaying the existing partitions.

```
(parted) print
```

- e. Remove the partitions listed in the previous step.

```
(parted) rm part#
```

- f. Create a new first partition.

```
(parted) mkpart primary 64s 1048639s
```

- g. Specify this is a bootable partition.

```
(parted) set 1 boot on
```

7. Create second primary (boot) and third primary (LVM) partitions.

- a. Create a second primary partition as a UEFI boot partition with `fat32`.

```
(parted) mkpart primary fat32 1048640s 1572927s
```

```
(parted) set 2 boot on
```

- b. Create a new third partition.

```
(parted) mkpart primary 1572928s -1s
```

- c. Configure the third partition as a physical volume.

```
(parted) set 3 lvm on
```

- d. Write the information to disk, then quit.

```
(parted) quit
```

## 8. Create the physical volume and volume group.

```
lvm pvcreate /dev/sda3
lvm vgcreate VGExaDb /dev/sda3
```

If the physical volume or volume group already exists, then remove and then re-create them as follows:

```
lvm vgremove VGExaDb
lvm pvremove /dev/sda3
lvm pvcreate /dev/sda3
lvm vgcreate VGExaDb /dev/sda3
```

## 9. Re-create the customized LVM partitions, then create and mount the file systems.

### Note:

Use the following information and examples as guidance for this step. You must make the necessary adjustments for customized LVM partitions and file systems. For example, you may need to adjust the names and sizes of various partitions to match your previous customizations, or you may need to create additional custom partitions.

### a. Create the logical volumes.

For example, to create logical volumes for the / (root) and /u01 file systems:

```
lvm lvcreate -n LVDbSys1 -L40G VGExaDb
lvm lvcreate -n LVDbOra1 -L100G VGExaDb
```

### b. Create the file systems.

- If your environment uses the `xfs` file system type, then you could use the following commands to create the / (root), /u01, and /boot file systems:

```
mkfs.xfs /dev/VGExaDb/LVDbSys1 -f
mkfs.xfs /dev/VGExaDb/LVDbOra1 -f
mkfs.xfs /dev/sda1 -f
```

- Alternatively, if your environment uses the `ext4` file system type, then you could use the following commands:

```
mkfs.ext4 /dev/VGExaDb/LVDbSys1
mkfs.ext4 /dev/VGExaDb/LVDbOra1
mkfs.ext4 /dev/sda1
```

### c. Label the file systems.

- If your environment uses the `xfs` file system type, then you could use the following commands to label the `/` (root), `/u01`, and `/boot` file systems:

```
xfs_admin -L DBSYS /dev/VGExaDb/LVDbSys1
xfs_admin -L DBORA /dev/VGExaDb/LVDbOra1
xfs_admin -L BOOT /dev/sda1
```

- Alternatively, if your environment uses the `ext4` file system type, then you could use the following commands:

```
e2label /dev/VGExaDb/LVDbSys1 DBSYS
e2label /dev/VGExaDb/LVDbOra1 DBORA
e2label /dev/sda1 BOOT
```

- d. Create mount points for all the partitions to mirror the original system, and mount the respective partitions.

For example, assuming that `/mnt` is used as the top level directory for the recovery operation, you could use the following commands to create the directories and mount the partitions:

```
mkdir -p /mnt/u01
mkdir -p /mnt/boot
mount /dev/VGExaDb/LVDbSys1 /mnt -t filesystem_type
mount /dev/VGExaDb/LVDbOra1 /mnt/u01 -t filesystem_type
mount /dev/sda1 /mnt/boot -t filesystem_type
```

In the preceding commands, specify `xfs` or `ext4` as the `filesystem_type` according to your system configuration.

10. Create the system swap space.

For Exadata Database Machine X7 and X8 systems, the default swap size is 24 GB.

For example:

```
lvm lvcreate -n LVDbSwap1 -L24G VGExaDb
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

11. Create `/mnt/boot/efi`, label `/dev/sda2`, and mount `/dev/sda2` on `/mnt/boot/efi` with type `vfat`.

```
mkdir /mnt/boot/efi
dosfslabel /dev/sda2 ESP
mount /dev/sda2 /mnt/boot/efi -t vfat
```

12. Bring up the network.

```
ip address add ip_address_for_eth0/netmask_for_eth0 dev eth0
ip link set up eth0
ip route add default via gateway_address dev eth0
```

13. Mount the NFS server where you have the backup.

The following example assumes that the backup is located in the `/export` directory of the NFS server with IP address `nfs_ip`.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/export /
root/mnt
```

#### 14. Restore from backup.

```
tar -pjxvf /root/mnt/mybackup.tar.bz2 -C /mnt
```

#### 15. Unmount the restored file systems.

For example:

```
umount /mnt/boot/efi
umount /mnt/boot
umount /mnt/tmp
umount /mnt/var/log/audit
umount /mnt/var/log
umount /mnt/var
umount /mnt/home
umount /mnt/u01
umount /mnt
```

#### 16. Detach the `diagnostics.iso` file.

#### 17. Check the boot devices and boot order for the `ExadataLinux_1` device.

##### a. Check the available boot devices.

```
efibootmgr
BootCurrent: 000C
Timeout: 1 seconds
BootOrder: 000C,0001,0002,0003,0004,0005,0007,0008,0009,000A,000B
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet
Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet
Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000B* Oracle Linux
Boot000C* USB:SUN
```

If the `Boot0000* ExadataLinux_1` device is not listed then create the device.

```
efibootmgr -c -d /dev/sda -p 2 -1 '\EFI\REDHAT\SHIM.EFI' -L
'ExadataLinux_1'
BootCurrent: 000C
Timeout: 1 seconds
```

```

BootOrder: 0000,000C,0001,0002,0003,0004,0005,0007,0008,0009,000A,000B
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000B* Oracle Linux
Boot000C* USB:SUN
Boot0000* ExadataLinux_1

```

- b. Configure the `Boot0000* ExadataLinux_1` device to be first in the boot order.

```

efibootmgr -o 0000
BootCurrent: 000B
Timeout: 1 seconds
BootOrder: 0000
Boot0000* ExadataLinux_1
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000B* USB:SUN
Boot000C* UEFI OS

```

18. Restart the system and update the boot order in the BIOS.

```
reboot
```

Modify the boot order to set the `ExadataLinux_1` boot device as the first device.

- a. Press F2 when booting the system.
- b. Go to the Setup Utility.
- c. Select **BOOT**.
- d. Set `ExadataLinux_1` for Boot Option #1.
- e. Exit the Setup Utility.

19. Run `reclaimdisks.sh` on restored database server.

```
/opt/oracle.SupportTools/reclaimdisks.sh -free -reclaim
```

20. If the recovery was on Oracle Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

## 2.11.6 Recovering Exadata X6 or Earlier Database Servers with Customized Partitions

This procedure describes how to recover Oracle Exadata Database Servers for Exadata Database Machine X6-2 or earlier running Oracle Linux from a snapshot-based backup when using customized partitions.

1. Prepare an NFS server to host the backup archive file (`mybackup.tar.bz2`).

The NFS server must be accessible by IP address.

For example, on an NFS server with the IP address `nfs_ip`, where the directory `/export` is exported as an NFS mount, put the backup file (`mybackup.tar.bz2`) in the `/export` directory.

2. Restart the recovery target system using the `diagnostics.iso` file.

See Booting a Server using the Diagnostic ISO File in *Oracle Exadata System Software User's Guide*.

3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. If required, use `/opt/MegaRAID/storcli/storcli64` (or `/opt/MegaRAID/MegaCli/MegaCli64` for releases earlier than Oracle Exadata System Software 19c) to configure the disk controller to set up the disks.

5. Unmount `/mnt/cell`

```
umount /mnt/cell
```

6. Create the boot partition.

- a. Start an interactive session using the `partd` command.

```
parted /dev/sda
```

- b. Assign a disk label.

- If you are running Oracle Exadata System Software release 11.2.3.3.0 or later:

```
(parted) mklabel gpt
```

- If you are running a release earlier than Oracle Exadata System Software release 11.2.3.3.0:

```
(parted) mklabel msdos
```

- c. Set the unit size as sector.

```
(parted) unit s
```

- d. Check the partition table by displaying the existing partitions.

```
(parted) print
```

- e. Remove the partitions that will be re-created.

```
(parted) rm <part#>
```

- f. Create a new first partition.

```
(parted) mkpart primary 63 1048639
```

- g. Specify this is a bootable partition.

```
(parted) set 1 boot on
```

7. Create an additional primary (LVM) partition.

- If using Oracle Exadata System Software release 18.1.0.0.0 or later — Create second primary (bios\_grub) and third primary (LVM) partitions:

- a. Create a new second partition.

```
(parted) mkpart primary 1048640 1050687
```

- b. Specify this is a GRUB BIOS partition.

```
(parted) set 2 bios_grub on
```

- c. Create a new third partition.

```
(parted) mkpart primary 1050688 1751949278
```

- d. Specify this is a physical volume.

```
(parted) set 3 lvm on
```

- e. Write the information to disk, then quit.

```
(parted) quit
```

- If using a release earlier than Oracle Exadata System Software release 18.1.0.0.0:



- a. Create a new second partition.

```
(parted) mkpart primary 1048640 -1
```

- b. Specify this is a physical volume.

```
(parted) set 2 lvm on
```

- c. Write the information to disk, then quit.

```
(parted) quit
```

8. Re-create the customized LVM partitions and create the file systems.

- a. Create the physical volume, volume group, and the logical volumes as follows:

```
lvm pvcreate /dev/sda2
lvm vgcreate VGExaDb /dev/sda2
```

- b. Create the logical volume for the / (root) directory, a file system, and label it.

- Create the logical volume:

```
lvm lvcreate -n LVDbSys1 -L40G VGExaDb
```

- If using Oracle Exadata System Software release 12.1.2.2.0 or later, then create the logical volume for the reserved partition.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L1G VGExaDb
```



**Note:**

Do not create any file system on this logical volume.

- Create the file system.

- If you previously had an `ext4` file system, use the `mkfs.ext4` command:

```
mkfs.ext4 /dev/VGExaDb/LVDbSys1
```

- If you previously had an `ext3` file system, use the `mkfs.ext3` command:

```
mkfs.ext3 /dev/VGExaDb/LVDbSys1
```

- Label the file system.

```
e2label /dev/VGExaDb/LVDbSys1 DBSYS
```

c. Create the system swap space.

```
lvm lvcreate -n LVDbSwap1 -L24G VGExaDb
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

d. Create the logical volume for the /u01 directory, and label it.

- Create the logical volume:

```
lvm lvcreate -n LVDbOra1 -L100G VGExaDb
```

- Create the file system.

- If you previously had an `ext4` file system, then use the `mkfs.ext4` command:

```
mkfs.ext4 /dev/VGExaDb/LVDbOra1
```

- If you previously had an `ext3` file system, then use the `mkfs.ext3` command:

```
mkfs.ext3 /dev/VGExaDb/LVDbOra1
```

- Label the file system.

```
e2label /dev/VGExaDb/LVDbOra1 DBORA
```

e. Create a file system on the /boot partition, and label it.

- Create the file system.

- If you previously had an `ext4` file system, use the `mkfs.ext4` command:

```
mkfs.ext4 /dev/sda1
```

- If you previously had an `ext3` file system, use the `mkfs.ext3` command:

```
mkfs.ext3 /dev/sda1
```

- Label the file system:

```
e2label /dev/sda1 BOOT
```

 **Note:**

For customized file system layouts, additional logical volumes can be created at this time. For customized layouts, different sizes may be used.

9. Create mount points for all the partitions to mirror the original system, and mount the respective partitions.

For example, assuming `/mnt` is used as the top level directory for this, the mounted list of partitions may look like the following:

```
/dev/VGExaDb/LVDbSys1 on /mnt
/dev/VGExaDb/LVDbOral on /mnt/u01
/dev/sda1 on /mnt/boot
```

 **Note:**

For customized file system layouts with additional logical volumes, additional mount points need to be created during this step.

The following is an example for Exadata Database Machine X6-2 and earlier systems of how to mount the `root` file system, and create two mount points. In the commands below, `filesystem_type` specifies the applicable file system type; either `ext3` or `ext4`.

```
mount /dev/VGExaDb/LVDbSys1 /mnt -t filesystem_type
mkdir /mnt/u01 /mnt/boot
mount /dev/VGExaDb/LVDbOral /mnt/u01 -t filesystem_type
mount /dev/sda1 /mnt/boot -t filesystem_type
```

#### 10. Bring up the network.

- If the operating system is Oracle Linux 6 or later:

```
ip address add ip_address_for_eth0/netmask_for_eth0 dev eth0
ip link set up eth0
ip route add default via gateway_address dev eth0
```

- If the operating system is Oracle Linux 5:

```
ifconfig eth0 ip_address_for_eth0 netmask netmask_for_eth0 up
```

#### 11. Mount the NFS server where you have the backup.

The following example assumes that the backup is located in the `/export` directory of the NFS server with IP address `nfs_ip`.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/export /
root/mnt
```

#### 12. Restore from backup.

```
tar -pjxvf /root/mnt/mybackup.tar.bz2 -C /mnt
```

#### 13. Unmount the restored file systems, and remount the `/boot` partition.

```
umount /mnt/u01
umount /mnt/boot
umount /mnt
```

```
mkdir /boot
mount /dev/sda1 /boot -t filesystem_type
```

#### 14. Set up the boot loader.

In the following instructions, `/dev/sda1` is the `/boot` area.

- If using Oracle Exadata System Software release 18.1.0.0.0 or later:

```
grub2-install /dev/sda
```

```
Installing for i386-pc platform.
Installation finished. No error reported.
```

- If using a release earlier than Oracle Exadata System Software release 18.1.0.0.0:

```
grub
grub> find /I_am_hd_boot
grub> root (hdX,0)
grub> setup (hdX)
grub> quit
```

In the preceding commands, the `find` command identifies the hard disk that contains the file `I_am_hd_boot`; for example `(hd0,0)`. Use the value that you observe to specify the `hdX` value in the GRUB `root` and `setup` commands.

#### 15. Detach the `diagnostics.iso` file.

#### 16. Unmount the `/boot` partition.

```
umount /boot
```

#### 17. Restart the system.

```
shutdown -r now
```

This completes the restoration procedure for the server.

- 18. If the recovery was on Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

## 2.11.7 Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery

After the Oracle Linux database server in Oracle Exadata Database Machine Eighth Rack has been re-imaged, restored, or rescued, you can then reconfigure the eighth rack.

- [Configuring Eighth Rack On X3-2 or Later Machines Running Oracle Exadata Storage Server Release 12.1.2.3.0 or Later](#)
- [Configuring Eighth Rack On X3-2 Machines Running Oracle Exadata Storage Server Release 12.1.2.2.3 or Earlier](#)

### 2.11.7.1 Configuring Eighth Rack On X3-2 or Later Machines Running Oracle Exadata Storage Server Release 12.1.2.3.0 or Later

The following procedure should be performed after Oracle Linux database server in Oracle Exadata Database Machine Eighth Rack has been re-imaged, restored, or rescued.

For X3-2 systems, use this method only if you are running Oracle Exadata System Software release 12.1.2.3.0 or later.

1. On the recovered server, check that the `resourcecontrol` utility exists in the `/opt/oracle.SupportTools` directory. If not, copy it from another database server to the recovered server.

2. Ensure proper permissions are set on the `resourcecontrol` utility.

```
chmod 740 /opt/oracle.SupportTools/resourcecontrol
```

3. Verify the current configuration.

```
dbmcli -e LIST DBSERVER ATTRIBUTES coreCount
```

See [Table 2-3](#) for the number of cores allowed for each machine configuration. If the correct value is shown, then no configuration changes are necessary. If that value is not shown, then continue to step 4 of this procedure.

4. Change the enabled core configuration.

```
dbmcli -e ALTER DBSERVER pendingCoreCount=new_core_count FORCE
```

*new\_core\_count* for an Eighth Rack is:

- X8-2: 24
- X7-2: 24
- X6-2: 22
- X5-2: 18
- X4-8: 60
- X4-2: 12

5. Restart the server.

```
reboot
```

6. Verify the changes to the configuration.

```
dbmcli -e LIST DBSERVER ATTRIBUTES coreCount
```

## 2.11.7.2 Configuring Eighth Rack On X3-2 Machines Running Oracle Exadata Storage Server Release 12.1.2.2.3 or Earlier

The following procedure should be performed after Oracle Linux database server in Oracle Exadata Database Machine Eighth Rack has been re-imaged, restored, or rescued.

1. Copy the `/opt/oracle.SupportTools/resourcecontrol` utility from another database server to the `/opt/oracle.SupportTools/resourcecontrol` directory on the recovered server.
2. Ensure proper permissions are set on the utility.

```
chmod 740 /opt/oracle.SupportTools/resourcecontrol
```

3. Verify the current configuration.

The output from the command is shown in this example.

```
/opt/oracle.SupportTools/resourcecontrol -show

Validated hardware and OS. Proceed.
Number of cores active: 8
```

For an eighth rack configuration, eight cores should be enabled. If that value is shown, then no configuration changes are necessary. If that value is not shown, then continue to step 4 of this procedure.

### Note:

If there is an error similar to the following after running the utility, then restarting the server one or more times usually clears the error:

```
Validated hardware and OS. Proceed.
Cannot get ubisconfig export. Cannot Proceed. Exit.
```

4. Change the configuration for enabled cores.

```
/opt/oracle.SupportTools/resourcecontrol -cores 8
```

5. Restart the server.

```
shutdown -r now
```

6. Verify the changes to the configuration.

```
/opt/oracle.SupportTools/resourcecontrol -show
```

The following is an example of the expected output from the command for the database server:

```
This is a Linux database server.
Validated hardware and OS. Proceed.
Number of cores active per socket: 4
```

## 2.12 Re-Imaging the Oracle Exadata Database Server

The re-image procedure is necessary when a database server needs to be brought to an initial state for various reasons.

Some examples scenarios for re-imaging the database server are:

- You want to install a new server and need to use an earlier release than is in the image already installed on the server.
- You need to replace a damaged database server with a new database server.
- Your database server had multiple disk failures causing local disk storage failure and you do not have a database server backup.
- You want to repurpose the server to a new rack.

During the re-imaging procedure, the other database servers on Exadata Database Machine are available. When the new server is added to the cluster, the software is copied from an existing database server to the new server. It is your responsibility to restore scripting, CRON jobs, maintenance actions, and non-Oracle software.

### Note:

The procedures in this section assume the database is Oracle Database 11g Release 2 (11.2) or later.

Starting with Oracle Exadata System Software release 19.1.0, Secure Eraser is automatically started during re-imaging if the hardware supports Secure Eraser. This significantly simplifies the re-imaging procedure while maintaining performance. Now, when re-purposing a rack, you only have to image the rack and the secure data erasure is taken care of transparently as part of the process.

The following tasks describes how to re-image an Oracle Exadata Database Server running Oracle Linux:

- [Contact Oracle Support Services](#)  
If a failed server is being replaced, open a support request with Oracle Support Services.
- [Download Latest Release of Cluster Verification Utility](#)  
The latest release of the cluster verification utility (`cluvfy`) is available from My Oracle Support.
- [Remove the Database Server from the Cluster](#)  
If you are reimaging a failed server or repurposing a server, follow the steps in this task to remove the server from the cluster before you reimage it. If you are reimaging the server for a different reason, skip this task and proceed with the reimaging task next.

- [Image the Database Server](#)  
After the database server has been installed or replaced, you can image the new database server.
- [Configure the Re-imaged Database Server](#)  
The re-imaged database server does not have any host names, IP addresses, DNS or NTP settings. The steps in this task describe how to configure the re-imaged database server.
- [Prepare the Re-imaged Database Server for the Cluster](#)  
This task describes how to ensure the changes made during initial installation are done to the re-imaged, bare metal database server.
- [Apply Oracle Exadata System Software Patch Bundles to the Replacement Database Server](#)  
Oracle periodically releases Oracle Exadata System Software patch bundles for Exadata Database Machine.
- [Clone Oracle Grid Infrastructure to the Replacement Database Server](#)  
This procedure describes how to clone Oracle Grid Infrastructure to the replacement database server.
- [Clone Oracle Database Homes to the Replacement Database Server](#)  
The following procedure describes how to clone the Oracle Database homes to the replacement server.

## 2.12.1 Contact Oracle Support Services

If a failed server is being replaced, open a support request with Oracle Support Services.

The support engineer will identify the failed server, and send a replacement. The support engineer will ask for the output from the `imagehistory` command run from a surviving database server. The output provides a link to the `computeImageMaker` file that was used to image the original database server, and provides a means to restore the system to the same level.

## 2.12.2 Download Latest Release of Cluster Verification Utility

The latest release of the cluster verification utility (`cluvfy`) is available from My Oracle Support.

See My Oracle Support note 316817.1 for download instructions and other information.

### Related Topics

- [My Oracle Support note 316817.1: Cluster Verification Utility \(CLUVfy\) FAQ](#)

## 2.12.3 Remove the Database Server from the Cluster

If you are reimaging a failed server or repurposing a server, follow the steps in this task to remove the server from the cluster before you reimage it. If you are reimaging the server for a different reason, skip this task and proceed with the reimaging task next.

The steps in this task are performed using a working database server in the cluster. In the following commands, *working\_server* is a working database server, and *failed\_server* is the database server you are removing, either because it failed or it is being repurposed.

1. Log in as the `oracle` or `grid` user on a database server in the cluster.



Log in as the user that owns the Oracle Grid Infrastructure software installation.

2. Disable the listener that runs on the failed server.

```
$ srvctl disable listener -n failed_server
$ srvctl stop listener -n failed_server
```

3. Delete the Oracle home from the Oracle inventory.

In the following command, *list\_of\_working\_servers* is a list of the servers that are still working in the cluster, such as dm01db02, dm01db03, and so on.

In the following command, replace /u01/app/oracle/product/12.1.0.2/dbhome\_1 with the location of your Oracle Database home directory.

```
$ cd $ORACLE_HOME/oui/bin
$./runInstaller -updateNodeList ORACLE_HOME= \
/u01/app/oracle/product/12.1.0.2/dbhome_1
"CLUSTER_NODES=list_of_working_servers"
```

4. Log in as the `grid` user on the database server.

The `grid` user refers to the operating system user that owns the Oracle Grid Infrastructure software installation. The `$ORACLE_HOME` variable should point to the location of the Grid home.

5. Verify the failed server is unpinned.

```
$ olsnodes -s -t
```

The following is an example of the output from the command:

dm01adm05	Inactive	Unpinned
dm01adm06	Active	Unpinned
dm01adm07	Active	Unpinned
dm01adm08	Active	Unpinned

6. Log in as the `root` user on the database server.
7. Stop and delete the VIP resources for the failed database server.

```
srvctl stop vip -i failed_server-vip
PRCC-1016 : failed_server-vip.example.com was already stopped

srvctl remove vip -i failed_server-vip
Please confirm that you intend to remove the VIPs failed_server-vip
(y/[n]) y
```

8. Delete the server from the cluster.

```
crsctl delete node -n failed_server
CRS-4661: Node dm01db01 successfully deleted.
```

If you receive an error message similar to the following, then relocate the voting disks.

```
CRS-4662: Error while trying to delete node dm01db01.
CRS-4000: Command Delete failed, or completed with errors.
```

To relocate the voting disks use the following steps:

- a. Determine the current location of the voting disks.

```
crsctl query css votedisk
```

The following is an example of the output from the command. The current location is DBFS\_DG.

```
STATE File Universal Id File Name Disk
group
-- ----- -

1. ONLINE 123456789abab (o/192.168.73.102/DATA_CD_00_dm01ce107)
[DBFS_DG]
2. ONLINE 123456789cdcd (o/192.168.73.103/DATA_CD_00_dm01ce108)
[DBFS_DG]
3. ONLINE 123456789efef (o/192.168.73.100/DATA_CD_00_dm01ce105)
[DBFS_DG]
Located 3 voting disk(s).
```

- b. Relocate the voting disks to another disk group.

```
./crsctl replace votedisk +DATA
```

```
Successful addition of voting disk 2345667aabbdd.
```

```
...
```

```
CRS-4266: Voting file(s) successfully replaced
```

- c. Relocate the voting disks to the original location using a command similar to the following:

```
./crsctl replace votedisk +DBFS_DG
```

- d. Delete the server from the cluster.

9. Log in as the `grid` user on the database server.

The `grid` user refers to the operating system user that owns the Oracle Grid Infrastructure software installation. The `$ORACLE_HOME` variable should point to the location of the Grid home.

10. Update the Oracle inventory.

In the following command, replace `/u01/app/12.1.0.2/grid` with the location of your Oracle Grid Infrastructure home directory.

```
$ cd $ORACLE_HOME/oui/bin
$./runInstaller -updateNodeList ORACLE_HOME=/u01/app/12.1.0.2/grid \
 "CLUSTER_NODES=list_of_working_servers" CRS=TRUE
```

#### 11. Verify the server was deleted successfully.

```
$ cluvfy stage -post nodedel -n failed_server -verbose
```

The following is an example of the output from the command:

```
Performing post-checks for node removal
Checking CRS integrity...
The Oracle clusterware is healthy on node "dm01db02"
CRS integrity check passed
Result:
Node removal check passed
Post-check for node removal was successful.
```

#### Related Topics

- *Oracle Real Application Clusters Administration and Deployment Guide*

## 2.12.4 Image the Database Server

After the database server has been installed or replaced, you can image the new database server.

You can use installation media on a USB thumb drive, or a touchless option using PXE or ISO attached to the ILOM. See *Imaging a New System in Oracle Exadata Database Machine Installation and Configuration Guide* for the details.

## 2.12.5 Configure the Re-imaged Database Server

The re-imaged database server does not have any host names, IP addresses, DNS or NTP settings. The steps in this task describe how to configure the re-imaged database server.

You need the following information prior to configuring the re-imaged database server:

- Name servers
- Time zone, such as Americas/Chicago
- NTP servers
- IP address information for the management network
- IP address information for the client access network
- IP address information for the RDMA Network Fabric
- Canonical host name
- Default gateway

The information should be the same on all database servers in Exadata Database Machine. The IP addresses can be obtained from DNS. In addition, a document with the information should have been provided when Exadata Database Machine was installed.

The following procedure describes how to configure the re-imaged database server:

1. Power on the replacement database server. When the system boots, it automatically runs the Configure Oracle Exadata routine, and prompts for information.
2. Enter the information when prompted, and confirm the settings. The start up process will continue.

 **Note:**

- If the database server does not use all network interfaces, then the configuration process stops, and warns that some network interfaces are disconnected. It prompts whether to retry the discovery process. Respond with `yes` or `no`, as appropriate for the environment.
- If bonding is used for the client access network, then it is set in the default active-passive mode at this time.

## 2.12.6 Prepare the Re-imaged Database Server for the Cluster

This task describes how to ensure the changes made during initial installation are done to the re-imaged, bare metal database server.

 **Note:**

For Oracle VM systems, follow the procedure in [Expanding an Oracle RAC Cluster on Oracle VM Using OEDACLI](#).

1. Copy or merge the contents of the following files using files on a working database server as reference:
  - a. Copy the contents of the `/etc/security/limits.conf` file.
  - b. Merge the contents of the `/etc/hosts` files.
  - c. Copy the `/etc/oracle/cell/network-config/cellinit.ora` file.
  - d. Update the `/etc/oracle/cell/network-config/cellinit.ora` file with the `IP_ADDRESS` of the `ifcfg-bondib0` interface (in case of active/passive bonding) or `ib0` and `ib1` interfaces (in case of active/active bonding) of the replacement server.
  - e. Copy the `/etc/oracle/cell/network-config/cellip.ora` file.

The content of the `cellip.ora` file should be the same on all database servers.
  - f. Configure additional network requirements, such as 10 GbE.
  - g. Copy the `modprobe` configuration.

The contents of the configuration file should be the same on all database servers.

    - Oracle Linux 5 or 6: The file is located at `/etc/modprobe.conf`.
    - Oracle Linux 7: The file is located at `/etc/modprobe.d/exadata.conf`.
  - h. Copy the `/etc/sysctl.conf` file.

The contents of the file should be the same on all database servers.

- i. Update the `cellroute.ora`.  
Make a copy of the `/etc/oracle/cell/network-config/cellroute.ora` file. Modify the contents on the replacement server to use the local InfiniBand interfaces on the new node.
  - j. Restart the database server so the network changes take effect.
2. Set up the users for the software owners on the replacement database server by adding groups.

If you are using role-separated management, then the users are usually `oracle` and `grid`. If you use a single software owner, then the user is usually `oracle`. The group information is available on a working database server.

- a. Obtain the current group information from a working database server.

```
id oracle
uid=1000(oracle) gid=1001(oinstall)
groups=1001(oinstall),1002(dba),1003(oper),1004(asmdba)
```

- b. Use the `groupadd` command to add the group information to the replacement database server.

```
groupadd -g 1001 oinstall
groupadd -g 1002 dba
groupadd -g 1003 oper
groupadd -g 1004 asmdba
```

- c. Obtain the current user information from a working database server.

```
id oracle uid=1000(oracle) gid=1001(oinstall) \
groups=1001(oinstall),1002(dba),1003(oper),1004(asmdba)
```

- d. Add the user information to the replacement database server.

```
useradd -u 1000 -g 1001 -G 1001,1002,1003,1004 -m -d /home/
oracle -s \
/bin/bash oracle
```

- e. Create the Oracle Base and Grid home directories, such as `/u01/app/oracle` and `/u01/app/12.2.0.1/grid`.

```
mkdir -p /u01/app/oracle
mkdir -p /u01/app/12.2.0.1/grid
chown -R oracle:oinstall /u01/app
```

- f. Change the ownership on the `cellip.ora` and `cellinit.ora` files.

The ownership is usually `oracle:oinstall`.

```
chown -R oracle:oinstall /etc/oracle/cell/network-config
```



replacement Oracle Exadata Database Server. Determine if a patch bundle has been applied as follows:

- Prior to Oracle Exadata System Software release 11.2.1.2.3, the database servers did not maintain version history information. To determine the release number, log in to Oracle Exadata Storage Server, and run the following command:

```
imageinfo -ver
```

If the command shows a different release than the release used by the `computeImageMaker` file, then Oracle Exadata System Software patch has been applied to Oracle Exadata Database Machine and must be applied to the replacement Oracle Exadata Database Server.

- Starting with Oracle Exadata System Software release 11.2.1.2.3, the `imagehistory` command exists on the Oracle Exadata Database Server. Compare information on the replacement Oracle Exadata Database Server to information on a working Oracle Exadata Database Server. If the working database has a later release, then apply the Oracle Exadata Storage Server patch bundle to the replacement Oracle Exadata Database Server.

#### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

## 2.12.8 Clone Oracle Grid Infrastructure to the Replacement Database Server

This procedure describes how to clone Oracle Grid Infrastructure to the replacement database server.

In the following commands, *working\_server* is a working database server, and *replacement\_server* is the replacement database server. The commands in this procedure are run from a working database server as the Grid home owner. When the `root` user is needed to run a command, it will be called out.

1. Verify the hardware and operating system installation using the cluster verification utility (`cluvfy`).

```
$ cluvfy stage -post hwos -n replacement_server,working_server -
verbose
```

The phrase `Post-check for hardware and operating system setup was successful` should appear at the end of the report. If the cluster verification utility fails to validate the storage on the replacement server, you can ignore those messages.

2. Verify peer compatibility.

```
$ cluvfy comp peer -refnode working_server -n replacement_server \
-orainv oinstall -osdba dba | grep -B 3 -A 2 mismatched
```

The following is an example of the output:

```
Compatibility check: Available memory [reference node: dm01db02]
Node Name Status Ref. node status Comment

dm01db01 31.02GB (3.2527572E7KB) 29.26GB (3.0681252E7KB) mismatched
Available memory check failed
Compatibility check: Free disk space for "/tmp" [reference node: dm01db02]
Node Name Status Ref. node status Comment

dm01db01 55.52GB (5.8217472E7KB) 51.82GB (5.4340608E7KB) mismatched
Free disk space check failed
```

If the only failed components are related to the physical memory, swap space and disk space, then it is safe to continue.

### 3. Perform the requisite checks for adding the server.

- a. Ensure the `GRID_HOME/network/admin/samples` directory has permissions set to 750.
- b. Validate the addition of the database server.

Run the following command as the `oracle` user. The command prompts for the password of the `root` user.

```
$ cluvfy stage -pre nodeadd -n replacement_server -fixup -method root
-verbose
Enter "ROOT" password:
```

If the only failed component is related to swap space, then it is safe to continue.

If the command returns an error, then set the following environment variable and rerun the command:

```
$ export IGNORE_PREADDDNODE_CHECKS=Y
```

### 4. Add the replacement database server to the cluster.

If you are using Oracle Grid Infrastructure release 12.1 or higher, include the `CLUSTER_NEW_NODE_ROLES` attribute, as shown in the following example.

```
$ cd GRID_HOME/addnode

$./addnode.sh -silent "CLUSTER_NEW_NODES={replacement_server}" \
 "CLUSTER_NEW_VIRTUAL_HOSTNAMES={replacement_server-vip}" \
 "CLUSTER_NEW_NODE_ROLES={hub}"
```

The second command causes Oracle Universal Installer to copy the Oracle Clusterware software to the replacement database server. A message similar to the following is displayed:

```
WARNING: A new inventory has been created on one or more nodes in this
session.
However, it has not yet been registered as the central inventory of this
system. To register the new inventory please run the script at
```



'/u01/app/oraInventory/orainstRoot.sh' with root privileges on nodes 'dm01db01'. If you do not register the inventory, you may not be able to update or patch the products you installed.

The following configuration scripts need to be executed as the "root" user in each cluster node:

```
/u01/app/oraInventory/orainstRoot.sh #On nodes dm01db01
```

```
/u01/app/12.1.0.2/grid/root.sh #On nodes dm01db01
```

##### 5. Run the configuration scripts.

As the `root` user, first disable HAIP, then run the `orainstRoot.sh` and `root.sh` scripts on the replacement database server using the commands shown in the following example.

```
export HAIP_UNSUPPORTED=true
/u01/app/oraInventory/orainstRoot.sh
Creating the Oracle inventory pointer file (/etc/oraInst.loc)
Changing permissions of /u01/app/oraInventory.
Adding read,write permissions for group.
Removing read,write,execute permissions for world.
Changing groupname of /u01/app/oraInventory to oinstall.
The execution of the script is complete.

GRID_HOME/root.sh
```

#### Note:

Check `GRID_HOME/install/ log` files for the output of `root.sh` script.

If you are running Oracle Grid Infrastructure release 11.2, then the output file created by the script reports that the listener resource on the replaced database server failed to start. This is the expected output.

```
/u01/app/11.2.0/grid/bin/srvctl start listener -n dm01db01 \
...Failed
/u01/app/11.2.0/grid/perl/bin/perl \
-I/u01/app/11.2.0/grid/perl/lib \
-I/u01/app/11.2.0/grid/crs/install \
/u01/app/11.2.0/grid/crs/install/rootcrs.pl execution failed
```

After the scripts are run, the following message is displayed:

```
The Cluster Node Addition of /u01/app/12.1.0.2/grid was successful.
Please check '/tmp/silentInstall.log' for more details.
```

**6. Check the cluster.**

```
$ GRID_HOME/bin/crsctl check cluster -all

node1:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online

node2:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online

node3:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
```

**7. If you are running Oracle Grid Infrastructure release 11.2, then re-enable the listener resource.**

Run the following commands on the replacement database server.

```
GRID_HOME/grid/bin/srvctl enable listener -l LISTENER \
-n replacement_server

GRID_HOME/grid/bin/srvctl start listener -l LISTENER \
-n replacement_server
```

**8. Start the disk groups on the replacement server.**

**a. Check disk group status.**

In the following example, notice that disk groups are offline on the replacement server.

```
$ crsctl stat res -t

Name Target State Server State details

Local Resources

ora.DATAAC1.dg
 ONLINE ONLINE node1 STABLE
 OFFLINE OFFLINE node2 STABLE
ora.DBFS_DG.dg
 ONLINE ONLINE node1 STABLE
 ONLINE ONLINE node2 STABLE
ora.LISTENER.lsnr
 ONLINE ONLINE node1 STABLE
 ONLINE ONLINE node2 STABLE
ora.RECOC1.dg
 ONLINE ONLINE node1 STABLE
 OFFLINE OFFLINE node2 STABLE
```

- b. For each offline disk group, run the `START DISKGROUP` command for each disk group that is offline from either the original server or the replacement server.

```
$ srvctl start diskgroup -diskgroup dgname
```

### Related Topics

- *Oracle Real Application Clusters Administration and Deployment Guide*

## 2.12.9 Clone Oracle Database Homes to the Replacement Database Server

The following procedure describes how to clone the Oracle Database homes to the replacement server.

Run the commands from a working database server as the `oracle` user. When the `root` user is needed to run a command, it will be called out.

1. Add the Oracle Database `ORACLE_HOME` to the replacement database server using the following commands:

```
$ cd /u01/app/oracle/product/12.1.0.2/dbhome_1/addnode
$./addnode.sh -silent "CLUSTER_NEW_NODES={replacement_server}"
```

The second command causes Oracle Universal Installer to copy the Oracle Database software to the replacement database server.

WARNING: The following configuration scripts need to be executed as the "root" user in each cluster node.

```
/u01/app/oracle/product/12.1.0.2/dbhome_1/root.sh #On nodes dm01db01
```

To execute the configuration scripts:  
Open a terminal window.  
Log in as root.  
Run the scripts on each cluster node.

After the scripts are finished, the following messages appear:

```
The Cluster Node Addition of /u01/app/oracle/product/12.1.0.2/
dbhome_1 was successful.
Please check '/tmp/silentInstall.log' for more details.
```

2. Run the following script on the replacement database server:

```
/u01/app/oracle/product/12.1.0.2/dbhome_1/root.sh
```

Check the `/u01/app/oracle/product/12.1.0.2/dbhome_1/install/root_replacement_server.com_date.log` file for the output of the script.

3. Run the Oracle Database Configuration Assistant (DBCA) in interactive mode to add database instances to the target nodes.

- a. Start up DBCA.

```
$ cd /u01/app/oracle/product/12.1.0.2/dbhome_1/bin
$./dbca
```

- b. On the Database Operation screen, select **Instance Management**. Click **Next**.
- c. On the Instance Operation screen, select **Add an instance**. Click **Next**.
- d. On the Database List screen, select the cluster database to which you want to add an instance.
- e. The List Instance screen displays the current instances. Click **Next** to add a new instance.
- f. The Add Instance screen displays the default name and the newly added node to the cluster. Accept the defaults and click **Next**.
- g. On the Summary screen, verify the plan and click **Finish**.
- h. On the Progress screen, watch for 100% completion.
- i. On the Finish screen, acknowledge the confirmation that the new instance was successfully added.

Verify that the instance has been added:

```
$ srvctl config database -db dbm01
```

Verify the administrative privileges on the target node:

```
$ cd /u01/app/oracle/product/12.1.0.2/dbhome_1/bin
$./cluvfy comp admprv -o db_config -d /u01/app/oracle/product/12.1.0.2/
dbhome_1 -n new_node
```

4. Ensure the instance parameters are set for the replaced database instance. The following is an example for the `CLUSTER_INTERCONNECTS` parameter.

```
SQL> SHOW PARAMETER cluster_interconnects
```

NAME	TYPE	VALUE
cluster_interconnects	string	

```
SQL> ALTER SYSTEM SET cluster_interconnects='192.168.73.90' SCOPE=spfile
SID='dbm1';
```

5. Validate the configuration files as follows:

- The `Oracle_home/dbs/initSID.ora` file points to the SPFILE in the Oracle ASM shared storage.
- The password file that is copied in the `Oracle_home/dbs` directory has been changed to `orapwSID`.

6. Check that any services that incorporated this instance before and ensure the services are updated to include this replacement instance.
7. If this procedure was performed on Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

**Related Topics**

- [Oracle Real Application Clusters Administration and Deployment Guide](#)

## 2.13 Changing Existing Elastic Configurations for Database Servers

Elastic configurations provide a flexible and efficient mechanism to change the server configuration of your Exadata Database Machine.

- [Adding a New Database Server to the Cluster](#)  
You can add a new database server to an existing Oracle Real Application Clusters (Oracle RAC) cluster running on Exadata Database Machine.
- [Moving an Existing Database Server to a Different Cluster](#)  
You can repurpose an existing database server and move it to a different cluster within the same Oracle Exadata Rack.
- [Dropping a Database Server from an Oracle RAC Cluster](#)  
You can remove a database server that is a member of an Oracle Real Application Clusters (Oracle RAC) cluster.

**Related Topics**

- [Changing Existing Elastic Configurations for Storage Cells](#)  
You can modify the capacity of your Exadata Database Machine using elastic configuration.

### 2.13.1 Adding a New Database Server to the Cluster

You can add a new database server to an existing Oracle Real Application Clusters (Oracle RAC) cluster running on Exadata Database Machine.

1. Determine if the new database server needs to be re-imaged or upgraded.  
Check the image label of the database servers in the cluster to which you want to add the new database server.
  - If the image label of the new database server is *later*, or a newer release, than the image label of the existing database servers in the cluster and you do not wish to upgrade the existing servers, then reimage the new database server to the older version by completing the following tasks:
    - [Download Latest Release of Cluster Verification Utility](#)
    - [Image the Database Server](#)
    - [Configure the Re-imaged Database Server](#)
  - If the image label of the new database server is *earlier*, or an older release, than the image label of the existing database servers in the cluster, then an

upgrade is required using patchmgr. See [Updating Exadata Software](#) for details.

2. Add the database server to the cluster by completing the following tasks:
  - For bare metal systems:
    - [Prepare the Re-imaged Database Server for the Cluster](#)
    - [Clone Oracle Grid Infrastructure to the Replacement Database Server](#)
    - [Clone Oracle Database Homes to the Replacement Database Server](#)
  - For Oracle VM systems:
    - [Expanding an Oracle RAC Cluster on Oracle VM Using OEDACLI](#)
3. Download and run the latest version of Oracle EXAchk to ensure that the resulting configuration implements the latest best practices for Exadata Database Machine.

## 2.13.2 Moving an Existing Database Server to a Different Cluster

You can repurpose an existing database server and move it to a different cluster within the same Oracle Exadata Rack.

1. Remove the database server from the existing Oracle Real Application Clusters (Oracle RAC) cluster.

- a. Stop Oracle Grid Infrastructure on the database server.

```
Grid_home/bin/crstl stop crs
```

- b. Remove the database server from the cluster by completing the steps in [Remove the Database Server from the Cluster](#).

2. Determine if the database server that is being repurposed needs to be reimaged.

Check the image label of the existing database servers in the cluster to which you want to add the database server. If the image label of the database server being added does not match the image label of the existing database servers in the cluster, then reimage the database server being added. Complete the following tasks:

- [Contact Oracle Support Services](#)
- [Download Latest Release of Cluster Verification Utility](#)
- [Image the Database Server](#)
- [Configure the Re-imaged Database Server](#)

If an upgrade is required, the upgrade may be performed using patchmgr. See [Updating Exadata Software](#) for details.

3. Add the database server to the cluster.

Complete the following tasks:

- [Prepare the Re-imaged Database Server for the Cluster](#)
- [Apply Oracle Exadata System Software Patch Bundles to the Replacement Database Server](#)
- [Clone Oracle Grid Infrastructure to the Replacement Database Server](#)
- [Clone Oracle Database Homes to the Replacement Database Server](#)

4. Download and run the latest version of Oracle EXAchk to ensure that the resulting configuration implements the latest best practices for Exadata Database Machine.

#### Related Topics

- [dbnodeupdate.sh and dbserver.patch.zip: Updating Exadata Database Server Software using the DBNodeUpdate Utility and patchmgr \(My Oracle Support Doc ID 1553103.1\)](#)

## 2.13.3 Dropping a Database Server from an Oracle RAC Cluster

You can remove a database server that is a member of an Oracle Real Application Clusters (Oracle RAC) cluster.

1. Stop Oracle Grid Infrastructure on the database server to be removed.

```
$ Grid_home/bin/crsctl stop crs
```

2. Remove the database server from the cluster by completing the steps in [Remove the Database Server from the Cluster](#).
3. Download and run the latest Oracle EXAchk to ensure that the resulting configuration implements the latest best practices for Exadata Database Machine.

## 2.14 Managing Quorum Disks for High Redundancy Disk Groups

Quorum disks allow for the Oracle RAC voting files to be stored in a high redundancy disk group on an Oracle Exadata Rack with fewer than five storage servers due to the presence of two extra failure groups.

- [Using Quorum Disks to Increase Fault Tolerance](#)  
Quorum disks are used to meet the minimum requirement of five failure groups for a high redundancy disk group on a system that does not have five storage servers.
- [Overview of Quorum Disk Manager](#)  
The Quorum Disk Manager utility, introduced in Oracle Exadata System Software release 12.1.2.3.0, helps you to manage the quorum disks.
- [Software Requirements for Quorum Disk Manager](#)  
You must satisfy the minimum software requirements to use the Quorum Disk Manager utility.
- [quorumdiskmgr Reference](#)
- [Add Quorum Disks to Database Nodes](#)  
You can add quorum disks to database nodes on an Oracle Exadata Rack with fewer than five storage servers that contains a high redundancy disk group.
- [Recreate Quorum Disks](#)  
In certain circumstances, you might need to recreate a quorum disk.
- [Use Cases](#)  
The following topics describe various configuration cases when using the quorum disk manager utility.

- [Reconfigure Quorum Disk After Restoring a Database Server](#)  
After restoring a database server, `lvdisplay` shows the quorum disk was not restored.

## 2.14.1 Using Quorum Disks to Increase Fault Tolerance

Quorum disks are used to meet the minimum requirement of five failure groups for a high redundancy disk group on a system that does not have five storage servers.

A **failure group** is a subset of the disks in a disk group, which could fail at the same time because they share hardware. Oracle recommends a minimum of three failure groups for normal redundancy disk groups and five failure groups for high redundancy disk groups to maintain the necessary number of copies of the Partner Status Table (PST) and to ensure robustness with respect to storage hardware failures. On Engineered Systems, these recommendations are enforced to ensure the highest availability of the system.

The PST contains status information about the Oracle Automatic Storage Management (Oracle ASM) disks in a disk group, such as the disk number, status (either online or offline), partner disk number, failure group info, and heartbeat info. To tolerate a single hardware failure, you need 3 total copies of the PST available to form a 2 of 3 majority. If there are two hardware failures, then you need a total for 5 copies of the PST so that after a double failure you still have a 3 of 5 majority.

A **quorum failure group** is a special type of failure group that does not contain user data. Quorum failure groups are used for storing the PST. A quorum failure group can also be used to store a copy of the voting file for Oracle Clusterware. Because disks in quorum failure groups do not contain user data, a quorum failure group is not considered when determining redundancy requirements in respect to storing user data.

In the event of a system failure, three failure groups in a normal redundancy disk group allow a comparison among three PSTs to accurately determine the most up to date and correct version of the PST, which could not be done with a comparison between only two PSTs. Similarly with a high redundancy disk group, if two failure groups are offline, then Oracle ASM would be able to make a comparison among the three remaining PSTs.

You can create quorum failure groups with Oracle Exadata Deployment Assistant (OEDA) when deploying Exadata or you can add them later using the Quorum Disk Manager Utility. The iSCSI quorum disks are created on database nodes and a voting file is created on those quorum disks. These additional quorum failure groups are used to meet the minimum requirement of five voting files and PSTs for a high redundancy disk group. Quorum disks are required when the following conditions exist:

- The Oracle Exadata Rack has fewer than five storage servers.
- The Oracle Exadata Rack has at least two database nodes.
- The Oracle Exadata Rack has at least one high redundancy disk group.

Quorum failure groups allow for a high redundancy disk group to exist on Oracle Exadata Racks with fewer than five storage servers by creating two extra failure groups. Without this feature, a disk group is vulnerable to a double partner storage server failure that results in the loss of the PST or voting file quorum, which can cause a complete cluster and database outage. Refer to My Oracle Support note 1339373.1 for how to restart the clusterware and databases in this scenario.

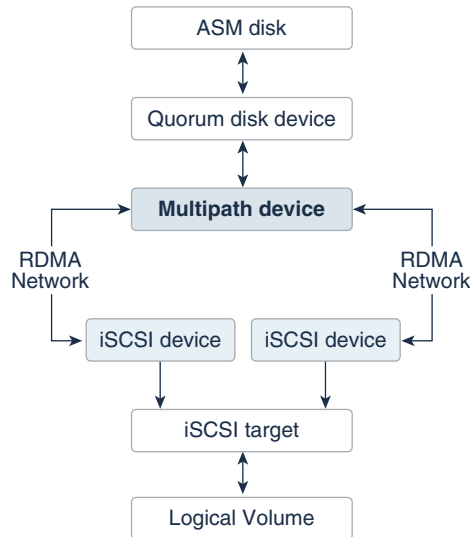
The iSCSI quorum disk implementation has high availability because the IP addresses on the RDMA Network Fabric are highly available using RDS.

Each iSCSI device shown in the figure below corresponds to a particular path to the iSCSI target. Each path corresponds to an RDMA Network Fabric port on the database node. For



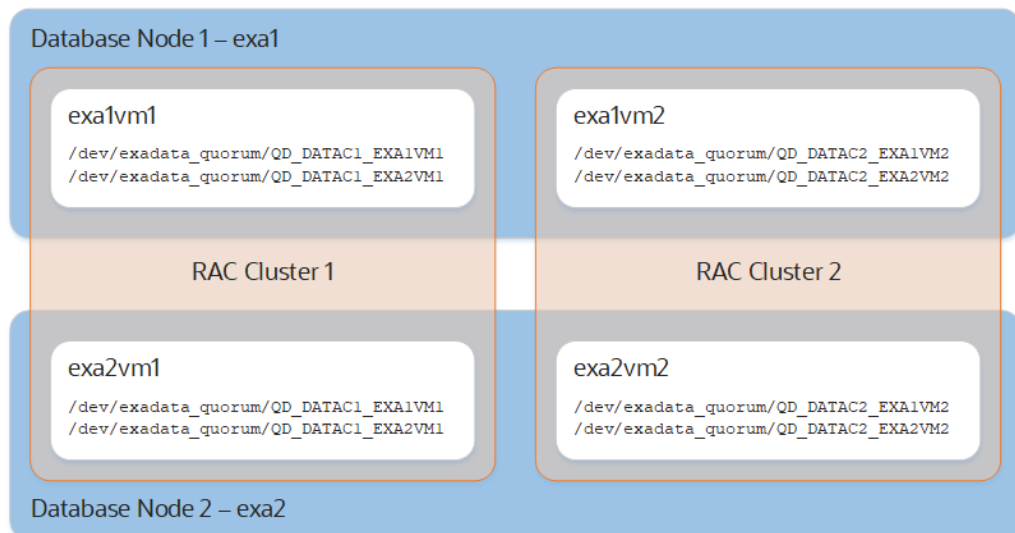
each multipath quorum disk device in an active–active system, there are two iSCSI devices, one for `ib0` or `re0` and the other for `ib1` or `re1`.

**Figure 2-1 Multipath Device Connects to Both iSCSI Devices in an Active-Active System**



Quorum disks can be used with bare metal Oracle Real Application Clusters (Oracle RAC) clusters and Oracle VM Oracle RAC clusters. For Oracle VM Oracle RAC clusters, the quorum disk devices reside in the Oracle RAC cluster nodes which are Oracle VM user domains as shown in the following figure.

**Figure 2-2 Quorum Disk Devices on Oracle VM Oracle RAC Cluster**



 **Note:**

For pkey-enabled environments, the interfaces used for discovering the targets should be the pkey interfaces used for the Oracle Clusterware communication. These interfaces are listed using the following command:

```
Grid_home/bin/oifcfg getif | grep cluster_interconnect | awk
'{print $1}'
```

**Related Topics**

- [Operational Steps for Recovery after Losing a Disk Group in an Exadata Environment \(Doc ID 1339373.1\)](#)
- Oracle ASM Failure Groups

## 2.14.2 Overview of Quorum Disk Manager

The Quorum Disk Manager utility, introduced in Oracle Exadata System Software release 12.1.2.3.0, helps you to manage the quorum disks.

This utility enables you to create an iSCSI quorum disk on two of the database nodes and store a voting file on those two quorum disks. These two additional voting files are used to meet the minimum requirement of five voting files for a high redundancy disk group.

The Quorum Disk Manager utility (`quorumdiskmgr`) is used to create and manage all the necessary components including the iSCSI configuration, the iSCSI targets, the iSCSI LUNs, and the iSCSI devices for implementing quorum disks.

**Related Topics**

- [quorumdiskmgr Reference](#)

## 2.14.3 Software Requirements for Quorum Disk Manager

You must satisfy the minimum software requirements to use the Quorum Disk Manager utility.

To use this feature, the following releases are required:

- Oracle Exadata System Software release 12.1.2.3.0 and above
- Patch 23200778 for all Oracle Database homes
- Oracle Grid Infrastructure release 12.1.0.2.160119 with patches 22722476 and 22682752, or Oracle Grid Infrastructure release 12.1.0.2.160419 and above

For new deployments, Oracle Exadata Deployment Assistant (OEDA) installs the patches automatically.

## 2.14.4 quorumdiskmgr Reference

The quorum disk manager utility (`quorumdiskmgr`) runs on each database server to enable you to create and manage iSCSI quorum disks on database servers. You use `quorumdiskmgr` to create, list, alter, and delete iSCSI quorum disks on database servers. The utility is installed on database servers when they are shipped.

- [Syntax for the Quorum Disk Manager Utility](#)
- [quorumdiskmgr Objects](#)
- [Creating a Quorum Disk Configuration \(--create --config\)](#)  
The `--create --config` action creates a quorum disk configuration.
- [Creating a Target \(--create --target\)](#)  
The `--create --target` action creates a target that will be used to create the devices to add to the specified Oracle ASM disk group.
- [Creating a Device \(--create --device\)](#)  
The `--create --device` action creates devices by discovering and logging into targets on database servers with an RDMA Network Fabric IP address in the specified list of IP addresses.
- [Listing Quorum Disk Configurations \(--list --config\)](#)  
The `--list --config` action lists the quorum disk configurations.
- [Listing Targets \(--list --target\)](#)  
The `--list --target` action lists the attributes of targets.
- [Listing Devices \(--list --device\)](#)  
The `--list --device` action lists the attributes of devices, including device path, size, host name and ASM disk group name.
- [Deleting Configurations \(--delete --config\)](#)  
The `--delete --config` action deletes quorum disk configurations.
- [Deleting Targets \(--delete --target\)](#)  
The `--delete --target` action deletes the targets created for quorum disks on database servers.
- [Deleting Devices \(--delete --device\)](#)  
The `--delete --device` command deletes quorum disk devices.
- [Changing Owner and Group Values \(--alter --config\)](#)  
The `--alter --config` action changes the owner and group configurations.
- [Changing the RDMA Network Fabric IP Addresses \(--alter --target\)](#)  
The `--alter --target` command changes the RDMA Network Fabric IP addresses of the database servers that have access to the local target created for the specified Oracle ASM disk group.

### 2.14.4.1 Syntax for the Quorum Disk Manager Utility

The quorum disk manager utility is a command-line tool. It has the following syntax:

```
quorumdiskmgr --verb --object [--options]
```

*verb* is an action performed on an object. It is one of: alter, create, delete, list.

*object* is an object on which the command performs an action.

*options* extend the use of a command combination to include additional parameters for the command.

When using the quorumdiskmgr utility, the following rules apply:

- Verbs, objects, and options are case-sensitive except where explicitly stated.

- Use the double quote character around the value of an option that includes spaces or punctuation.

## 2.14.4.2 quorumdiskmgr Objects

Object	Description
config	The quorum disk configurations include the owner and group of the ASM instance to which the iSCSI quorum disks will be added, and the list of network interfaces through which local and remote iSCSI quorum disks will be discovered.
target	A target is an endpoint on each database server that waits for an iSCSI initiator to establish a session and provides required IO data transfer.
device	A device is an iSCSI device created by logging into a local or remote target.

## 2.14.4.3 Creating a Quorum Disk Configuration (--create --config)

The `--create --config` action creates a quorum disk configuration.

The configuration must be created before any targets or devices can be created.

### Syntax

```
quorumdiskmgr --create --config [--owner owner --group group]
 --network-iface-list network-iface-list
```

### Parameters

The following table lists the parameters for the `--create --config` action:

Parameter	Description
<code>--owner</code>	Specifies the owner of the Oracle ASM instance to which the iSCSI quorum disks will be added. This is an optional parameter. The default value is <code>grid</code> .
<code>--group</code>	Specifies the group of the Oracle ASM instance to which the iSCSI quorum disks will be added. This is an optional parameter. The default value is <code>dba</code> .
<code>--network-iface-list</code>	Specifies the list of RDMA Network Fabric interface names through which the local and remote targets will be discovered.

### Example 2-7 Create a Quorum Disk Configuration for a System with InfiniBand Network Fabric

```
quorumdiskmgr --create --config --owner=oracle --group=dba --network-iface-
list="ib0, ib1"
```

### Example 2-8 Create a Quorum Disk Configuration for a System with RoCE Network Fabric

```
quorumdiskmgr --create --config --owner=oracle --group=dba --network-iface-
list="re0, re1"
```

## 2.14.4.4 Creating a Target (--create --target)

The `--create --target` action creates a target that will be used to create the devices to add to the specified Oracle ASM disk group.

The `--create --target` action creates a target that can be accessed by database servers with an RDMA Network Fabric IP address in the specified IP address list.

After a target is created, the `asm-disk-group`, `host-name`, and `size` attributes cannot be changed.

### Syntax

```
quorumdiskmgr --create --target --asm-disk-group asm_disk_group --
visible-to ip_list
 [--host-name host_name] [--size size]
```

### Parameters

Parameter	Description
<code>--asm-disk-group</code>	Specifies the Oracle ASM disk group to which the device created from the target will be added. The value of <code>asm-disk-group</code> is not case-sensitive.
<code>--visible-to</code>	Specifies a list of RDMA Network Fabric IP addresses. Database servers with an RDMA Network Fabric IP address in the list will have access to the target.
<code>--host-name</code>	Specifies the host name of the database server on which <code>quorumdiskmgr</code> runs. The total length of the values for <code>asm-disk-group</code> and <code>host-name</code> cannot exceed 26 characters. If the host name is too long, a shorter host name can be specified as long as a different host name is specified for each database server in the rack.  This is an optional parameter. The default value is the host name of the database server on which <code>quorumdiskmgr</code> runs. The value of <code>host-name</code> is not case-sensitive.
<code>--size</code>	Specifies the size of the target. This is an optional parameter. The default value is 128 MB.

### Example 2-9 Creating a Target For Oracle ASM Disk Group Devices

This example shows how to create a target for devices added to the `DATA1` disk group. That target is only visible to database servers that have an RDMA Network Fabric IP address of 192.168.10.45 or 192.168.10.46.

```
quorumdiskmgr --create --target --asm-disk-group=data1 --visible-
to="192.168.10.45, 192.168.10.46"
 --host-name=db01
```

### 2.14.4.5 Creating a Device (--create --device)

The `--create --device` action creates devices by discovering and logging into targets on database servers with an RDMA Network Fabric IP address in the specified list of IP addresses.

The created devices will be automatically discovered by the Oracle ASM instance with the owner and group specified during configuration creation.

#### Syntax

```
quorumdiskmgr --create --device --target-ip-list target_ip_list
```

#### Parameters

- `--target-ip-list`: Specifies a list of RDMA Network Fabric IP addresses.  
`quorumdiskmgr` discovers targets on database servers that have an IP address in the list, then logs in to those targets to create devices.

#### Example

##### Example 2-10 Creating Devices From a Target For an Oracle ASM Disk Group

This example shows how to create devices using targets on database servers that have an IP address of 192.168.10.45 or 192.168.10.46.

```
quorumdiskmgr --create --device --target-ip-list="192.168.10.45,
192.168.10.46"
```

### 2.14.4.6 Listing Quorum Disk Configurations (--list --config)

The `--list --config` action lists the quorum disk configurations.

#### Syntax

```
quorumdiskmgr --list --config
```

#### Sample Output

##### Example 2-11 Listing the quorum disk configuration on rack with InfiniBand Network Fabric

```
$ quorumdiskmgr --list --config
Owner: grid
Group: dba
ifaces: exadata_ib1 exadata_ib0
```

**Example 2-12 Listing the quorum disk configuration on rack with RoCE Network Fabric**

```
$ quorumdiskmgr --list --config
Owner: grid
Group: dba
ifaces: exadata_re1 exadata_re0
```

### 2.14.4.7 Listing Targets (--list --target)

The `--list --target` action lists the attributes of targets.

The target attributes listed include target name, size, host name, Oracle ASM disk group name, the list of IP addresses (a `visible-to` IP address list) indicating which database servers have access to the target, and the list of IP addresses (a `discovered-by` IP address list) indicating which database servers have logged into the target.

If an Oracle ASM disk group name is specified, the action lists all local targets created for the specified Oracle ASM disk group. Otherwise, the action lists all local targets created for quorum disks.

**Syntax**

```
quorumdiskmgr --list --target [--asm-disk-group asm_disk_group]
```

**Parameters**

- `--asm-disk-group`: Specifies the Oracle ASM disk group. `quorumdiskmgr` displays all local targets for this Oracle ASM disk group. The value of `asm-disk-group` is not case-sensitive.

**Example 2-13 Listing the Target Attributes for a Specific Oracle ASM Disk Group**

This example shows how to list the attributes of the target for the `DATA1` disk group.

```
quorumdiskmgr --list --target --asm-disk-group=data1
Name: iqn.2015-05.com.oracle:qd--data1_db01
Size: 128 MB
Host name: DB01
ASM disk group name: DATA1
Visible to: iqn.1988-12.com.oracle:192.168.10.23,
iqn.1988-12.com.oracle:192.168.10.24,
iqn.1988-12.com.oracle:1b48248af770, iqn.1988-12.com.oracle:7a4a399566
Discovered by: 192.168.10.47, 192.168.10.46
```



**Note:**

For systems installed using a release prior to Oracle Exadata System Software 19.1.0, the Name might appear as `iqn.2015-05.com.oracle:QD_DATAAC1_DB01`. Also, the Visible to field displays IP addresses instead of names.

### 2.14.4.8 Listing Devices (--list --device)

The `--list --device` action lists the attributes of devices, including device path, size, host name and ASM disk group name.

- If only the Oracle ASM disk group name is specified, then the output includes all the devices that have been added to the Oracle ASM disk group.
- If only the host name is specified, then the output includes all the devices created from the targets on the host.
- If both an Oracle ASM disk group name and a host name are specified, then the output includes a single device created from the target on the host that has been added to the Oracle ASM disk group.
- If neither an Oracle ASM disk group name or a host name is specified, then the output includes all quorum disk devices.

**Syntax**

```
quorumdiskmgr --list --device [--asm-disk-group asm_disk_group] [--host-name host_name]
```

**Parameters**

Parameter	Description
<code>--asm-disk-group</code>	Specifies the Oracle ASM disk group to which devices have been added. The value of <code>asm-disk-group</code> is not case-sensitive.
<code>--host-name</code>	Specifies the host name of the database server from whose targets devices are created. The value of <code>host-name</code> is not case-sensitive.

**Example 2-14 Listing Device Attributes for an Oracle ASM Disk Group**

This example shows how to list the attributes for devices used by the `DATAAC1` disk group.

```
$ quorumdiskmgr --list --device --asm-disk-group dataac1
Device path: /dev/exadata_quorum/QD_DATAAC1_DB01
Size: 128 MB
Host name: DB01
ASM disk group name: DATAAC1

Device path: /dev/exadata_quorum/QD_DATAAC1_DB02
Size: 128 MB
Host name: DB02
ASM disk group name: DATAAC1
```



### 2.14.4.9 Deleting Configurations (--delete --config)

The `--delete --config` action deletes quorum disk configurations.

The configurations can only be deleted when there are no targets or devices present.

#### Syntax

```
quorumdiskmgr --delete --config
```

### 2.14.4.10 Deleting Targets (--delete --target)

The `--delete --target` action deletes the targets created for quorum disks on database servers.

If an Oracle ASM disk group name is specified, then this command deletes all the local targets created for the specified Oracle ASM disk group. Otherwise, this command deletes all local targets created for quorum disks.

#### Syntax

```
quorumdiskmgr --delete --target [--asm-disk-group asm_disk_group]
```

#### Parameters

- `--asm-disk-group`: Specifies the Oracle ASM disk group. Local targets created for this disk group will be deleted.

The value of `asm-disk-group` is not case-sensitive.

#### Example 2-15 Deleting Targets Created for an Oracle ASM Disk Group

This example shows how to delete targets created for the `DATA1` disk group.

```
quorumdiskmgr --delete --target --asm-disk-group=data1
```

### 2.14.4.11 Deleting Devices (--delete --device)

The `--delete --device` command deletes quorum disk devices.

- If only an Oracle ASM disk group name is specified, then the command deletes all the devices that have been added to the Oracle ASM disk group.
- If only a host name is specified, then the command deletes all the devices created from the targets on the host.
- If both an Oracle ASM disk group name and a host name are specified, then the command deletes a single device created from the target on the host and that has been added to the Oracle ASM disk group.
- If neither an Oracle ASM disk group name nor a host name is specified, then the command deletes all quorum disk devices.

**Syntax**

```
quorumdiskmgr --delete --device [--asm-disk-group asm_disk_group] [--host-name host_name]
```

**Parameters**

Parameter	Description
--asm-disk-group	Specifies the Oracle ASM disk group whose device you want to delete. The value of <code>asm-disk-group</code> is not case-sensitive.
--host-name	Specifies the host name of the database server. Devices created from targets on this host will be deleted. The value of <code>host-name</code> is not case-sensitive.

**Example 2-16 Deleting Quorum Disk Devices Created from Targets on a Specific Host**

This example shows how to delete all the quorum disk devices that were created from the targets on the host `DB01`.

```
quorumdiskmgr --delete --device --host-name=db01
```

### 2.14.4.12 Changing Owner and Group Values (--alter --config)

The `--alter --config` action changes the owner and group configurations.

**Syntax**

```
quorumdiskmgr --alter --config --owner owner --group group
```

**Parameters**

Parameter	Description
--owner	Specifies the new owner for the quorum disk configuration. This parameter is optional. If not specified, the owner is unchanged.
--group	Specifies the new group for the quorum disk configuration. This parameter is optional. If not specified, the group is unchanged.

**Example 2-17 Changes the Owner and Group Configuration for Quorum Disk Devices**

This example shows how to change the assigned owner and group for quorum disk devices.

```
quorumdiskmgr --alter --config --owner=grid --group=dba
```

### 2.14.4.13 Changing the RDMA Network Fabric IP Addresses (--alter --target)

The `--alter --target` command changes the RDMA Network Fabric IP addresses of the database servers that have access to the local target created for the specified Oracle ASM disk group.

#### Syntax

```
quorumdiskmgr --alter --target --asm-disk-group asm_disk_group --
visible-to ip_list
```

#### Parameters

Parameter	Description
<code>--asm-disk-group</code>	Specifies the Oracle ASM disk group to which the device created from the target will be added. The value of <code>asm-disk-group</code> is not case-sensitive.
<code>--visible-to</code>	Specifies a list of RDMA Network Fabric IP addresses. Database servers with an RDMA Network Fabric IP address in the list will have access to the target.

#### Example 2-18 Changing the RDMA Network Fabric IP Addresses for Accessing Targets

This example shows how to change the RDMA Network Fabric IP address list that determines which database servers have access to the local target created for `DATA1` disk group

```
quorumdiskmgr --alter --target --asm-disk-group=data1 --visible-
to="192.168.10.45, 192.168.10.47"
```

## 2.14.5 Add Quorum Disks to Database Nodes

You can add quorum disks to database nodes on an Oracle Exadata Rack with fewer than five storage servers that contains a high redundancy disk group.

Oracle strongly recommends having quorum disks in all high redundancy disk groups with less than five failure groups. Having five quorum disks is important to mirror ASM metadata in any high redundancy disk group, and not just for the disk group housing the voting files.

The example in this section creates quorum disks for an Oracle Exadata Rack that has two database servers: `db01` and `db02`.

Typically, there are two RDMA Network Fabric ports on each database server:

- For systems with InfiniBand Network Fabric the ports are: `ib0` and `ib1`
- For systems with RoCE Network Fabric the ports are: `re0` and `re1`

On each cluster node, the network interfaces to be used for communication with the iSCSI devices can be found using the following command:

```
$ oifcfg getif | grep cluster_interconnect | awk '{print $1}'
```

The IP address of each interface can be found using the following command:

```
ip addr show interface_name
```

The RDMA Network Fabric IP addresses for this example are as follows:

On db01:

- Network interface: `ib0` or `re0`, IP address: 192.168.10.45
- Network interface: `ib1` or `re1`, IP address: 192.168.10.46

On db02:

- Network interface: `ib0` or `re0`, IP address: 192.168.10.47
- Network interface: `ib1` or `re1`, IP address: 192.168.10.48

The Oracle ASM disk group to which the quorum disks will be added is `DATA1`. The Oracle ASM owner is `grid`, and the user group is `dba`.

In this example, we will move the voting files from the normal redundancy disk group `RECO1` to `DATA1` after it has been augmented with quorum disks to yield five failure groups. The example shows the cluster voting files moving from `RECO1` to `DATA1`, but there is no need to relocate voting files if you are just adding quorum disks to a high redundancy disk group and you already have your voting files in some other high redundancy disk group.

Initially, the voting files reside on a normal redundancy disk group `RECO1`:

```
$ Grid_home/bin/crsctl query css votedisk
STATE File Universal Id File Name Disk group
-- -
 1. ONLINE 21f5507a28934f77bf3b7ecf88b26c47 (o/
192.168.76.187;192.168.76.188/RECO1_CD_00_celadm12) [RECO1]
 2. ONLINE 387f71ee81f14f38bfbd0693451e328 (o/
192.168.76.189;192.168.76.190/RECO1_CD_00_celadm13) [RECO1]
 3. ONLINE 6f7fab62e6054fb8bf167108cdbc2f64 (o/
192.168.76.191;192.168.76.192/RECO1_CD_00_celadm14) [RECO1]
Located 3 voting disk(s).
```

1. Log into the database servers (for example, `db01` and `db02`) as the `root` user.
2. Check if quorum disks are already configured on your system. If so, you can skip the next step.

Run the following `quorumdiskmgr` command on the database servers.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --config
```

If quorum disks are already configured on your system, your output should resemble one of the following:

- For Oracle Exadata System Software release 18.x or earlier, the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib1 exadata_ib0
```

- If you have upgraded to Oracle Exadata System Software release 19.1.0 or later from an earlier release, then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib0
Initiator name: iqn.1988-12.com.oracle:7a4a399566
```

- If you have a system that was imaged with Oracle Exadata System Software release 19.1.0 or later (not upgraded), then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib0
Initiator name: iqn.1988-12.com.oracle:192.168.18.205
```

- If your rack uses RoCE Network Fabric, then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_re0
Initiator name: iqn.1988-12.com.oracle:192.168.18.205
```

3. If the previous step didn't show an existing quorum disk configuration, run the `quorumdiskmgr` command with the `--create --config` options to create quorum disk configurations on the database servers.

- For systems with InfiniBand Network Fabric:

```
/opt/oracle.SupportTools/quorumdiskmgr --create --config --
owner=grid --group=dba --network-iface-list="ib0, ib1"
```

- For systems with RoCE Network Fabric:

```
/opt/oracle.SupportTools/quorumdiskmgr --create --config --
owner=grid --group=dba --network-iface-list="re0, re1"
```

4. Run the `quorumdiskmgr` command with the `--list --config` options to verify that the configurations have been successfully created on the database servers.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --config
```

Refer to step 2 for examples of the expected output.

5. Run the `quorumdiskmgr` command with the `--create --target` options to create a target on the database servers for Oracle ASM disk group `DATA1`, and make the target visible to the database servers.

In this example scenario, the following command would be run on the database servers db01 and db02:

```
/opt/oracle.SupportTools/quorumdiskmgr --create --target --asm-disk-
group=datac1
--visible-to="192.168.10.45, 192.168.10.46, 192.168.10.47, 192.168.10.48"
```

6. Run the `quorumdiskmgr` command with the `--list --target` options to verify the target has been successfully created on the database servers.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --target
```

- If you are running Oracle Exadata System Software release 18.x or earlier, then the output should look like this for each node:

```
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB01
Size: 128 MB
Host name: DB01
ASM disk group name: DATAAC1
Visible to: 192.168.10.45, 192.168.10.46, 192.168.10.47, 192.168.10.48
Discovered by:
```

```
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB02
Size: 128 MB
Host name: DB02
ASM disk group name: DATAAC1
Visible to: 192.168.10.45, 192.168.10.46, 192.168.10.47, 192.168.10.48
Discovered by:
```

- If you are running Oracle Exadata System Software release 19.x or later, then the output should look like this for each node:

```
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB01
Size: 128 MB
Host name: DB01
ASM disk group name: DATAAC1
Visible to: 192.168.10.45, 192.168.10.46, 192.168.10.47,
192.168.10.48, iqn.1988-12.com.oracle:ee657eb81b53,
iqn.1988-12.com.oracle:db357ba82b24
```

```
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB02
Size: 128 MB
Host name: DB02
ASM disk group name: DATAAC1
Visible to: 192.168.10.45, 192.168.10.46, 192.168.10.47,
192.168.10.48, iqn.1988-12.com.oracle:ee657eb81b53,
iqn.1988-12.com.oracle:db357ba82b24
```

 **Note:**

The output shows both IP addresses and initiator names in the `Visible to list` only if you have a system that was upgraded from a release older than Oracle Exadata System Software release 19.1.0. Otherwise, the `Visible to list` shows only IP addresses in it.

7. Run the `quorumdiskmgr` command with the `--create --device` options to create devices on the database servers using the previously created targets.

In this example scenario, the following command would be run on the database servers `db01` and `db02`:

```
/opt/oracle.SupportTools/quorumdiskmgr --create --device --target-
ip-list="192.168.10.45, 192.168.10.46,
192.168.10.47, 192.168.10.48"
```

8. Run the `quorumdiskmgr` command with the `--list --device` options to verify the devices have been successfully created on the database servers.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --device
```

For this example scenario, the output should look like:

```
Device path: /dev/exadata_quorum/QD_DATAAC1_DB01
Size: 128 MB
Host name: DB01
ASM disk group name: DATAAC1

Device path: /dev/exadata_quorum/QD_DATAAC1_DB02
Size: 128 MB
Host name: DB02
ASM disk group name: DATAAC1
```

9. Switch to the `grid` user on one of the database servers (for example, `db01` or `db02`) and set the environment to access Oracle ASM.
10. Alter the `asm_diskstring` initialization parameter and add `/dev/exadata_quorum/*` to the existing string:

For example:

```
SQL> ALTER SYSTEM SET asm_diskstring='o/*/DATAAC1_*','o/*/
RECO1_*','/dev/exadata_quorum/*' scope=both sid='*';
```

11. Verify that the quorum disk devices are automatically discovered by Oracle ASM.

For example:

```
SQL> set linesize 200
SQL> col path format a50
SQL> SELECT inst_id, label, path, mode_status, header_status
FROM gv$asm_disk WHERE path LIKE '/dev/exadata_quorum/%';
```

The output should look like:

```

INST_ID LABEL PATH MODE_STATUS
HEADER_STATUS

1 QD_DATAAC1_DB01 /dev/exadata_quorum/QD_DATAAC1_DB01 ONLINE
CANDIDATE
1 QD_DATAAC1_DB02 /dev/exadata_quorum/QD_DATAAC1_DB02 ONLINE
CANDIDATE
2 QD_DATAAC1_DB01 /dev/exadata_quorum/QD_DATAAC1_DB01 ONLINE
CANDIDATE
2 QD_DATAAC1_DB02 /dev/exadata_quorum/QD_DATAAC1_DB02 ONLINE
CANDIDATE

```

**12.** Add the quorum disk devices to a high redundancy Oracle ASM disk group.

If there is no high redundancy disk group, create a high redundancy disk group and include the new quorum disks. For example:

```

SQL> CREATE DISKGROUP DATAAC1 HIGH REDUNDANCY ADD
 QUORUM FAILGROUP db01 DISK '/dev/exadata_quorum/QD_DATAAC1_DB01'
 QUORUM FAILGROUP db02 DISK '/dev/exadata_quorum/QD_DATAAC1_DB02' ...

```

If a high redundancy disk group already exists, add the new quorum disks. For example:

```

SQL> ALTER DISKGROUP DATAAC1 ADD
 QUORUM FAILGROUP db01 DISK '/dev/exadata_quorum/QD_DATAAC1_DB01'
 QUORUM FAILGROUP db02 DISK '/dev/exadata_quorum/QD_DATAAC1_DB02';

```

 **Tip:**

The failure group name should match the last part of the device path following `QD_disk_group_name_`. In this example, the failure group names are `DB01` and `DB02`.

**13.** Verify that the status of the quorum disks in Oracle ASM is changed from `CANDIDATE` to `MEMBER`.

For example:

```

SQL> set linesize 200
SQL> col path format a50
SQL> SELECT inst_id, label, path, mode_status, header_status
 FROM gv$asm_disk WHERE path LIKE '/dev/exadata_quorum/%';

```

The output should look like:

```

INST_ID LABEL PATH MODE_STATUS
HEADER_STATUS


```



```

1 QD_DATAAC1_DB01 /dev/exadata_quorum/QD_DATAAC1_DB01
ONLINE MEMBER
1 QD_DATAAC1_DB02 /dev/exadata_quorum/QD_DATAAC1_DB02
ONLINE MEMBER
2 QD_DATAAC1_DB01 /dev/exadata_quorum/QD_DATAAC1_DB01
ONLINE MEMBER
2 QD_DATAAC1_DB02 /dev/exadata_quorum/QD_DATAAC1_DB02
ONLINE MEMBER

```

- 14.** If you are moving voting files from a normal redundancy disk group to a high redundancy disk group with 5 five quorum disks, you can now relocate the voting files.

For example:

```
$ Grid_home/bin/crsctl replace votedisk +DATAAC1
```

- 15.** Verify that the voting disks have been successfully relocated to the high redundancy disk group and that five voting files exist.

```
$ crsctl query css votedisk
```

In this example, the output shows three voting disks on storage servers and two voting disks on database servers:

```

STATE File Universal Id File Name Disk group
-- -----
1. ONLINE ca2f1b57873f4ff4bf1dfb78824f2912 (o/192.168.10.42/
DATAAC1_CD_09_celadm12) [DATAAC1]
2. ONLINE a8c3609a3dd44f53bf17c89429c6ebe6 (o/192.168.10.43/
DATAAC1_CD_09_celadm13) [DATAAC1]
3. ONLINE cafb7e95a5be4f00bf10bc094469cad9 (o/192.168.10.44/
DATAAC1_CD_09_celadm14) [DATAAC1]
4. ONLINE 4dca8fb7bd594f6ebf8321ac23e53434 (/dev/exadata_quorum/QD_
DATAAC1_DB01) [DATAAC1]
5. ONLINE 4948b73db0514f47bf94ee53b98fdb51 (/dev/exadata_quorum/QD_
DATAAC1_DB02) [DATAAC1]
Located 5 voting disk(s).

```

- 16.** Move the Oracle ASM password file to the high redundancy disk group.

- a.** Get the source Oracle ASM password file location.

```
$ asmcmd pwget --asm
```

- b.** Move the Oracle ASM password file to the high redundancy disk group.

```
$ asmcmd pmove --asm full_path_of_source_file
full_path_of_destination_file
```

For example:

```
$ asmcmd pmove --asm +recocl/ASM/PASSWORD/pwdasm.256.898960531
+dataac1/aspwdfile
```

**17.** Move the Oracle ASM SPFILE to the high redundancy disk group.

- a.**
- Get the Oracle ASM SPFILE in use.

```
$ asmcmd spget
```

- b.**
- Copy the Oracle ASM SPFILE to the high redundancy disk group.

```
$ asmcmd sscopy full_path_of_source_file full_path_of_destination_file
```

- c.**
- Modify the Oracle Grid Infrastructure configuration to use the relocated SPFILE upon next restart.

```
$ asmcmd spset full_path_of_destination_file
```

- d.**
- If it can be tolerated, restart Oracle Grid Infrastructure.

```
Grid_home/bin/crsctl stop crs
```

```
Grid_home/bin/crsctl start crs
```

If restarting Oracle Grid Infrastructure is not permitted, repeat this step (17) every time the Oracle ASM SPFILE is modified until Oracle Grid Infrastructure is restarted.

**18.** Relocate the MGMTDB to the high redundancy disk group.

Move the MGMTDB (if running) to the high redundancy disk group using [How to Move/Recreate GI Management Repository to Different Shared Storage \(Diskgroup, CFS or NFS etc\) \(My Oracle Support Doc ID 1589394.1\)](#).

Configure the MGMTDB to not use hugepages using the steps below:

```
export ORACLE_SID=-MGMTDB
export ORACLE_HOME=$GRID_HOME
sqlplus "sys as sysdba"
SQL> ALTER SYSTEM SET use_large_pages=false scope=spfile sid='*';
```

**19.** Optional: Restart Oracle Grid Infrastructure.

```
Grid_home/bin/crsctl stop crs
```

```
Grid_home/bin/crsctl start crs
```

**20.** Optional: Convert the normal redundancy disk group to a high redundancy disk group.

Refer to [How To Move The Database To Different Diskgroup \(Change Diskgroup Redundancy\) \(My Oracle Support Doc ID 438580.1\)](#) for details on how to change disk group redundancy.

**Related Topics**

- [Exadata Oracle Linux 6 to Oracle Linux 7 Quorumdisk Migration FAQ \(My Oracle Support Doc ID 2453054.1\)](#)

## 2.14.6 Recreate Quorum Disks

In certain circumstances, you might need to recreate a quorum disk.

Some examples of when you might need to recreate a quorum disk are:

- When recreating a guest domU
  - If you deleted the quorum disks without first dropping the quorum disks from the Oracle ASM disk group
1. Force drop the lost quorum disk.

```
ALTER DISKGROUP dg_name DROP QUORUM DISK disk_name FORCE;
```

2. Follow the instructions in "Adding Quorum Disks to Database Nodes" to add a new quorum disk.

### Related Topics

- [Add Quorum Disks to Database Nodes](#)  
You can add quorum disks to database nodes on an Oracle Exadata Rack with fewer than five storage servers that contains a high redundancy disk group.

## 2.14.7 Use Cases

The following topics describe various configuration cases when using the quorum disk manager utility.

- [New Deployments on Oracle Exadata 12.1.2.3.0 or Later](#)
- [Upgrading to Oracle Exadata Release 12.1.2.3.0 or Later](#)
- [Downgrading to a Pre-12.1.2.3.0 Oracle Exadata Release](#)
- [Managing Quorum Disks When Changing Elastic Configurations](#)  
When modifying the elastic configuration of an Oracle Exadata Rack, you might have to perform additional actions if you use quorum disks.

### 2.14.7.1 New Deployments on Oracle Exadata 12.1.2.3.0 or Later

For new deployments on Oracle Exadata release 12.1.2.3.0 and above, OEDA implements this feature by default when all of the following requirements are satisfied:

- The system has at least two database nodes and fewer than five storage servers.
- You are running OEDA release February 2016 or later.
- You meet the software requirements listed in [Software Requirements for Quorum Disk Manager](#).
- Oracle Database is 11.2.0.4 and above.
- The system has at least one high redundancy disk group.

If the system has three storage servers in place, then two quorum disks will be created on the first two database nodes of the cluster picked by OEDA.

If the system has four storage servers in place, then one quorum disk will be created on the first database node picked by OEDA.

## 2.14.7.2 Upgrading to Oracle Exadata Release 12.1.2.3.0 or Later

If the target Exadata system has fewer than five storage servers, at least one high redundancy disk group, and two or more database nodes, you can implement this feature manually using `quorumdiskmgr`.

### Related Topics

- [Add Quorum Disks to Database Nodes](#)

## 2.14.7.3 Downgrading to a Pre-12.1.2.3.0 Oracle Exadata Release

Rolling back to a pre-12.1.2.3.0 Oracle Exadata release, which does not support quorum disks, from a release that supports quorum disks, which is any release 12.1.2.3.0 and later, requires quorum disk configuration to be removed if the environment has quorum disk implementation in place. You need to remove the quorum disk configuration before performing the Exadata software rollback.

To remove quorum disk configuration, perform these steps:

1. Ensure there is at least one normal redundancy disk group in place. If not, create one.
2. Relocate the voting files to a normal redundancy disk group:

```
$GI_HOME/bin/crsctl replace votedisk +normal_redundancy_diskgroup
```

3. Drop the quorum disks from ASM. Run the following command for each quorum disk:

```
SQL> alter diskgroup diskgroup_name drop quorum disk quorum_disk_name force;
```

Wait for the rebalance operation to complete. You can tell it is complete when `v$asm_operation` returns no rows for the disk group.

4. Delete the quorum devices. Run the following command from each database node that has quorum disks in place:

```
/opt/oracle.SupportTools/quorumdiskmgr --delete --device [--asm-disk-group
asm_disk_group] [--host-name host_name]
```

5. Delete the targets. Run the following command from each database node that has quorum disks in place:

```
/opt/oracle.SupportTools/quorumdiskmgr --delete --target [--asm-disk-group
asm_disk_group]
```

6. Delete the configuration. Run the following command from each database node that has quorum disks in place:

```
/opt/oracle.SupportTools/quorumdiskmgr --delete -config
```

## 2.14.7.4 Managing Quorum Disks When Changing Elastic Configurations

When modifying the elastic configuration of an Oracle Exadata Rack, you might have to perform additional actions if you use quorum disks.

- [Adding a Database Node if Using Quorum Disks](#)
- [Removing a Database Node When Using Quorum Disks](#)

If database node being removed hosted a quorum disk containing a voting file and there are fewer than five storage servers in the Oracle Real Application Clusters (Oracle RAC)

cluster, then a quorum disk must be created on a different database node before the database node is removed.

- [Adding an Oracle Exadata Storage Server and Expanding an Existing High Redundancy Disk Group](#)
- [Removing an Oracle Exadata Storage Server When Using Quorum Disks](#)

#### 2.14.7.4.1 Adding a Database Node if Using Quorum Disks

If the existing Oracle Real Application Clusters (Oracle RAC) cluster has fewer than two database nodes and fewer than five storage servers, and the voting files are not stored in a high redundancy disk group, then Oracle recommends adding quorum disks to the database node(s) and relocating the voting files to a high redundancy disk group.



#### Note:

The requirements listed in "[Software Requirements for Quorum Disk Manager](#)" must be met.

If the existing Oracle RAC cluster already has quorum disks in place, the quorum disks need to be made visible to the newly added node prior to adding the node to the Oracle RAC cluster using the `addnode.sh` procedure.

1. Log in to the 2 database nodes that contain the quorum devices as the `root` user.
2. Retrieve the quorum disk iSCSI target configuration.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --target
```

The output of this command should be similar to the following (where the host name is `db01` and the `diskgroup_name` is `DATA`):

```
Name: iqn.2015-05.com.oracle:QD_DATA_DB01
Host name: DB01
ASM disk group name: DATA
Size: 128 MB
Visible to: IP_address1, IP_address2, IP_address3, IP_address4...
IP_address2n
Discovered by: IP_address1, IP_address2, IP_address3, IP_address4
```

*IP\_address1, IP\_address2, IP\_address3, IP\_address4...IP\_address2n* above refer to the IP addresses of the RDMA Network Fabric interfaces of all the existing cluster nodes. In the example above, the number of number of nodes in the cluster is *n*.

3. Modify the target in each node to make the device target visible to the node being added.

In this command, use the IP addresses shown in the `visibleToList` field in the previous step in `IP_list` and append the list with the IP addresses of the node being added.

```
/opt/oracle.SupportTools/quorumdiskmgr --alter --target --asm-disk-group
asm_diskgroupname --visible-to 'IP_list,
IP_addressX, IP_addressY'
```

`IP_addressX` and `IP_addressY` in the previous command refer to the IP addresses of the 2 RDMA Network Fabric interfaces of the node being added.

4. Run `/opt/oracle.SupportTools/quorumdiskmgr --list -target` in the two database nodes that contain the quorum devices and make sure the two IP addresses of the node being added are seen in the `Visible to list`.
5. Log in as the `root` user on the node being added.
6. Run the `quorumdiskmgr` command with the `--list --config` option to verify that the configurations have been successfully created on the node.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --config
```

Your output should resemble one of the following:

- For Oracle Exadata System Software release 18.x or earlier, the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib1 exadata_ib0
```

- If you have upgraded to Oracle Exadata System Software release 19.1.0 or later from an earlier release, then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib0
Initiator name: iqn.1988-12.com.oracle:7a4a399566
```

- If you have a system that was imaged with Oracle Exadata System Software release 19.1.0 or later (not upgraded), then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib0
Initiator name: iqn.1988-12.com.oracle:192.168.18.205
```

- If your rack uses RoCE Network Fabric, then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_re0
Initiator name: iqn.1988-12.com.oracle:192.168.18.205
```

7. Run the `quorumdiskmgr` command with the `--create --device` option to create the quorum devices on the node being added pointing to the targets for the existing quorum devices.

In the following command, *IP\_List* is a command delimited list of the IP addresses obtained in step 4.

```
/opt/oracle.SupportTools/quorumdiskmgr --create --device --target-
ip-list='IP_List'
```

8. Run the `quorumdiskmgr` command with the `--list --device` options to verify the existing quorum devices have been successfully discovered and are visible on the node being added.

```
/opt/oracle.SupportTools/quorumdiskmgr --list -device
```

On the newly added node the output should be similar to the following and it should be the same as in any of the existing cluster nodes:

```
Device path: /dev/exadata_quorum/QD_DATAAC1_DB01
Size: 128 MB
Host name: DB01
ASM disk group name: DATA
Device path: /dev/exadata_quorum/QD_DATAAC1_DB02
Size: 128 MB
Host name: DB02
ASM disk group name: DATA
```

### Related Topics

- [Add Quorum Disks to Database Nodes](#)  
You can add quorum disks to database nodes on an Oracle Exadata Rack with fewer than five storage servers that contains a high redundancy disk group.

## 2.14.7.4.2 Removing a Database Node When Using Quorum Disks

If database node being removed hosted a quorum disk containing a voting file and there are fewer than five storage servers in the Oracle Real Application Clusters (Oracle RAC) cluster, then a quorum disk must be created on a different database node before the database node is removed.

If the database node being removed did not host a quorum disk, then no action is required. Otherwise, use these steps to create a quorum disk on a database node that does not currently host a quorum disk.

1. Log into `db01` and `db02` as the `root` user.
2. Run the `quorumdiskmgr` command with the `--create --config` options to create quorum disk configurations on both `db01` and `db02`.
  - For InfiniBand Network Fabric:

```
/opt/oracle.SupportTools/quorumdiskmgr --create --config --
owner=grid --group=dba --network-iface-list="ib0, ib1"
```

- For RoCE Network Fabric:

```
/opt/oracle.SupportTools/quorumdiskmgr --create --config --
owner=grid --group=dba --network-iface-list="re0, re1"
```

3. Run the `quorumdiskmgr` command with the `--list --config` options to verify that the configurations have been successfully created on both `db01` and `db02`.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --config
```

Your output should resemble one of the following:

- For Oracle Exadata System Software release 18.x or earlier, the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib1 exadata_ib0
```

- If you have upgraded to Oracle Exadata System Software release 19.1.0 or later from an earlier release, then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib0
Initiator name: iqn.1988-12.com.oracle:7a4a399566
```

- If you have a system that was imaged with Oracle Exadata System Software release 19.1.0 or later (not upgraded), then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_ib0
Initiator name: iqn.1988-12.com.oracle:192.168.18.205
```

- If your rack uses RoCE Network Fabric, then the output should look like this:

```
Owner: grid
Group: dba
ifaces: exadata_re0
Initiator name: iqn.1988-12.com.oracle:192.168.18.205
```

4. Run the `quorumdiskmgr` command with the `--create --target` options to create a target on both `db01` and `db02` for Oracle ASM disk group `DATA1` and make the target visible to both `db01` and `db02`.

For example, you would use a command similar to this:

```
/opt/oracle.SupportTools/quorumdiskmgr --create --target --asm-disk-
group=data1
--visible-to="192.168.10.45, 192.168.10.46, 192.168.10.47, 192.168.10.48"
```



5. Run the `quorumdiskmgr` command with the `--list --target` options to verify the target has been successfully created on both `db01` and `db02`.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --target
```

The output shows both IP addresses and initiator names in the `Visible to list` only if you have a system that was upgraded from a release older than Oracle Exadata System Software release 19.1.0. Otherwise, the `Visible to list` shows only IP addresses in it.

- If you are running Oracle Exadata System Software release 18.x or earlier, then the output should look like this for each node:

```
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB01
Size: 128 MB
Host name: DB01
ASM disk group name: DATAAC1
Visible to: 192.168.10.45, 192.168.10.46, 192.168.10.47,
192.168.10.48
Discovered by:
```

```
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB02
Size: 128 MB
Host name: DB02
ASM disk group name: DATAAC1
Visible to: 192.168.10.45, 192.168.10.46, 192.168.10.47,
192.168.10.48
Discovered by:
```

- If you are running Oracle Exadata System Software release 19.x or later, then the output should look like this for each node:

```
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB01
Size: 128 MB
Host name: DB01
ASM disk group name: DATAAC1
Visible to: 192.168.10.45, 192.168.10.46, 192.168.10.47,
192.168.10.48, iqn.1988-12.com.oracle:ee657eb81b53,
iqn.1988-12.com.oracle:db357ba82b24
```

```
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB02
Size: 128 MB
Host name: DB02
ASM disk group name: DATAAC1
Visible to: 192.168.10.45, 192.168.10.46, 192.168.10.47,
192.168.10.48, iqn.1988-12.com.oracle:ee657eb81b53,
iqn.1988-12.com.oracle:db357ba82b24
```

6. Run the `quorumdiskmgr` command with the `--create --device` options to create devices on both `db01` and `db02` from targets on both `db01` and `db02`.

For example, you would use a command similar to this:

```
/opt/oracle.SupportTools/quorumdiskmgr --create --device --target-ip-
list="192.168.10.45, 192.168.10.46,
192.168.10.47, 192.168.10.48"
```

7. Run the `quorumdiskmgr` command with the `--list --device` options to verify the devices have been successfully created on both `db01` and `db02`.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --device
```

On both `db01` and `db02`, the output should look like:

```
Device path: /dev/exadata_quorum/QD_DATAAC1_DB01
Size: 128 MB
Host name: DB01
ASM disk group name: DATAAC1
```

```
Device path: /dev/exadata_quorum/QD_DATAAC1_DB02
Size: 128 MB
Host name: DB02
ASM disk group name: DATAAC1
```

8. Add the two quorum disk devices to a high redundancy Oracle ASM disk group.

If there is no high redundancy disk group, create a high redundancy disk group and include the two new quorum disks. For example:

```
SQL> CREATE DISKGROUP DATAAC1 HIGH REDUNDANCY ADD QUORUM FAILGROUP db01
DISK '/dev/exadata_quorum/QD_DATAAC1_DB01'
QUORUM FAILGROUP db02 DISK '/dev/exadata_quorum/QD_DATAAC1_DB02' ...
```

If a high redundancy disk group already exists, add the two new quorum disks. For example:

```
SQL> ALTER DISKGROUP dataac1 ADD QUORUM FAILGROUP db01 DISK '/dev/
exadata_quorum/QD_DATAAC1_DB01'
QUORUM FAILGROUP db02 DISK '/dev/exadata_quorum/QD_DATAAC1_DB02';
```

9. Remove the database node.

After the database node is removed, its voting file will get relocated automatically to the quorum disk added in the previous steps.

### 2.14.7.4.3 Adding an Oracle Exadata Storage Server and Expanding an Existing High Redundancy Disk Group

When you add a storage server that uses quorum disks, Oracle recommends relocating a voting file from a database node to the newly added storage server.

1. Add the Exadata storage server. See [Adding a Cell Node](#) for details.  
In the example below, the new storage server added is called "celadm04".
2. After the storage server is added, verify the new fail group from `v$asm_disk`.

```
SQL> select distinct failgroup from v$asm_disk;
FAILGROUP

ADM01
ADM02
CELADM01
CELADM02
CELADM03
CELADM04
```

### 3. Verify at least one database node has a quorum disk containing a voting file.

```
$ crsctl query css votedisk
STATE File Universal Id File Name Disk group
-- -----
 1. ONLINE 834ee5a8f5054f12bf47210c51ecb8f4 (o/
192.168.12.125;192.168.12.126/DATAC5_CD_00_celadm01) [DATAC5]
 2. ONLINE f4af2213d9964f0bbfa30b2ba711b475 (o/
192.168.12.127;192.168.12.128/DATAC5_CD_00_celadm02) [DATAC5]
 3. ONLINE ed61778df2964f37bf1d53ea03cd7173 (o/
192.168.12.129;192.168.12.130/DATAC5_CD_00_celadm03) [DATAC5]
 4. ONLINE bfe1c3aa91334f16bf78ee7d33ad77e0 (/dev/exadata_quorum/
QD_DATAC5_ADM01) [DATAC5]
 5. ONLINE a3a56e7145694f75bf21751520b226ef (/dev/exadata_quorum/
QD_DATAC5_ADM02) [DATAC5]
Located 5 voting disk(s).
```

The example above shows there are two quorum disks with voting files on two database nodes.

### 4. Drop one of the quorum disks.

```
SQL> alter diskgroup datac5 drop quorum disk QD_DATAC5_ADM01;
```

The voting file on the dropped quorum disk will be relocated automatically to the newly added storage server by the Grid Infrastructure as part of the voting file refresh. You can verify this as follows:

```
$ crsctl query css votedisk
STATE File Universal Id File Name Disk group
-- -----
 1. ONLINE 834ee5a8f5054f12bf47210c51ecb8f4 (o/
192.168.12.125;192.168.12.126/DATAC5_CD_00_celadm01) [DATAC5]
 2. ONLINE f4af2213d9964f0bbfa30b2ba711b475 (o/
192.168.12.127;192.168.12.128/DATAC5_CD_00_celadm02) [DATAC5]
 3. ONLINE ed61778df2964f37bf1d53ea03cd7173 (o/
192.168.12.129;192.168.12.130/DATAC5_CD_00_celadm03) [DATAC5]
 4. ONLINE a3a56e7145694f75bf21751520b226ef (/dev/exadata_quorum/
QD_DATAC5_ADM02) [DATAC5]
 5. ONLINE ab5aefd60cf84fe9bff6541b16e33787 (o/
192.168.12.131;192.168.12.132/DATAC5_CD_00_celadm04) [DATAC5]
```

## 2.14.7.4.4 Removing an Oracle Exadata Storage Server When Using Quorum Disks

If removing a storage server results in the number of storage servers being used by the Oracle RAC cluster to fewer than five, and the voting files reside in a high redundancy disk group, then Oracle recommends adding quorum disks to the database nodes, if not in place already.

Prior to removing the storage server, add the quorum disks so that five copies of the voting files are available immediately after removing the storage server.

**Related Topics**

- [Add Quorum Disks to Database Nodes](#)

## 2.14.8 Reconfigure Quorum Disk After Restoring a Database Server

After restoring a database server, `lvdisplay` shows the quorum disk was not restored.

When you restore a database server, Exadata image rescue mode restores the layout of disks and file systems, with the exception of custom partitions, including quorum disks. These files must be recreated after being restored from backup.

The logical volumes created for quorum disks are in `/dev/VGExaDb` and have the name-prefix `LVDbVd*`.

1. Using the configuration backed up under `/etc/lvm/archive`, make a logical volume (LV) for the quorum disk on every node.

For example, you would use a command similar to the following, but using values from the backup configuration information.

```
lvcreate -L 128MB -n <LVName> VGExaDb
```

2. Reboot all database servers.

```
shutdown -r now
```

3. After the servers restart, verify the quorum disks are restored.

```
/opt/oracle.SupportTools/quorumdiskmgr --list --config
Owner: grid
Group: dba
ifaces: exadata_ib1 exadata_ib0
```

```
/opt/oracle.SupportTools/quorumdiskmgr --list --target
Name: iqn.2015-05.com.oracle:QD_DATAAC1_DB01
Host name: DB01
ASM disk group name: DATAAC1
Size: 128 MB
Visible to: 192.168.10.45, 192.168.10.46
Discovered by: 192.168.10.45, 192.168.10.46
```

```
/opt/oracle.SupportTools/quorumdiskmgr --list --device
Device path: /dev/exadata_quorum/QD_DATAAC1_DB01
Host name: DB01
ASM disk group name: DATAAC1
Size: 128 MB
```

```
Device path: /dev/exadata_quorum/QD_DATAAC1_DB01
Host name: DB01
ASM disk group name: DATAAC1
Size: 128 MB
```

4. Query the cluster to see if all voting disks are available.

```
crsctl query css votedisk
STATE File Universal Id File
Name Disk group
-- -

1. ONLINE ca2f1b57873f4ff4bf1dfb78824f2912 (o/192.168.10.42/
DATA_C1_CD_09_celadm12) [DATA_C1]
2. ONLINE a8c3609a3dd44f53bf17c89429c6ebe6 (o/192.168.10.43/
DATA_C1_CD_09_celadm13) [DATA_C1]
3. ONLINE 4948b73db0514f47bf94ee53b98fdb51 (/dev/
exadata_quorum/QD_DATA_C1_DB02) [DATA_C1]
4. ONLINE cafb7e95a5be4f00bf10bc094469cad9 (o/192.168.10.44/
DATA_C1_CD_09_celadm14) [DATA_C1]
Located 4 voting disk(s).
```

Notice that there is one voting disk missing, for the recovered database server (DB01).

If you query `V$ASM_DISK`, you can see that the quorum disk was offlined by the recovery process.

```
SQL> SELECT label, path, mode_status, header_status, mount_status
 FROM v$asm_disk
 WHERE path LIKE '/dev/%';
```

```
LABEL PATH
MODE_STATUS
HEADER_STATUS MOUNT_S

QD_DATA_C1_DB01 /dev/exadata_quorum/QD_DATA_C1_
ONLINE
CANDIDATE CLOSED

QD_DATA_C1_DB02 /dev/exadata_quorum/QD_DATA_C1_
ONLINE
MEMBER CACHED
```

5. Drop the unavailable quorum disk from the Oracle ASM disk group using the `FORCE` option.

This step is only required if the previous step shows the quorum disk as part of the ASM disk group in the `V$ASM_DISK` query output.

```
SQL> alter diskgroup DATA_C1 drop quorum disk QD_DATA_C1_DB01 force;
```

6. Add the quorum disk back to the Oracle ASM disk group.

```
SQL> alter diskgroup DATA_C1 add quorum failgroup DB01
 disk '/dev/exadata_quorum/QD_DATA_C1_DB01';
```

If the command fails with an error indicating that the disk already belongs to the disk group, run the command with the `FORCE` option.

```
SQL> alter diskgroup DATA_C1 add quorum failgroup DB01
 disk '/dev/exadata_quorum/QD_DATAAC1_DB01' force;
```

7. Requery `v$asm_disk` to verify that the quorum disks are available.

```
SQL> SELECT label, path, mode_status, header_status, mount_status
 FROM v$asm_disk
 WHERE path LIKE '/dev/%';
```

LABEL	PATH	MODE_STATUS
QD_DATAAC1_DB01	/dev/exadata_quorum/QD_DATAAC1_	ONLINE
MEMBER		CACHED
QD_DATAAC1_DB02	/dev/exadata_quorum/QD_DATAAC1_	ONLINE
MEMBER		CACHED

8. Query the cluster to verify all voting disks are now available.

```
crsctl query css votedisk
STATE File Universal Id File
Name Disk group
-- -

 1. ONLINE ca2f1b57873f4ff4bf1dfb78824f2912 (o/192.168.10.42/
DATAAC1_CD_09_celadm12) [DATAAC1]
 2. ONLINE a8c3609a3dd44f53bf17c89429c6ebe6 (o/192.168.10.43/
DATAAC1_CD_09_celadm13) [DATAAC1]
 3. ONLINE 4948b73db0514f47bf94ee53b98fdb51 (/dev/exadata_quorum/QD_
DATAAC1_DB02) [DATAAC1]
 4. ONLINE cafb7e95a5be4f00bf10bc094469cad9 (o/192.168.10.44/
DATAAC1_CD_09_celadm14) [DATAAC1]
 5. ONLINE 4dca8fb7bd594f6ebf8321ac23e53434 (/dev/exadata_quorum/QD_
DATAAC1_DB01) [DATAAC1]
Located 5 voting disk(s).
```

## 2.15 Using vmetrics

The vmetrics package enables you to display system statistics gathered by the vmetrics service.

- [About the vmetrics Package](#)  
The vmetrics service collects the statistics required for SAP monitoring of Oracle VM domains.
- [Installing and Starting the vmetrics Service](#)
- [Files in the vmetrics Package](#)
- [Displaying the Statistics](#)

- [Adding Metrics to vmetrics](#)

## 2.15.1 About the vmetrics Package

The vmetrics service collects the statistics required for SAP monitoring of Oracle VM domains.

You can access the system statistics from the management domain (dom0) or the user domain (domU). The vmetrics service runs on the management domain, collects the statistics, and pushes them to the xenstore. This allows the user domains to access the statistics.

System statistics collected by the vmetrics service are shown below, with sample values:

```
com.sap.host.host.VirtualizationVendor=Oracle Corporation;
com.sap.host.host.VirtProductInfo=Oracle VM 3;
com.sap.host.host.PagedInMemory=0;
com.sap.host.host.PagedOutMemory=0;
com.sap.host.host.PageRates=0;
com.sap.vm.vm.uuid=2b80522b-060d-47ee-8209-2ab65778eb7e;
com.sap.host.host.HostName=sc10adm01.example.com;
com.sap.host.host.HostSystemInfo=sc10adm01;
com.sap.host.host.NumberOfPhysicalCPUs=24;
com.sap.host.host.NumCPUs=4;
com.sap.host.host.TotalPhyMem=98295;
com.sap.host.host.UsedVirtualMemory=2577;
com.sap.host.host.MemoryAllocatedToVirtualServers=2577;
com.sap.host.host.FreeVirtualMemory=29788;
com.sap.host.host.FreePhysicalMemory=5212;
com.sap.host.host.TotalCPUTime=242507.220000;
com.sap.host.host.Time=1453150151;
com.sap.vm.vm.PhysicalMemoryAllocatedToVirtualSystem=8192;
com.sap.vm.vm.ResourceMemoryLimit=8192;
com.sap.vm.vm.TotalCPUTime=10160.1831404;
com.sap.vm.vm.ResourceProcessorLimit=4;
```

## 2.15.2 Installing and Starting the vmetrics Service

To install the vmetrics service, run the `install.sh` script as the root user on dom0:

```
[root@scac10adm01]# cd /opt/oracle.SupportTools/vmetrics
[root@scac10adm01]# ./install.sh
```

The `install.sh` script verifies that it is running on dom0, stops any vmetrics services currently running, copies the package files to `/opt/oracle.vmetrics`, and copies `vmetrics.svc` to `/etc/init.d`.

To start the vmetrics service on dom0, run the following command as the root user on dom0:

```
[root@scac10adm01 vmetrics]# service vmetrics.svc start
```

The commands to gather the statistics are run every 30 seconds.

## 2.15.3 Files in the vmetrics Package

The vmetrics package contains the following files:

File	Description
install.sh	This file installs the package.
vm-dump-metrics	This script reads the statistics from the xenstore and displays them in XML format.
vmetrics	This Python script runs the system commands and uploads them to the xenstore. The system commands are listed in the <code>vmetrics.conf</code> file.
vmetrics.conf	This XML file specifies the metrics that the dom0 should push to the xenstore, and the system commands to run for each metric.
vmetrics.svc	The <code>init.d</code> file that makes vmetrics a Linux service.

## 2.15.4 Displaying the Statistics

Once the statistics have been pushed to the xenstore, you can view the statistics on dom0 and domU by running either of the following commands:

### Note:

On domU's, ensure that the `xenstoreprovider` and `ovmd` packages are installed.

`xenstoreprovider` is the library which communicates with the `ovmapi` kernel infrastructure.

`ovmd` is a daemon that handles configuration and reconfiguration events and provides a mechanism to send/receive messages between the VM and the Oracle VM Manager.

The following command installs the necessary packages on Oracle Linux 5 and 6 to support the Oracle VM API.

```
yum install ovmd xenstoreprovider
```

- The `/usr/sbin/ovmd -g vmhost` command displays the statistics on one line. The `sed` command breaks up the line into multiple lines, one statistic per line. You need to run this command as the root user.

```
root@scac10db01vm04 ~]# /usr/sbin/ovmd -g vmhost |sed 's/; */;\n/g;s/:"/:"\n/g'
com.sap.host.host.VirtualizationVendor=Oracle Corporation;
com.sap.host.host.VirtProductInfo=Oracle VM 3;
com.sap.host.host.PagedInMemory=0;
com.sap.host.host.PagedOutMemory=0;
com.sap.host.host.PageRates=0;
com.sap.vm.vm.uuid=2b80522b-060d-47ee-8209-2ab65778eb7e;
com.sap.host.host.HostName=scac10adm01.example.com;
com.sap.host.host.HostSystemInfo=scac10adm01;
com.sap.host.host.NumberOfPhysicalCPUs=24;
com.sap.host.host.NumCPUs=4;
...
```



- The `vm-dump-metrics` command displays the metrics in XML format.

```
[root@scac10db01vm04 ~]# ./vm-dump-metrics
<metrics>
<metric type='real64' context='host'>
<name>TotalCPUTime</name>
<value>242773.600000</value>
</metric>
<metric type='uint64' context='host'>
<name>PagedOutMemory</name>
<value>0</value>
</metric>
...
```

Note that you have copy the `vm-dump-metrics` command to the domU's from which you want to run the command.

## 2.15.5 Adding Metrics to vmetrics

You can add your own metric to be collected by the vmetrics service.

1. In `/opt/oracle.SupportTools/vmetrics/vmetrics.conf`, add the new metric and the system commands to retrieve and parse that metric. For example:

```
<metric type="uint32" context="host">
 <name>NumCPUs</name>
 <action>grep -c processor /proc/cpuinfo</action>
 <action2>xm list | grep '^Domain-0' |awk '{print $4}'</action2>
</metric>
```

In the `<name>` element, enter the name of the new metric.

In the `<action>` and `<action2>` elements, specify the system command for the new metric. You only need to have `<action2>`, but you can use `<action>` as a fallback in case `<action2>` does not work on some systems.

Note that any action that needs the name of the vm should be done with `scas07client07vm01`. When vmetrics runs, it swaps out this dummy name for the actual domU names that are running in the dom0.

2. In `/opt/oracle.SupportTools/vmetrics/vmetrics`, add the metric in the list `gFieldsList`. Prefix the metric name with "host" if the metric is about the host (dom0) or with "vm" if the metric is about the vm (domU). For example:

Suppose the `gFieldsList` looks like this:

```
gFieldsList = ['host.VirtualizationVendor',
 'host.VirtProductInfo',
 'host.PagedInMemory',
 'vm.ResourceProcessorLimit']
```

If you are adding a new metric called "NumCPUs" (as shown in the example in step 1), and this metric is intended to tell the domU how many cpu's the dom0 has available, then `gFieldsList` would now look like:

```
gFieldsList = ['host.VirtualizationVendor',
 'host.VirtProductInfo',
 'host.PagedInMemory',
 'vm.ResourceProcessorLimit',
 'host.NumCPUs']
```

3. (optional) In `/opt/oracle.SupportTools/vmetrics/vm-dump-metrics`, add the new metric if you want the new metric to be included in the XML output.

If you skip this step, you can view the new metric using the `ovmd -g vmhost` command.

## 2.16 Using FIPS mode

On database servers running Oracle Linux 7, you can enable the kernel to run in FIPS mode.

Starting with Oracle Exadata System Software release 20.1.0, you can enable and disable the Federal Information Processing Standards (FIPS) compatibility mode on Exadata Database Machine database servers running Oracle Linux 7.

After you enable or disable FIPS mode, you must reboot the server for the action to take effect.

To enable, disable, and get status information about FIPS mode, use the utility at `/opt/oracle.cellos/host_access_control` with the `fips-mode` option:

- To display the current FIPS mode setting, run:

```
/opt/oracle.cellos/host_access_control fips-mode --status
```

- To enable FIPS mode, run:

```
/opt/oracle.cellos/host_access_control fips-mode --enable
```

Then, reboot the server to finalize the action.

- To disable FIPS mode, run:

```
/opt/oracle.cellos/host_access_control fips-mode --disable
```

Then, reboot the server to finalize the action.

- To display information about FIPS mode, run:

```
/opt/oracle.cellos/host_access_control fips-mode --info
```

The following example shows the typical command sequence and command output for enabling and disabling FIPS mode on a server.

```
/opt/oracle.cellos/host_access_control fips-mode --status
[2020-04-14 09:19:45 -0700] [INFO] [IMG-SEC-1101] FIPS mode is disabled

/opt/oracle.cellos/host_access_control fips-mode --enable
[2020-04-14 09:30:10 -0700] [INFO] [IMG-SEC-1107] Using only FIPS compliant
SSH host keys and sshd configuration updated in /etc/ssh/sshd_config
[2020-04-14 09:30:10 -0700] [INFO] [IMG-SEC-1103] FIPS mode is set to
enabled. A reboot is required to effect this change.

/opt/oracle.cellos/host_access_control fips-mode --status
[2020-04-14 09:30:14 -0700] [INFO] [IMG-SEC-1101] FIPS mode is configured but
not activated. A reboot is required to activate.
```

```
reboot

...

/opt/oracle.cellos/host_access_control fips-mode --status
[2020-04-14 09:23:15 -0700] [INFO] [IMG-SEC-1103] FIPS mode is
configured and
active

/opt/oracle.cellos/host_access_control fips-mode --disable
[2020-04-14 09:40:37 -0700] [INFO] [IMG-SEC-1103] FIPS mode is set to
disabled. A reboot is required to effect this change.

/opt/oracle.cellos/host_access_control fips-mode --status
[2020-04-14 09:40:37 -0700] [INFO] [IMG-SEC-1103] FIPS mode is
disabled but
is active. A reboot is required to deactivate FIPS mode.

reboot

...

/opt/oracle.cellos/host_access_control fips-mode --status
[2020-04-14 09:46:22 -0700] [INFO] [IMG-SEC-1101] FIPS mode is disabled
```

## 2.17 LED Status Descriptions

The LEDs on the Oracle Exadata Rack components help you identify the component that needs servicing.

- [Exadata Database Server X5-2 and Later LEDs](#)
- [Sun Server X4-2 Oracle Database Server LEDs](#)
- [Sun Server X3-2 Oracle Database Server LEDs](#)
- [Sun Fire X4170 M2 Oracle Database Server LEDs](#)
- [Sun Fire X4170 Oracle Database Server LEDs](#)
- [Exadata Database Server X5-8 and Later LEDs](#)
- [Sun Server X4-8 Oracle Database Server LEDs](#)
- [Sun Server X2-8 Oracle Database Server LEDs](#)
- [Sun Fire X4800 Oracle Database Server LEDs](#)

### 2.17.1 Exadata Database Server X5-2 and Later LEDs

The following table describes the color codes of the LEDs on Oracle Server X5-2 and later Oracle Database Servers.

**Table 2-4 Oracle Server X5-2 and Later Oracle Database Server LED Status Descriptions**

Component	LED Status
Fan module	<ul style="list-style-type: none"> <li>• Top Fan Fault LED is off: The system is powered on and the fan module is functioning correctly.</li> <li>• Top Fan Fault LED is amber: The fan module is faulty. The front and rear panel Fault-Service Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Power supply	<ul style="list-style-type: none"> <li>• AC OK LED is green: Power supply can be removed during a hot-swap procedure.</li> <li>• Fault-Service Required LED is amber: The power supply is faulty. The front and rear panel Fault-Service Required LEDs are also lit if the system detects a power supply fault.</li> </ul>
Server chassis	<ul style="list-style-type: none"> <li>• OK to Remove LED is blue: The storage device can be removed safely during a hot-swap procedure.</li> <li>• Fault-Service Required LED is amber: The system is running but the storage device is faulty. The front and rear panel Fault-Service Required LEDs are also lit if the system detects a storage device fault.</li> <li>• OK/Activity LED is green: Data is being read from or written to the storage device.</li> </ul>

The following table describes the storage device status based on the LEDs.

**Table 2-5 Storage Device Status of Oracle Server X5-2 and later Oracle Database Server Based on LEDs**

LED	Predictive Failure	Critical
Service Action Required (Amber)	On	On
OK to Remove (Blue)	On	On

## 2.17.2 Sun Server X4-2 Oracle Database Server LEDs

[Table 2-6](#) describes the color codes of the LEDs on Sun Server X4-2 Oracle Database Servers.

**Table 2-6 Sun Fire X4170 M2 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Server X4-2 Oracle Database Server fan module	<ul style="list-style-type: none"> <li>Fan Status LED is green: The system is powered on and the fan module is functioning correctly.</li> <li>Fan Status LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Sun Server X4-2 Oracle Database Server power supply	<ul style="list-style-type: none"> <li>OK to Remove LED is green: The power supply can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> </ul>
Sun Server X4-2 Oracle Database Server servers	<ul style="list-style-type: none"> <li>OK to Remove LED is blue: The storage drive can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The system is running but the storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>

Table 2-7 describes the disk status based on the LEDs.

**Table 2-7 Disk Status of Sun Server X4-2 Oracle Database Server Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	On	On	On
OK to Remove (Blue)	Off	On	On

### 2.17.3 Sun Server X3-2 Oracle Database Server LEDs

Table 2-8 describes the color codes of the LEDs on Sun Server X3-2 Oracle Database Servers.

**Table 2-8 Sun Server X3-2 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Server X3-2 Oracle Database Server fan module	<ul style="list-style-type: none"> <li>Fan Status LED is green: The system is powered on and the fan module is functioning correctly.</li> <li>Fan Status LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Sun Server X3-2 Oracle Database Server power supply	<ul style="list-style-type: none"> <li>OK to Remove LED is green: The power supply can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> </ul>
Sun Server X3-2 Oracle Database Server servers	<ul style="list-style-type: none"> <li>OK to Remove LED is blue: The storage drive can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The system is running but the storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>

[Table 2-9](#) describes the disk status based on the LEDs.

**Table 2-9 Disk Status of Sun Server X3-2 Oracle Database Server Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	On	On	On
OK to Remove (Blue)	Off	On	On

## 2.17.4 Sun Fire X4170 M2 Oracle Database Server LEDs

[Table 2-10](#) describes the color codes of the LEDs on Sun Fire X4170 M2 Oracle Database Servers.

**Table 2-10 Sun Fire X4170 M2 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Fire X4170 M2 Oracle Database Server fan module	<ul style="list-style-type: none"> <li>Fan Status LED is green: The system is powered on and the fan module is functioning correctly.</li> <li>Fan Status LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Sun Fire X4170 M2 Oracle Database Server power supply	<ul style="list-style-type: none"> <li>OK to Remove LED is green: The power supply can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> </ul>
Sun Fire X4170 M2 Oracle Database Server servers	<ul style="list-style-type: none"> <li>OK to Remove LED is blue: The storage drive can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The system is running but the storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>

[Table 2-11](#) describes the disk status based on the LEDs.

**Table 2-11 Disk Status of Sun Fire X4170 M2 Oracle Database Server Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	On	On	On
OK to Remove (Blue)	Off	On	On

## 2.17.5 Sun Fire X4170 Oracle Database Server LEDs

[Table 2-12](#) describes the color codes of the LEDs on Sun Fire X4170 Oracle Database Servers.

**Table 2-12 Sun Fire X4170 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Fire X4170 Oracle Database Server fan module	<ul style="list-style-type: none"> <li>Power/OK LED is green: The system is powered on and the fan module is functioning correctly.</li> <li>Service Action Required LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Sun Fire X4170 Oracle Database Server power supply	<ul style="list-style-type: none"> <li>OK to Remove LED is green: The power supply can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> </ul>
Sun Fire X4170 Oracle Database Server servers	<ul style="list-style-type: none"> <li>OK to Remove LED is blue: A storage drive can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The system is running and the storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>

Table 2-13 describes the disk status based on the LEDs.

**Table 2-13 Disk Status of Sun Fire X4170 Oracle Database Server Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	Flashing	Flashing	On
OK to Remove (Blue)	Off	Flashing	Flashing

## 2.17.6 Exadata Database Server X5-8 and Later LEDs

The color codes of the LEDs on Exadata Database Server X5-8 and later models are as follows:

- Dual PCIe Card Carriers (DPCC):
  - Fault-Service Required LED is amber. If lit, there is a fault condition that requires service. This LED blinks rapidly when you use Oracle ILOM to activate the LED as a locator.
  - OK/Activity LED is green. A steady light indicates the PCIe is ready, and in use by the operating system. A flashing green light indicates the DPCC is booting. If the LED is not lit, then power is not present.



- Hard drives:
  - OK to Remove LED is blue. If this LED is lit, then the hard drive can safely be removed.
  - Fault-Service Required LED is amber. If this LED is lit, then a fault was detected in the hard drive. If flashing, this LED indicates the drive is not functioning correctly, which is related to a status of predictive failure, poor performance, or critical.
  - OK/Activity LED is green: This LED flashes to indicate drive activity. The rate at which the LED blinks can vary by activity. If the LED is a steady green, then the storage drive is functioning normally. If the LED is not lit, then power is not present or the Oracle ILOM boot is not complete.
- Power supplies:
  - PSU Fault-Service Required LED is amber. If this LED is lit, then a fault was detected in the power supply.
  - PSU OK LED is green. If the LED is a steady green, then the power supply is functioning normally.
  - AC OK LED is green. If the LED is a steady green, then the power supply is connected to a properly rated AC power source.
- Service processor (SP) modules:
  - Network Activity LEDs: A steady green light indicates a live network. This LED lights when the network port is active. If the LED is not lit, it indicates that there is no activity, and the link is not operational. When there is network traffic, the LED flashes green.
  - Network Speed LEDs: If the LED is not lit, it indicates a 10BASE-T link (if link up) (10 Gigabit Ethernet 10GBASE-T). If there is a steady amber light, it indicates a 100BASE-T link (Fast Ethernet 100 BASE-TX). If the light is a steady green light, it indicates a 1000BASE-T link (Gigabit Ethernet 1000BASE-T).
  - Chassis Fault-Service Required LED is amber. If this LED is lit, then a fault was detected in the system module.
  - Power/System OK LED: Green indicates full power mode. It flashes quickly during server start up and when AC power cords are connected to the server. It flashes slowly in standby power mode.
  - Temperature Fault LED is amber. If lit, this LED indicates the internal server temperature exceeds the upper threshold.
  - SP OK LED is green. When this LED is lit, it indicates the service processor and Oracle ILOM are operational. If the LED is flashing, it indicates the SP is booting.
  - Locator LED is white: This LED is lit when activated by Oracle ILOM or the recessed Locate button. This LED enables you to locate a server quickly and easily.

## 2.17.7 Sun Server X4-8 Oracle Database Server LEDs

[Table 2-14](#) describes the color codes of the LEDs on Sun Server X4-8 Oracle Database Server.

**Table 2-14 Sun Server X4-8 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Server X4-8 Oracle Database Server Back Panel PCIe Card Carriers (DPCC)	<ul style="list-style-type: none"> <li>• ATTN Button LED: Press to remove and to add the PCIe to the operating system.</li> <li>• Service Action Required (fault) LED is amber: There is a fault condition.</li> <li>• OK LED is green: PCIe is ready, and in use by the operating system.</li> </ul>
Sun Server X4-8 Oracle Database Server hard drives	<ul style="list-style-type: none"> <li>• Hot Swap LED is blue: The hard drive can safely be removed.</li> <li>• Fault is amber: The hard drive is faulty.</li> <li>• Activity is green: Flashing indicates drive activity, and standby.</li> </ul>
Sun Server X4-8 Oracle Database Server power supply	<ul style="list-style-type: none"> <li>• PSU Fault LED is amber: The power supply has a fault condition.</li> <li>• PSU OK LED is green: The power supply is on.</li> <li>• AC LED is green: AC is connected to the power supply.</li> </ul>
Sun Server X4-8 Oracle Database Server service processor (SP) modules	<ul style="list-style-type: none"> <li>• 10/1000/1000Base-T Ethernet LED (left): Green indicates link is established at 1 gigabit. Amber and on indicates link is established at 100 megabits. Amber and off indicates link is established at 10 megabits.</li> <li>• 10/1000/1000Base-T Ethernet LED (bottom) is green: There is activity on the link.</li> <li>• Chassis Service Action Required (fault) LED is amber: There is a fault condition.</li> <li>• Chassis Power OK LED: Green indicates full power mode. It flashes quickly during SP boot when AC power cords are connected to the server. It flashes slowly in standby power mode.</li> <li>• Chassis Over-temperature LED is amber: The internal server temperature exceeds the upper threshold.</li> <li>• SP module OK LED: It is green and on when SP is functional. It flashes green three times when SP module first receives power. It is yellow when SP is not functional.</li> <li>• Chassis Locate LED is white: It has been activated by the ILOM or Locate button.</li> </ul>

[Table 2-15](#) describes the disk status based on the LEDs.

**Table 2-15 Disk Status of Sun Server X4-8 Oracle Database Server Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	Flashing	Flashing	Flashing
OK to Remove (Blue)	Off	Off	Off

## 2.17.8 Sun Server X2-8 Oracle Database Server LEDs

Table 2-16 describes the color codes of the LEDs on Sun Server X2-8 Oracle Database Servers.

**Table 2-16 Sun Server X2-8 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Server X2-8 Oracle Database Server Back Panel PCIe Express Modules (EM)	<ul style="list-style-type: none"> <li>• ATTN Button LED: Press to remove and to add the EM to the operating system.</li> <li>• Service Action Required (fault) LED is amber: There is a fault condition.</li> <li>• OK LED is green: EM is ready, and in use by the operating system.</li> </ul>
Sun Server X2-8 Oracle Database Server hard drives	<ul style="list-style-type: none"> <li>• Hot Swap LED is blue: The hard drive can safely be removed.</li> <li>• Fault is amber: The hard drive is faulty.</li> <li>• Activity is green: Flashing indicates drive activity, and standby.</li> </ul>
Sun Server X2-8 Oracle Database Server Network Express Modules (NEM)	<ul style="list-style-type: none"> <li>• 10 Gb Ethernet port Activity LED: Green indicates link is established at 100 megabits. Amber indicates link is at 10 megabits, and not at full capacity.</li> <li>• 10 Gb Ethernet port Link LED is green: There is activity on the link.</li> <li>• 10/1000/1000Base-T Ethernet LED (top): Green indicates link is established at 1 gigabit. Amber and on indicates link is established at 100 megabits. Amber and off indicates link is established at 10 megabits.</li> <li>• 10/1000/1000Base-T Ethernet LED (bottom) is green: There is activity on the link.</li> <li>• NEM Locate LED is white: It has been activated by the ILOM or Locate button.</li> <li>• OK to Remove LED is blue: Not used.</li> <li>• Service Action Required (fault) LED is amber: There is a fault condition.</li> <li>• Power OK LED is green: The system is powered on.</li> </ul>
Sun Server X2-8 Oracle Database Server power supply	<ul style="list-style-type: none"> <li>• PSU Fault LED is amber: The power supply has a fault condition.</li> <li>• PSU OK LED is green: The power supply is on.</li> <li>• AC LED is green: AC is connected to the power supply.</li> </ul>

**Table 2-16 (Cont.) Sun Server X2-8 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Server X2-8 Oracle Database Server service processor (SP) modules	<ul style="list-style-type: none"> <li>10/1000/1000Base-T Ethernet LED (left): Green indicates link is established at 1 gigabit. Amber and on indicates link is established at 100 megabits. Amber and off indicates link is established at 10 megabits.</li> <li>10/1000/1000Base-T Ethernet LED (bottom) is green: There is activity on the link.</li> <li>Chassis Service Action Required (fault) LED is amber: There is a fault condition.</li> <li>Chassis Power OK LED: Green indicates full power mode. It flashes quickly during SP boot when AC power cords are connected to the server. It flashes slowly in standby power mode.</li> <li>Chassis Over-temperature LED is amber: The internal server temperature exceeds the upper threshold.</li> <li>SP module OK LED: It is green and on when SP is functional. It flashes green three times when SP module first receives power. It is yellow when SP is not functional.</li> <li>Chassis Locate LED is white: It has been activated by the ILOM or Locate button.</li> </ul>

Table 2-17 describes the disk status based on the LEDs.

**Table 2-17 Disk Status of Sun Server X2-8 Oracle Database Server Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	Flashing	Flashing	Flashing
OK to Remove (Blue)	Off	Off	Off

## 2.17.9 Sun Fire X4800 Oracle Database Server LEDs

Table 2-18 describes the color codes of the LEDs on Sun Fire X4800 Oracle Database Servers.

**Table 2-18 Sun Fire X4800 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Fire X4800 Oracle Database Server Back Panel PCIe Express Modules (EM)	<ul style="list-style-type: none"> <li>ATTN Button LED: Press to remove and to add the EM to the operating system.</li> <li>Service Action Required (fault) LED is amber: There is a fault condition.</li> <li>OK LED is green: EM is ready, and in use by the operating system.</li> </ul>

**Table 2-18 (Cont.) Sun Fire X4800 Oracle Database Server LED Status Descriptions**

Component	LED Status
Sun Fire X4800 Oracle Database Server hard drives	<ul style="list-style-type: none"> <li>• Hot Swap LED is blue: The hard drive can safely be removed.</li> <li>• Fault is amber: The hard drive is faulty.</li> <li>• Activity is green: Flashing indicates drive activity, and standby.</li> </ul>
Sun Fire X4800 Oracle Database Server Network Express Modules (NEM)	<ul style="list-style-type: none"> <li>• 10 Gb Ethernet port Activity LED: Green indicates link is established at 100 megabits. Amber indicates link is at 10 megabits, and not at full capacity.</li> <li>• 10 Gb Ethernet port Link LED is green: There is activity on the link.</li> <li>• 10/1000/1000Base-T Ethernet LED (top): Green indicates link is established at 1 gigabit. Amber and on indicates link is established at 100 megabits. Amber and off indicates link is established at 10 megabits.</li> <li>• 10/1000/1000Base-T Ethernet LED (bottom) is green: There is activity on the link.</li> <li>• NEM Locate LED is white: It has been activated by the ILOM or Locate button.</li> <li>• OK to Remove LED is blue: Not used.</li> <li>• Service Action Required (fault) LED is amber: There is a fault condition.</li> <li>• Power OK LED is green: The system is powered on.</li> </ul>
Sun Fire X4800 Oracle Database Server power supply	<ul style="list-style-type: none"> <li>• PSU Fault LED is amber: The power supply has a fault condition.</li> <li>• PSU OK LED is green: The power supply is on.</li> <li>• AC LED is green: AC is connected to the power supply.</li> </ul>
Sun Fire X4800 Oracle Database Server service processor (SP) modules	<ul style="list-style-type: none"> <li>• 10/1000/1000Base-T Ethernet LED (left): Green indicates link is established at 1 gigabit. Amber and on indicates link is established at 100 megabits. Amber and off indicates link is established at 10 megabits.</li> <li>• 10/1000/1000Base-T Ethernet LED (bottom) is green: There is activity on the link.</li> <li>• Chassis Service Action Required (fault) LED is amber: There is a fault condition.</li> <li>• Chassis Power OK LED: Green indicates full power mode. It flashes quickly during SP boot when AC power cords are connected to the server. It flashes slowly in standby power mode.</li> <li>• Chassis Over-temperature LED is amber: The internal server temperature exceeds the upper threshold.</li> <li>• SP module OK LED: It is green and on when SP is functional. It flashes green three times when SP module first receives power. It is yellow when SP is not functional.</li> <li>• Chassis Locate LED is white: It has been activated by the ILOM or Locate button.</li> </ul>

Table 2-19 describes the disk status based on the LEDs.

**Table 2-19 Disk Status of Sun Fire X4800 Oracle Database Server Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	Flashing	Flashing	Flashing
OK to Remove (Blue)	Off	Off	Off

## 2.18 Exadata Database Server Images

The Exadata database server models have different external layouts and physical appearance.

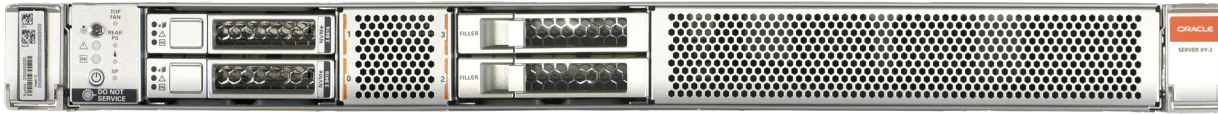
- [Oracle Server X9-2 Database Server Images](#)  
Oracle Server X9-2 is used as the database server in Exadata Database Machine X9M-2.
- [Oracle Server X8-2 Database Server Images](#)  
Oracle Server X8-2 is used as the database server in Exadata Database Machine X8M-2 and X8-2.
- [Oracle Server X7-2 Oracle Database Server Images](#)
- [Oracle Server X6-2 Oracle Database Server Images](#)
- [Oracle Server X5-2 Oracle Database Server Images](#)
- [Sun Server X4-2 Oracle Database Server Images](#)
- [Sun Server X3-2 Oracle Database Server Images](#)
- [Sun Fire X4170 M2 Oracle Database Server Images](#)
- [Sun Fire X4170 Oracle Database Server Images](#)
- [Oracle Server X8-8 Database Server Images](#)  
Oracle Server X8-8 is used as the database server in Exadata Database Machine X9M-8, X8M-8, and X8-8.
- [Oracle Server X7-8 Oracle Database Server Images](#)
- [Oracle Server X5-8 and X6-8 Oracle Database Server Images](#)
- [Sun Server X4-8 Oracle Database Server Images](#)
- [Sun Server X2-8 Oracle Database Server Images](#)
- [Sun Fire X4800 Oracle Database Server Images](#)

### 2.18.1 Oracle Server X9-2 Database Server Images

Oracle Server X9-2 is used as the database server in Exadata Database Machine X9M-2.

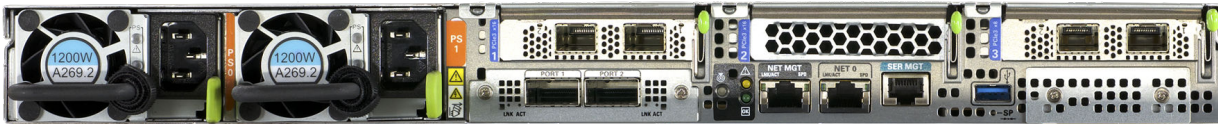
The following image shows the front view of Oracle Server X9-2 Database Servers.

**Figure 2-3 Front View of Oracle Server X9-2 Database Servers**



The following image shows a rear view of Oracle Server X9-2. This image shows a server with two dual-port 25 Gb/s network interface cards (in PCI slot 1 and slot3).

**Figure 2-4 Rear View of Oracle Server X9-2 Database Servers**



## 2.18.2 Oracle Server X8-2 Database Server Images

Oracle Server X8-2 is used as the database server in Exadata Database Machine X8M-2 and X8-2.

The following image shows the front view of Oracle Server X8-2 Database Servers.

**Figure 2-5 Front View of Oracle Server X8-2 Database Servers**



The following image shows the rear view of the Oracle Server.

**Figure 2-6 Rear View of Oracle Server X8-2 Database Servers**



## 2.18.3 Oracle Server X7-2 Oracle Database Server Images

The following image shows the front view of Oracle Server X7-2 Oracle Database Server.

**Figure 2-7 Front View of Oracle Server X7-2 Oracle Database Server**





The following image shows the rear view of Oracle Server.

**Figure 2-8 Rear View of X7-2 Oracle Database Server**



## 2.18.4 Oracle Server X6-2 Oracle Database Server Images

The following image shows the front view of Oracle Server X6-2 Oracle Database Server.

**Figure 2-9 Front View of Oracle Server X6-2 Oracle Database Server**



The following image shows the rear view of Oracle Server X6-2 Oracle Database Server.

The top hard disk drives are, from left to right HDD1, and HDD3. The lower drives are, from left to right, HDD0, and HDD2.

**Figure 2-10 Rear View of Oracle Server X6-2 Oracle Database Server**



## 2.18.5 Oracle Server X5-2 Oracle Database Server Images

The following image shows the front view of Oracle Server X5-2 Oracle Database Server.

**Figure 2-11 Front View of Oracle Server X5-2 Oracle Database Server**



The following image shows the rear view of Oracle Server X5-2 Oracle Database Server.

The top hard disk drives are, from left to right HDD1, and HDD3. The lower drives are, from left to right, HDD0, and HDD2.



**Figure 2-12 Rear View of Oracle Server X5-2 Oracle Database Server**



## 2.18.6 Sun Server X4-2 Oracle Database Server Images

The following image shows the front view of Sun Server X4-2 Oracle Database Server.

**Figure 2-13 Front View of Sun Server X4-2 Oracle Database Server**



The following image shows the rear view of Sun Server X4-2 Oracle Database Server.

**Figure 2-14 Rear View of Sun Server X4-2 Oracle Database Server**



## 2.18.7 Sun Server X3-2 Oracle Database Server Images

The following image shows the front view of Sun Server X3-2 Oracle Database Server.

**Figure 2-15 Front View of Sun Server X3-2 Oracle Database Server**



The following image shows the rear view of Sun Server X3-2 Oracle Database Server.

**Figure 2-16 Rear View of Sun Server X3-2 Oracle Database Server**



## 2.18.8 Sun Fire X4170 M2 Oracle Database Server Images

The following image shows the front view of Sun Fire X4170 M2 Oracle Database Server.

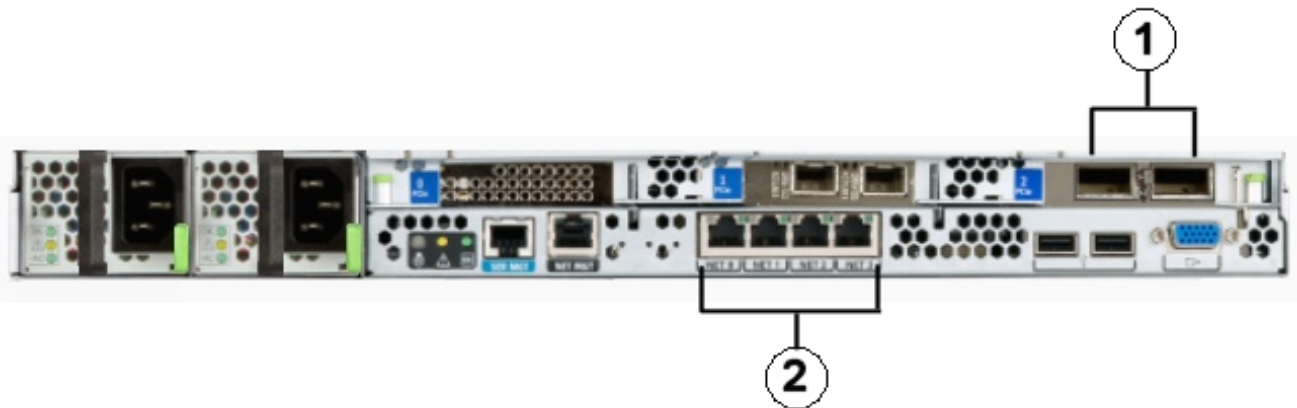
**Figure 2-17 Front View of Sun Fire X4170 M2 Oracle Database Server**



1. Hard disk drives. The top drives are, from left to right HDD1, and HDD3. The lower drives are, from left to right, HDD0, and HDD2.

The following image shows the rear view of Sun Fire X4170 M2 Oracle Database Server.

**Figure 2-18 Rear View of Sun Fire X4170 M2 Oracle Database Server**



1. InfiniBand host channel adapter
2. Gigabit Ethernet ports

## 2.18.9 Sun Fire X4170 Oracle Database Server Images

The following image shows the front view of Sun Fire X4170 Oracle Database Server.

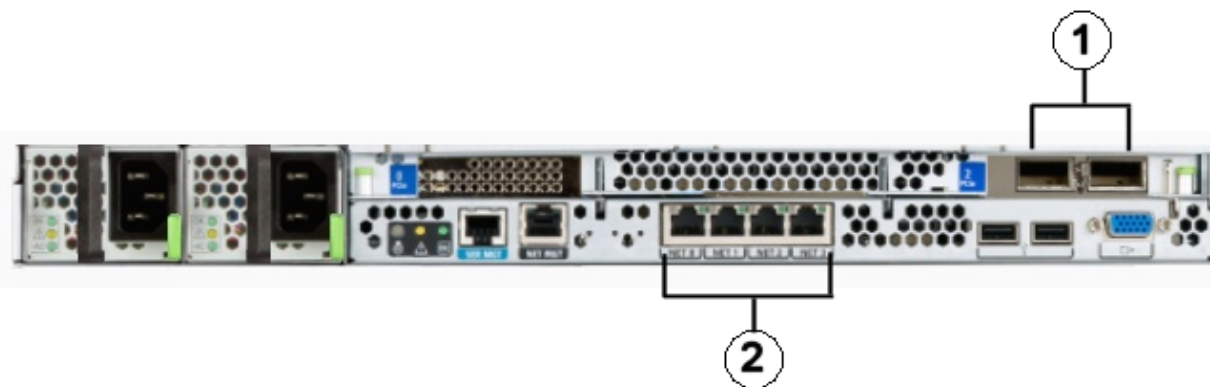
Figure 2-19 Front View of Sun Fire X4170 Oracle Database Server



1. Hard disk drives. The top drives are, from left to right HDD1, and HDD3. The lower drives are, from left to right, HDD0, and HDD2.

The following image shows the rear view of Sun Fire X4170 Oracle Database Server.

Figure 2-20 Rear View of Sun Fire X4170 Oracle Database Server



1. RDMA Network Fabric host channel adapter
2. Gigabit Ethernet ports

## 2.18.10 Oracle Server X8-8 Database Server Images

Oracle Server X8-8 is used as the database server in Exadata Database Machine X9M-8, X8M-8, and X8-8.

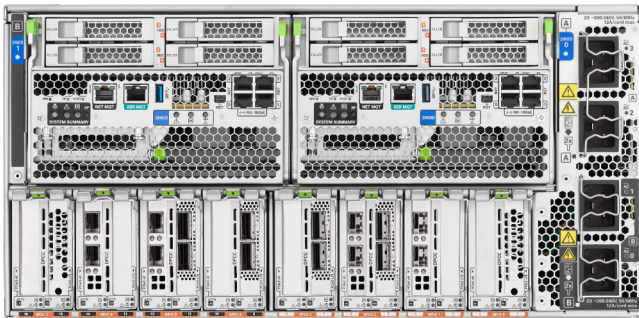
The following image shows the front view of Oracle Server X8-8 Database Server.

**Figure 2-21 Front View of Oracle Database Server X8-8**



The following image shows the rear view of Oracle Database Server X8-8.

**Figure 2-22 Rear View of Oracle Database Server X8-8**



## 2.18.11 Oracle Server X7-8 Oracle Database Server Images

The following image shows the front view of Oracle Server X7-8 Oracle Database Server.

**Figure 2-23 Front View of Oracle Server X7-8 Oracle Database Server**



The following image shows the rear view of Oracle Server X7-8 Oracle Database Server.



**Figure 2-24 Rear View of Oracle Server X7-8 Oracle Database Server**



## 2.18.12 Oracle Server X5-8 and X6-8 Oracle Database Server Images

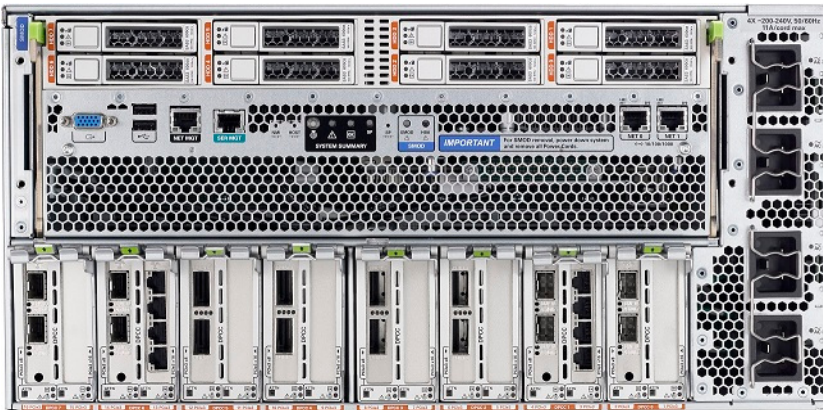
The following image shows the front view of Oracle Server X5-8 Oracle Database Server.

**Figure 2-25 Front View of Oracle Server X5-8 Oracle Database Server**



The following image shows the back view of Oracle Server X5-8 Oracle Database Server.

**Figure 2-26 Back View of Oracle Server X5-8 Oracle Database Server**



### 2.18.13 Sun Server X4-8 Oracle Database Server Images

The following image shows the front view of Sun Server X4-8 Oracle Database Server.

**Figure 2-27 Front View of Sun Server X4-8 Oracle Database Server**



The following image shows the rear view of Sun Server X4-8 Oracle Database Server.

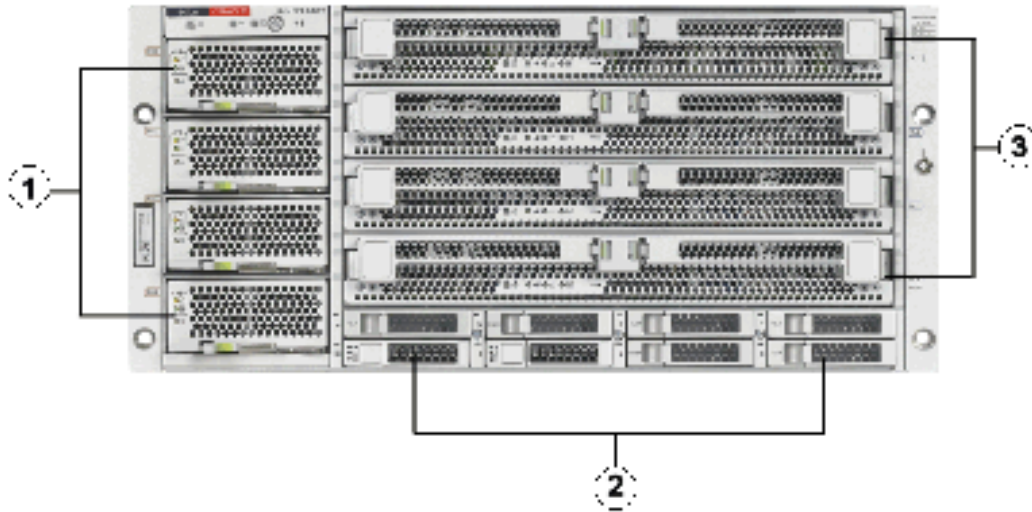
**Figure 2-28 Rear View of Sun Server X4-8 Oracle Database Server**



## 2.18.14 Sun Server X2-8 Oracle Database Server Images

The following image shows the front view of Sun Server X2-8 Oracle Database Server.

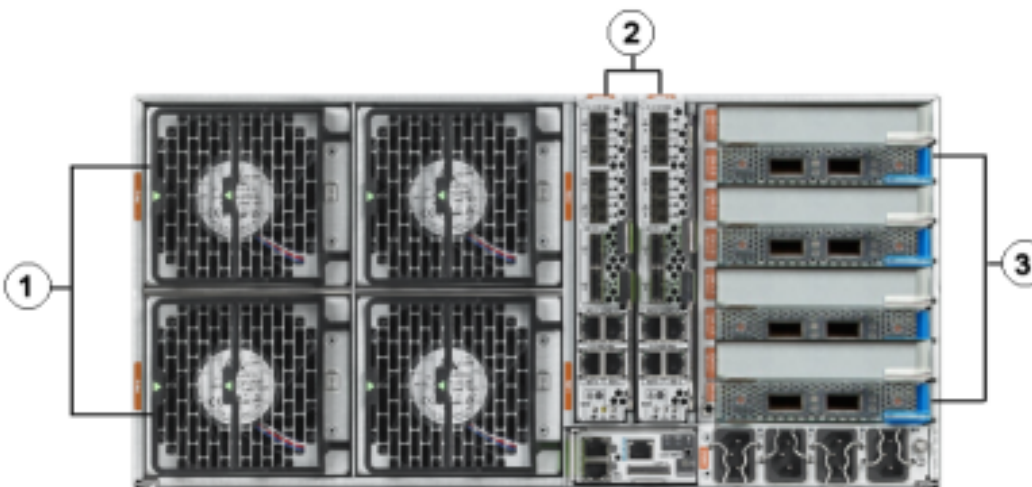
**Figure 2-29 Front View of Sun Server X2-8 Oracle Database Server**



1. Power supplies.
2. Hard disk drives. The top drives are, from left to right, XL4, XL5, XL6, and XL7. The lower drives are, from left to right, XL0, XL1, XL2, and XL3.
3. CPU modules. The modules are, from bottom to top, BL0, BL1, BL2, and BL3.

The following image shows the rear view of Sun Fire X4800 Oracle Database Server.

**Figure 2-30 Rear View of Sun Server X2-8 Oracle Database Server**



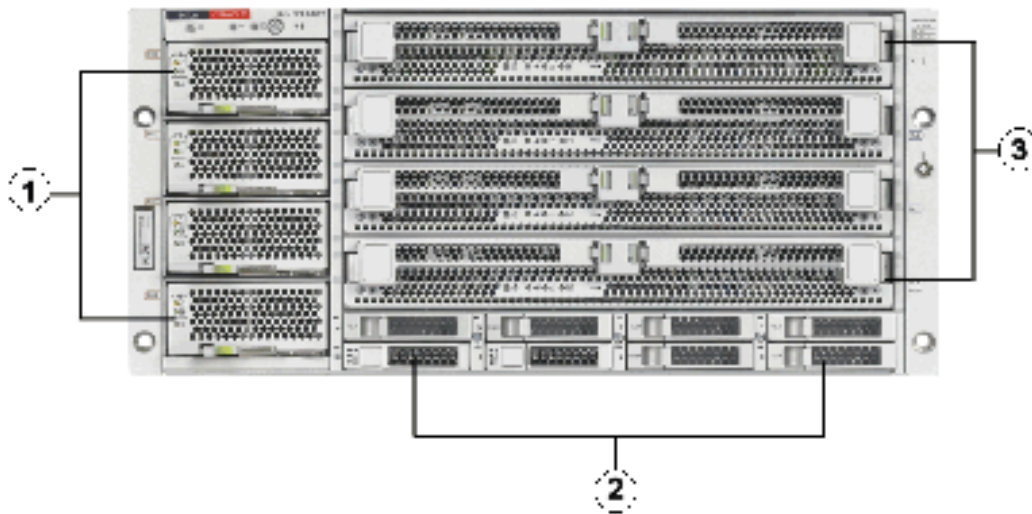


1. Fan modules.
2. Network Express Module.
3. InfiniBand EM (CX2) dual port PCI Express modules.

## 2.18.15 Sun Fire X4800 Oracle Database Server Images

The following image shows the front view of Sun Fire X4800 Oracle Database Server.

**Figure 2-31 Front View of Sun Fire X4800 Oracle Database Server**

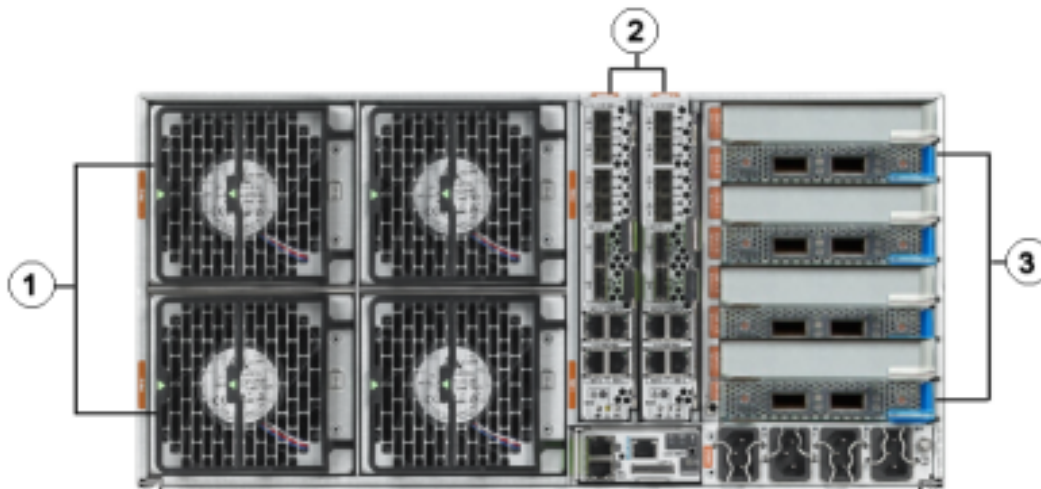


1. Power supplies.
2. Hard disk drives. The top drives are, from left to right, XL4, XL5, XL6, and XL7. The lower drives are, from left to right, XL0, XL1, XL2, and XL3.
3. CPU modules. The modules are, from bottom to top, BL0, BL1, BL2, and BL3.

The following image shows the rear view of Sun Fire X4800 Oracle Database Server.



Figure 2-32 Rear View of Sun Fire X4800 Oracle Database Server



1. Fan modules.
2. Network Express Module.
3. InfiniBand EM (CX2) dual port PCI Express modules.

# 3

## Maintaining Oracle Exadata Storage Servers

Oracle Exadata Storage Servers contain disks and memory devices that might require maintenance.

### Note:

- All procedures in this chapter are applicable to Exadata Database Machine and Oracle Exadata Storage Expansion Rack.
- For ease of reading, the name "Oracle Exadata Rack" is used when information refers to both Exadata Database Machine and Oracle Exadata Storage Expansion Rack.

- [Maintaining Oracle Exadata Storage Servers](#)  
This section describes how to perform maintenance on Oracle Exadata Storage Servers.
- [Using Exadata Extended \(XT\) Storage Servers](#)  
Oracle Exadata Extended (XT) Storage Server offers a lower cost storage option that can be used for infrequently accessed, older, or regulatory data.
- [Maintaining the Hard Disks of Oracle Exadata Storage Servers](#)
- [Maintaining Flash Disks on Oracle Exadata Storage Servers](#)
- [Maintaining PMEM Devices on Oracle Exadata Storage Servers](#)  
Starting with Oracle Exadata X8M, High Capacity (HC) and Extreme Flash (EF) storage servers contain persistent memory (PMEM) devices.
- [Maintaining the M.2 Disks of Oracle Exadata Storage Server](#)  
Exadata Database Machine X7 and later systems come with two internal M.2 devices that contain the system area.
- [Managing the RAM Cache on the Storage Servers](#)  
Cell RAM Cache is a cache in front of the Flash Cache and is an extension of the database cache. It is faster than the Flash Cache, but has smaller capacity.
- [Resizing Grid Disks](#)  
You can resize grid disks and Oracle ASM disk groups to shrink one with excess free space and increase the size of another that is near capacity.
- [Using the Oracle Exadata System Software Rescue Procedure](#)  
In the rare event that the system disks fail simultaneously, you must use the Oracle Exadata Storage Server rescue procedure to recover the system.
- [Changing Existing Elastic Configurations for Storage Cells](#)  
You can modify the capacity of your Exadata Database Machine using elastic configuration.
- [Managing Disk Controller Batteries](#)  
This section applies only to Exadata systems prior to X6 that use batteries. Newer systems have CVPM02 (Cache Vault), which is a super cap and not a battery.

- [Managing F20 PCIe Energy Storage Modules](#)  
Sun Flash Accelerator F20 PCIe cards are used in Exadata Database Machine X3 models.
- [LED Status Descriptions](#)  
The LEDs on the Oracle Exadata Rack components help you identify the component that needs servicing.
- [Exadata Storage Server Images](#)  
The Exadata Storage Server models have different external layouts and physical appearance.

## 3.1 Maintaining Oracle Exadata Storage Servers

This section describes how to perform maintenance on Oracle Exadata Storage Servers.

- [Shutting Down Exadata Storage Server](#)  
When performing maintenance on Exadata Storage Servers, it may be necessary to power down or restart the cell.
- [Checking Status of a Rebalance Operation](#)  
When dropping or adding a disk, you can check the status of the Oracle ASM rebalance operation.
- [Enabling Network Connectivity with the Diagnostic ISO](#)  
If a storage server does not restart, the diagnostic ISO may be needed to access the cell so it can be manually repaired.

### 3.1.1 Shutting Down Exadata Storage Server

When performing maintenance on Exadata Storage Servers, it may be necessary to power down or restart the cell.

If Exadata Storage Server is to be shut down when one or more databases are running, then you must verify that taking Exadata Storage Server offline will not impact Oracle ASM disk group and database availability. The ability to take Exadata Storage Server offline without affecting database availability depends on the level of Oracle ASM redundancy used on the affected disk groups. Availability also depends on the current status of disks in other Exadata Storage Servers that have mirror copies of data for the Exadata Storage Server that you are taking offline.

1. Optional: Configure the grid disks to remain offline after restarting the cell.

If you are planning to have multiple restarts, or you want to control when the Exadata Storage Server becomes active again, then you can perform this step. Making the grid disks inactive allows you to verify the planned maintenance activity was successful before making the grid disks available again.

- a. Set the grid disks to inactive.

```
CellCLI> ALTER GRIDDISK ALL INACTIVE
```

- b. Wait at least 30 seconds, or until Oracle ASM has completed taking the corresponding Oracle ASM disks offline.

This step is very important if you are using versions of Oracle Exadata System Software before release 18.1. If you put the commands into a script, then make sure to add a `sleep` command with a value over 30 seconds.

 **Note:**

If you set the grid disks to inactive, then you must complete step 6 later to activate the grid disks.

2. Stop the cell services.

```
CellCLI> ALTER CELL SHUTDOWN SERVICES ALL
```

The preceding command checks if any disks are offline, in predictive failure status, or need to be copied to its mirror. If Oracle ASM redundancy is intact, then the command takes the grid disks offline in Oracle ASM, and then stops the cell services. If the following error is displayed, then it may not be safe to stop the cell services because a disk group may be forced to dismount due to redundancy.

```
Stopping the RS, CELLSRV, and MS services...
The SHUTDOWN of ALL services was not successful.
CELL-01548: Unable to shut down CELLSRV because disk group DATA, RECO may
be
forced to dismount due to reduced redundancy.
Getting the state of CELLSRV services... running
Getting the state of MS services... running
Getting the state of RS services... running
```

If the `CELL-01548` error occurs, then restore Oracle ASM disk group redundancy and retry the command when disk status is back to normal for all the disks.

3. Shut down the Exadata Storage Server.

See [Powering Off Oracle Exadata Storage Servers](#)

4. After performing the maintenance, restart the Exadata Storage Server. The cell services are started automatically. As part of the Exadata Storage Server startup, all grid disks are automatically changed to `ONLINE` in Oracle ASM.

5. Verify that all grid disks have been successfully brought online.

```
CellCLI> LIST GRIDDISK ATTRIBUTES name, asmmodestatus
```

Wait until `asmmodestatus` shows `ONLINE` or `UNUSED` for all grid disks.

6. Optional: Change the grid disks status to `ONLINE`.

This step is only necessary when step 1 has been performed. If step 1 was not performed, then the grid disks were set to online automatically when the Exadata Storage Server was restarted.

```
CellCLI> ALTER GRIDDISK ALL ACTIVE
```

## 3.1.2 Checking Status of a Rebalance Operation

When dropping or adding a disk, you can check the status of the Oracle ASM rebalance operation.

- The rebalance operation may have completed successfully. Check the Oracle ASM alert logs to confirm.
- The rebalance operation may be currently running. Check the `GV$ASM_OPERATION` view to determine if the rebalance operation is still running.
- The rebalance operation may have failed. Check the `V$ASM_OPERATION.ERROR` view to determine if the rebalance operation failed.
- Rebalance operations from multiple disk groups can be done on different Oracle ASM instances in the same cluster if the physical disk being replaced contains Oracle ASM disks from multiple disk groups. One Oracle ASM instance can run one rebalance operation at a time. If all Oracle ASM instances are busy, then rebalance operations are queued.

### Note:

For Oracle Exadata Storage Servers running Oracle Exadata System Software release 12.1.2.0 with Oracle Database release 12.1.0.2 with BP4, Oracle ASM sends an e-mail about the status of a rebalance operation. In earlier releases, the administrator had to check the status of the operation.

## 3.1.3 Enabling Network Connectivity with the Diagnostic ISO

If a storage server does not restart, the diagnostic ISO may be needed to access the cell so it can be manually repaired.

The diagnostic ISO should be used after other boot methods, such as using the USB, do not work.

The following procedure enables networking with the diagnostic ISO so files can be transferred to repair the cell:

1. Restart the system using the `diagnostics.iso` file.  
See Booting a Server using the Diagnostic ISO File in *Oracle Exadata System Software User's Guide*.
2. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
```

```
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

3. Use the following command to avoid pings:

```
alias ping="ping -c"
```

4. Make a directory named `/etc/network`.
5. Make a directory named `/etc/network/if-pre-up.d`.
6. Add the following lines to the `/etc/network/interfaces` file:

```
iface eth0 inet static
address IP_address_of_cell
netmask netmask_of_cell
gateway gateway_IP_address_of_cell
```

7. Bring up the `eth0` interface using the following command:

```
ifup eth0
```

There may be some warning messages, but the interface is operational.

8. Use either FTP or the `wget` command to retrieve the files to repair the cell.

## 3.2 Using Exadata Extended (XT) Storage Servers

Oracle Exadata Extended (XT) Storage Server offers a lower cost storage option that can be used for infrequently accessed, older, or regulatory data.

- [About Oracle Exadata Extended \(XT\) Storage Servers](#)  
Oracle Exadata Extended (XT) Storage Servers help you extend the operational and management benefits of Exadata Database Machine to rarely accessed data that must be kept online.
- [What Data Can Be Stored on Oracle Exadata Extended \(XT\) Storage Servers?](#)  
Oracle Exadata Extended (XT) Storage Servers are intended provide lower-cost storage for infrequently accessed, older, or regulatory data.
- [Enabling Smart Scan on Exadata Extended \(XT\) Storage Servers](#)  
If you purchase Oracle Exadata System Software licenses for your Oracle Exadata Extended (XT) Storage Servers, then you can enable features such as Smart Scan and Storage Indexes to improve performance.

### 3.2.1 About Oracle Exadata Extended (XT) Storage Servers

Oracle Exadata Extended (XT) Storage Servers help you extend the operational and management benefits of Exadata Database Machine to rarely accessed data that must be kept online.

Each XT storage server includes twelve high-capacity disk drives. However, to achieve a lower cost, flash devices are not included. Also, Oracle Exadata System Software licensing is optional, with some software features disabled unless licensed. Hybrid Columnar Compression is included without requiring Oracle Exadata System Software licenses.

XT storage servers use the same RDMA Network Fabric as the other servers in your Oracle Exadata Rack. XT storage servers add storage capacity while remaining transparent to applications, transparent to SQL, and retaining the same operational model. You can use the same security model and encryption used for your other Exadata storage servers.

You can add XT storage servers to Oracle Exadata Rack X4 or newer, including Eighth Rack configurations. You must add at least 2 servers initially. After the initial 2 servers, you can add XT storage servers as needed. To implement high redundancy, you must have a minimum of 3 XT storage servers. XT storage servers follow the same placement patterns as High Capacity (HC) and Extreme Flash (EF) storage servers.

An Oracle ASM disk group should use storage provided by only one type of storage server (HC, EF, or XT). After adding the XT storage servers to your rack, create new disk groups to use the storage. The default disk group name for XT storage servers is XTND. However, you can use a different name as required.

XT storage servers provide fully integrated storage for Oracle Database. You can use the new disk group with database features such as Oracle Partitioning, Oracle Automatic Data Optimization, and Oracle Advanced Compression.

## 3.2.2 What Data Can Be Stored on Oracle Exadata Extended (XT) Storage Servers?

Oracle Exadata Extended (XT) Storage Servers are intended provide lower-cost storage for infrequently accessed, older, or regulatory data.

XT storage servers help you to keep all your required data online and available for queries. This includes data such as:

- Historical data
- Images, BLOBs, contracts, and other large table-based objects
- Compliance and regulatory data
- Local backups

The XT storage servers can also provide storage for development databases which have less stringent performance requirements compared to production databases.

## 3.2.3 Enabling Smart Scan on Exadata Extended (XT) Storage Servers

If you purchase Oracle Exadata System Software licenses for your Oracle Exadata Extended (XT) Storage Servers, then you can enable features such as Smart Scan and Storage Indexes to improve performance.

1. Procure or transfer Oracle Exadata System Software licenses.

All drives must be licensed to enable Smart Scan.

2. Modify the `enableSmartStorage` attribute for the XT storage servers.

You do not need to stop the storage servers first. Simply run the following command on each XT storage server that is licensed:

```
cellcli -e ALTER CELL enableSmartStorage=true
```

3. Verify the cell has been modified.

```
cellcli -e "LIST CELL ATTRIBUTES name, status, enableSmartStorage"
```

## 3.3 Maintaining the Hard Disks of Oracle Exadata Storage Servers

Every Oracle Exadata Storage Server in Oracle Exadata Rack has a system area, which is where the Oracle Exadata System Software system software resides. In Exadata Database Machine X7 and later systems, two internal M.2 devices contain the system area. In all other systems, the first two disks of Oracle Exadata Storage Server are system disks and the portions on these system disks are referred to as the system area.

In Exadata Database Machine X7 and later systems, all the hard disks in the cell are data disks. In systems prior to Exadata Database Machine X7, the non-system area of the system disks, referred to as data partitions, is used for normal data storage. All other disks in the cell are data disks.

Starting in Oracle Exadata System Software release 11.2.3.2.0, if there is a disk failure, then Oracle Exadata System Software sends an alert stating that the disk can be replaced, and, after all data has been rebalanced out from that disk, turns on the blue OK to Remove LED for the hard disk with predictive failure. In Oracle Exadata System Software releases earlier than 11.2.3.2.0, the amber Fault-Service Required LED was turned on for a hard disk with predictive failure, but not the blue LED. In these cases, it is necessary to manually check if all data has been rebalanced out from the disk before proceeding with disk replacement.

Starting with Oracle Exadata System Software release 18.1.0.0.0 and Exadata Database Machine X7 systems, there is an additional white Do Not Service LED that indicates when redundancy is reduced to inform system administrators or field engineers that the storage server should not be powered off for services. When redundancy is restored, Oracle Exadata System Software automatically turns off the Do Not Service LED to indicate that the cell can be powered off for services.

For a hard disk that has failed, both the blue OK to Remove LED and the amber Fault-Service Required LED are turned on for the drive indicating that disk replacement can proceed. The behavior is the same in all releases. The drive LED light is a solid light in Oracle Exadata System Software releases 11.2.3.2.0 and later; the drive LED blinks in earlier releases.



### Note:

Oracle Exadata Rack is online and available while replacing the Oracle Exadata Storage Server physical disks.

This section contains the following topics:



- [Monitoring the Status of Hard Disks](#)  
You can monitor the status of a hard disk by checking its attributes with the `CellCLI LIST PHYSICALDISK` command.
- [Monitoring Hard Disk Controller Write-through Caching Mode](#)  
The hard disk controller on each Oracle Exadata Storage Server periodically performs a discharge and charge of the controller battery. During the operation, the write cache policy changes from write-back caching to write-through caching.
- [Replacing a Hard Disk Due to Disk Failure](#)
- [Replacing a Hard Disk Due to Disk Problems](#)  
You may need to replace a hard disk because the disk is in `warning - predictive failure` status.
- [Replacing a Hard Disk Due to Bad Performance](#)  
A single bad hard disk can degrade the performance of other good disks. It is better to remove the bad disk from the system than let it remain.
- [Replacing a Hard Disk Proactively](#)  
Exadata Storage software has a complete set of automated operations for hard disk maintenance, when a hard disk has failed or has been flagged as a problematic disk. But there are situations where a hard disk has to be removed proactively from the configuration.
- [Moving All Drives to Another Exadata Storage Server](#)  
It may necessary to move all drives from one Exadata Storage Server to another Exadata Storage Server.
- [Repurposing a Hard Disk](#)  
You may want to delete all data on a disk, and then use the disk for another purpose.
- [Removing and Replacing the Same Hard Disk](#)  
What happens if you accidentally remove the wrong hard disk?
- [Re-Enabling a Hard Disk That Was Rejected](#)  
If a physical disk was rejected because it was inserted into the wrong slot, you can re-enable the disk.

#### Related Topics

- [LED Status Descriptions](#)  
The LEDs on the Oracle Exadata Rack components help you identify the component that needs servicing.
- [Oracle Maximum Availability Architecture \(MAA\) Web site](#)
- [Bare Metal Restore Procedure for Compute Nodes on an Exadata Environment \(My Oracle Support Doc ID 1084360.1\)](#)

### 3.3.1 Monitoring the Status of Hard Disks

You can monitor the status of a hard disk by checking its attributes with the `CellCLI LIST PHYSICALDISK` command.

For example, a hard disk status equal to `failed` (the status for failed hard disks was `critical` in earlier releases), or `warning - predictive failure` is probably having problems and needs to be replaced. The disk firmware maintains the error counters, and marks a drive with `Predictive Failure` when internal thresholds are exceeded. The drive, not the cell software, determines if it needs replacement.

- Use the CellCLI command `LIST PHYSICALDISK` to determine the status of a hard disk:

```
CellCLI> LIST PHYSICALDISK WHERE disktype=harddisk AND status!=normal
DETAIL
 name: 8:4
 deviceId: 12
 deviceName: /dev/sde
 diskType: HardDisk
 enclosureDeviceId: 8
 errOtherCount: 0
 luns: 0_4
 makeModel: "HGST H7280A520SUN8.0T"
 physicalFirmware: PD51
 physicalInsertTime: 2016-11-30T21:24:45-08:00
 physicalInterface: sas
 physicalSerial: PA9TVR
 physicalSize: 7.153663907200098T
 slotNumber: 4
 status: failed
```

When disk I/O errors occur, Oracle ASM performs bad extent repair for read errors due to media errors. The disks will stay online, and no alerts are sent. When Oracle ASM gets a read error on a physically-addressed metadata block, it does not have mirroring for the blocks, and takes the disk offline. Oracle ASM then drops the disk using the `FORCE` option.

The Oracle Exadata Storage Server hard disk statuses are as follows:

- Oracle Exadata System Software release 11.2.3.3 and later:
  - normal
  - normal - dropped for replacement
  - normal - confinedOnline
  - normal - confinedOnline - dropped for replacement
  - not present
  - failed
  - failed - dropped for replacement
  - failed - rejected due to incorrect disk model
  - failed - rejected due to incorrect disk model - dropped for replacement
  - failed - rejected due to wrong slot
  - failed - rejected due to wrong slot - dropped for replacement
  - warning - confinedOnline
  - warning - confinedOnline - dropped for replacement
  - warning - peer failure
  - warning - poor performance
  - warning - poor performance - dropped for replacement
  - warning - poor performance, write-through caching
  - warning - predictive failure, poor performance

- warning - predictive failure, poor performance - dropped for replacement
- warning - predictive failure, write-through caching
- warning - predictive failure
- warning - predictive failure - dropped for replacement
- warning - predictive failure, poor performance, write-through caching
- warning - write-through caching
- Oracle Exadata System Software release 11.2.3.2:
  - normal
  - normal - confinedOnline
  - not present
  - failed
  - failed - rejected due to incorrect disk model
  - failed - rejected due to wrong slot
  - warning - confinedOnline
  - warning - peer failure
  - warning - poor performance
  - warning - poor performance, write-through caching
  - warning - predictive failure, poor performance
  - warning - predictive failure, write-through caching
  - warning - predictive failure
  - warning - predictive failure, poor performance, write-through caching
  - warning - write-through caching
- Oracle Exadata System Software release 11.2.3.1.1 and earlier:
  - normal
  - critical
  - poor performance
  - predictive failure
  - not present

#### Related Topics

- LIST PHYSICALDISK
- About Oracle ASM Disk Repair Timer

### 3.3.2 Monitoring Hard Disk Controller Write-through Caching Mode

The hard disk controller on each Oracle Exadata Storage Server periodically performs a discharge and charge of the controller battery. During the operation, the write cache policy changes from write-back caching to write-through caching.

Write-through cache mode is slower than write-back cache mode. However, write-back cache mode has a risk of data loss if the Oracle Exadata Storage Server loses power

or fails. For Oracle Exadata System Software releases earlier than release 11.2.1.3, the operation occurs every month. For Oracle Exadata System Software release 11.2.1.3.0 and later, the operation occurs every three months, for example, at 01:00 on the 17th day of January, April, July and October.

- To change the start time for the learn cycle, use a command similar to the following:

```
CellCLI> ALTER CELL bbuLearnCycleTime="2013-01-22T02:00:00-08:00"
```

The time reverts to the default learn cycle time after the cycle completes.

- To see the time for the next learn cycle, use the following command:

```
CellCLI> LIST CELL ATTRIBUTES bbuLearnCycleTime
```

Oracle Exadata Storage Server generates an informational alert about the status of the caching mode for logical drives on the cell, similar to the following:

```
HDD disk controller battery on disk controller at adapter 0 is going into
a learn
cycle. This is a normal maintenance activity that occurs quarterly and
runs for
approximately 1 to 12 hours. The disk controller cache might go into
WriteThrough
caching mode during the learn cycle. Disk write throughput might be
temporarily
lower during this time. The message is informational only, no action is
required.
```

- To view the status of the battery, use a command similar to the following example:

 **Note:**

If you are running Oracle Exadata System Software 19.1.0 or later, substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following command:

```
/opt/MegaRAID/MegaCli/MegaCli64 -AdpBbuCmd -GetBbuStatus -a0
BBU status for Adapter: 0
```

```
BatteryType: iBBU08
Voltage: 3721 mV
Current: 541 mA
Temperature: 43 C
```

```
BBU Firmware Status:
Charging Status : Charging
Voltage : OK
Temperature : OK
Learn Cycle Requested : No
Learn Cycle Active : No
Learn Cycle Status : OK
```

```
Learn Cycle Timeout : No
I2c Errors Detected : No
Battery Pack Missing : No
Battery Replacement required : No
Remaining Capacity Low : Yes
Periodic Learn Required : No
Transparent Learn : No
```

Battery state:

```
GasGUAGEStatus:
Fully Discharged : No
Fully Charged : No
Discharging : No
Initialized : No
Remaining Time Alarm : Yes
Remaining Capacity Alarm: No
Discharge Terminated : No
Over Temperature : No
Charging Terminated : No
Over Charged : No
```

**Relative State of Charge: 7 %**

```
Charger System State: 1
Charger System Ctrl: 0
Charging current: 541 mA
```

**Absolute state of charge: 0 %**

```
Max Error: 0 %
```

```
Exit Code: 0x00
```

### 3.3.3 Replacing a Hard Disk Due to Disk Failure

A hard disk outage can cause a reduction in performance and data redundancy. Therefore, the disk should be replaced with a new disk as soon as possible. When the disk fails, the Oracle ASM disks associated with the grid disks on the hard disk are automatically dropped with the `FORCE` option, and an Oracle ASM rebalance follows to restore the data redundancy.

An Exadata alert is generated when a disk fails. The alert includes specific instructions for replacing the disk. If you have configured the system for alert notifications, then the alert is sent by e-mail to the designated address.

After the hard disk is replaced, the grid disks and cell disks that existed on the previous disk in that slot are re-created on the new hard disk. If those grid disks were part of an Oracle ASM group, then they are added back to the disk group, and the data is rebalanced on them, based on the disk group redundancy and `ASM_POWER_LIMIT` parameter.

 **Note:**

For storage servers running Oracle Exadata System Software release 12.1.2.0 with Oracle Database release 12.1.0.2 with BP4, Oracle ASM sends an e-mail about the status of a rebalance operation. In earlier releases, the administrator had to check the status of the operation.

For earlier releases, check the rebalance operation status as described in [Checking Status of a Rebalance Operation](#).

The following procedure describes how to replace a hard disk due to disk failure:

1. Determine the failed disk using the following command:

```
CellCLI> LIST PHYSICALDISK WHERE diskType=HardDisk AND status=failed
DETAIL
```

The following is an example of the output from the command. The slot number shows the location of the disk, and the status shows that the disk has failed.

```
CellCLI> LIST PHYSICALDISK WHERE diskType=HardDisk AND status=failed
DETAIL
```

```

name: 28:5
deviceId: 21
diskType: HardDisk
enclosureDeviceId: 28
errMediaCount: 0
errOtherCount: 0
foreignState: false
luns: 0_5
makeModel: "SEAGATE ST360057SSUN600G"
physicalFirmware: 0705
physicalInterface: sas
physicalSerial: A01BC2
physicalSize: 558.9109999993816G
slotNumber: 5
status: failed
```

2. Ensure the blue OK to Remove LED on the disk is lit before removing the disk.
3. Replace the hard disk on Oracle Exadata Storage Server and wait for three minutes. The hard disk is hot-pluggable, and can be replaced when the power is on.
4. Confirm the disk is online.

When you replace a hard disk, the disk must be acknowledged by the RAID controller before you can use it. This does not take long.

Use the `LIST PHYSICALDISK` command similar to the following to ensure the status is NORMAL.

```
CellCLI> LIST PHYSICALDISK WHERE name=28:5 ATTRIBUTES status
```

5. Verify the firmware is correct using the `ALTER CELL VALIDATE CONFIGURATION` command.

In rare cases, the automatic firmware update may not work, and the LUN is not rebuilt. This can be confirmed by checking the `ms-odl.trc` file.

 **See Also:**

- *Oracle Database Reference* for information about the `V$ASM_OPERATION` view
- *Oracle Automatic Storage Management Administrator's Guide* for information about the rebalance operation

### 3.3.4 Replacing a Hard Disk Due to Disk Problems

You may need to replace a hard disk because the disk is in `warning - predictive failure status`.

The `predictive failure` status indicates that the hard disk will soon fail, and should be replaced at the earliest opportunity. The Oracle ASM disks associated with the grid disks on the hard drive are automatically dropped, and an Oracle ASM rebalance relocates the data from the predictively failed disk to other disks.

If the drop did not complete before the hard drive dies, then refer to [Replacing a Hard Disk Due to Disk Failure](#).

An alert is sent when the disk is removed. After replacing the hard disk, the grid disks and cell disks that existed on the previous disk in the slot are re-created on the new hard disk. If those grid disks were part of an Oracle ASM disk group, then they are added back to the disk group, and the data is rebalanced based on disk group redundancy and the `ASM_POWER_LIMIT` parameter.

 **Note:**

On Oracle Exadata Storage Servers running Oracle Exadata System Software release 12.1.2.0 with Oracle Database release 12.1.0.2 with BP4, Oracle ASM sends an e-mail about the status of a rebalance operation. In earlier releases, the administrator had to check the status of the operation.

For earlier releases, check the rebalance operation status as described in [Checking Status of a Rebalance Operation](#).

1. Determine which disk is the failing disk.

```
CellCLI> LIST PHYSICALDISK WHERE diskType=HardDisk AND status= \
"warning - predictive failure" DETAIL
```

The following is an example of the output. The slot number shows the location of the disk, and the status shows the disk is expected to fail.

```
CellCLI> LIST PHYSICALDISK WHERE diskType=HardDisk AND status= \
"warning - predictive failure" DETAIL
name: 28:3
deviceId: 19
diskType: HardDisk
enclosureDeviceId: 28
errMediaCount: 0
errOtherCount: 0
foreignState: false
luns: 0_3
makeModel: "SEAGATE ST360057SSUN600G"
physicalFirmware: 0705
physicalInterface: sas
physicalSerial: E07L8E
physicalSize: 558.9109999993816G
slotNumber: 3
status: warning - predictive failure
```

2. Ensure the blue OK to Remove LED on the disk is lit before removing the disk.
3. Wait until the Oracle ASM disks associated with the grid disks on the hard disk have been successfully dropped. To determine if the grid disks have been dropped, query the `V$ASM_DISK_STAT` view on the Oracle ASM instance.

### Caution:

On all systems prior to Oracle Exadata Database Machine X7, the disks in the first two slots are system disks which store the operating system and Oracle Exadata System Software. One system disk must be in working condition to keep up the server.

Wait until `ALTER CELL VALIDATE CONFIGURATION` shows no `mdadm` errors, which indicates the system disk resynchronization has completed, before replacing the other system disk.

4. Replace the hard disk on Oracle Exadata Storage Server and wait for three minutes. The hard disk is hot-pluggable, and can be replaced when the power is on.
5. Confirm the disk is online.

When you replace a hard disk, the disk must be acknowledged by the RAID controller before you can use it. This does not take long. Use the `LIST PHYSICALDISK` command to ensure the status is `NORMAL`.

```
CellCLI> LIST PHYSICALDISK WHERE name=28:3 ATTRIBUTES status
```

6. Verify the firmware is correct using the `ALTER CELL VALIDATE CONFIGURATION` command.



 **See Also:**

- *Oracle Database Reference* for information about the `V$ASM_OPERATION` view
- *Oracle Automatic Storage Management Administrator's Guide* for information about tuning rebalance operations.
- *Oracle Database Reference* for information about querying the `V$ASM_DISK_STAT` view
- `ALTER CELL VALIDATE CONFIGURATION` command in *Oracle Exadata System Software User's Guide*

### 3.3.5 Replacing a Hard Disk Due to Bad Performance

A single bad hard disk can degrade the performance of other good disks. It is better to remove the bad disk from the system than let it remain.

Starting with Oracle Exadata System Software release 11.2.3.2, an underperforming disk is automatically identified and removed from active configuration. Oracle Exadata Database Machine then runs a set of performance tests. When poor disk performance is detected by CELLSRV, the cell disk status changes to `normal - confinedOnline`, and the hard disk status changes to `warning - confinedOnline`.

The following conditions trigger disk confinement:

- Disk stopped responding. The cause code in the storage alert log is `CD_PERF_HANG`.
- Slow cell disk such as the following:
  - High service time threshold (cause code `CD_PERF_SLOW_ABS`)
  - High relative service time threshold (cause code `CD_PERF_SLOW_RLTV`)
- High read or write latency such as the following:
  - High latency on writes (cause code `CD_PERF_SLOW_LAT_WT`)
  - High latency on reads (cause code `CD_PERF_SLOW_LAT_RD`)
  - High latency on reads and writes (cause code `CD_PERF_SLOW_LAT_RW`)
  - Very high absolute latency on individual I/Os happening frequently (cause code `CD_PERF_SLOW_LAT_ERR`)
- Errors such as I/O errors (cause code `CD_PERF_IOERR`).

If the disk problem is temporary and passes the tests, then it is brought back into the configuration. If the disk does not pass the tests, then it is marked as `poor performance`, and Oracle Auto Service Request (ASR) submits a service request to replace the disk. If possible, Oracle ASM takes the grid disks offline for testing. If Oracle ASM cannot take the disks offline, then the cell disk status stays at `normal - confinedOnline` until the disks can be taken offline safely.

The disk status change is associated with the following entry in the cell alert history:

```
MESSAGE ID date_time info "Hard disk entered confinement status. The LUN
n_m changed status to warning - confinedOnline. CellDisk changed status to
normal
- confinedOnline. Status: WARNING - CONFINEDONLINE Manufacturer: name
Model
Number: model Size: size Serial Number: serial_number Firmware:
fw_release
Slot Number: m Cell Disk: cell_disk_name Grid Disk: grid disk 1, grid
disk 2
... Reason for confinement: threshold for service time exceeded"
```

The following would be logged in the storage cell alert log:

```
CDHS: Mark cd health state change cell_disk_name with newState HEALTH_BAD_
ONLINE pending HEALTH_BAD_ONLINE ongoing INVALID cur HEALTH_GOOD
Celldisk entering CONFINED ACTIVE state with cause CD_PERF_SLOW_ABS
activeForced: 0
inactiveForced: 0 trigger HistoryFail: 0, forceTestOutcome: 0 testFail: 0
global conf related state: numHDsConf: 1 numFDsConf: 0 numHDsHung: 0
numFDsHung: 0
...
```

#### Note:

In releases earlier than Oracle Exadata System Software release 11.2.3.2, use the `CALIBRATE` command to identify a bad hard disk, and look for very low throughput and IOPS for each hard disk.

The following procedure describes how to remove a hard disk once the bad disk has been identified:

1. Illuminate the hard drive service LED to identify the drive to be replaced using a command similar to the following, where `disk_name` is the name of the hard disk to be replaced, such as `20:2`:

```
cellcli -e 'alter physicaldisk disk_name serviceled on'
```

2. Find all the grid disks on the bad disk.

For example:

```
[root@exa05celadm03 ~]# cellcli -e "list physicaldisk 20:11 attributes
name, id"
 20:11 RD58EA
[root@exa05celadm03 ~]# cellcli -e "list celldisk where
physicalDisk='RD58EA'"
 CD_11_exa05celadm03 normal
[root@exa05celadm03 ~]# cellcli -e "list griddisk where
cellDisk='CD_11_exa05celadm03'"
```

```
DATA_CD_11_exa05celadm03 active
DBFS_CD_11_exa05celadm03 active
RECO_CD_11_exa05celadm03 active
TPCH_CD_11_exa05celadm03 active
```

3. Direct Oracle ASM to stop using the bad disk immediately.

```
SQL> ALTER DISKGROUP diskgroup_name DROP DISK asm_disk_name;
```

4. Ensure the blue OK to Remove LED on the disk is lit before removing the disk.
5. Ensure that the Oracle ASM disks associated with the grid disks on the bad disk have been successfully dropped by querying the `V$ASM_DISK_STAT` view.
6. Remove the badly-performing disk. An alert is sent when the disk is removed.
7. When a new disk is available, install the new disk in the system. The cell disks and grid disks are automatically created on the new hard disk.

#### Note:

When a hard disk is replaced, the disk must be acknowledged by the RAID controller before it can be used. The acknowledgement does not take long, but use the `LIST PHYSICALDISK` command to ensure the status is `NORMAL`.

#### See Also:

- *Oracle Exadata System Software User's Guide*
- *Oracle Automatic Storage Management Administrator's Guide* for information about dropping a disk from a disk group

## 3.3.6 Replacing a Hard Disk Proactively

Exadata Storage software has a complete set of automated operations for hard disk maintenance, when a hard disk has failed or has been flagged as a problematic disk. But there are situations where a hard disk has to be removed proactively from the configuration.

In the `CellCLI ALTER PHYSICALDISK` command, the `DROP FOR REPLACEMENT` option checks if a normal functioning hard disk can be removed safely without the risk of data loss. However, after the execution of the command, the grid disks on the hard disk are inactivated on the storage cell and set to offline in the Oracle ASM disk groups.

To reduce the risk of having a disk group without full redundancy and proactively replace a hard disk, follow this procedure:

1. Identify the LUN, cell disk, and grid disk associated with the hard disk.

Use a command similar to the following where, X:Y identifies the hard disk name of the drive you are replacing.

```
cellcli -e "list diskmap" | grep 'X:Y'
```

The output should be similar to the following:

```
20:5 KEBTDJ 5 normal
559G
CD_05_exaceladm01 /dev/sdf
"DATAAC1_CD_05_exaceladm01, DBFS_DG_CD_05_exaceladm01,
RECO1_CD_05_exaceladm01"
```

To get the LUN, issue a command similar to the following:

```
CellCLI> list lun where deviceName='/dev/sdf/'
0_5 0_5 normal
```

## 2. Drop the disk.

- If you are using at least Oracle Exadata System Software release 21.2.0, use the following command to drop the physical disk while maintaining redundancy:

```
CellCLI> alter physicaldisk X:Y drop for replacement maintain
redundancy
```

Wait for the operation to complete before continuing.

- If you are using an Oracle Exadata System Software release before 21.2.0, do the following:
  - a. Drop the affected grid disks from the Oracle ASM disk groups in normal mode.

```
SQL> ALTER DISKGROUP diskgroup_name DROP DISK asm_disk_name;
```

- b. Wait for the ASM rebalance operation to complete before continuing.
- c. Drop the physical disk.

Use a command similar to the following where, X:Y identifies the hard disk name of the drive you are replacing.

```
CellCLI> alter physicaldisk X:Y drop for replacement
```

3. Ensure that the blue OK to Remove LED on the disk is lit before removing the disk.
4. Replace the new hard disk.
5. Verify the LUN, cell disk and grid disk associated with the hard disk were created.

```
CellCLI> list lun lun_name
CellCLI> list celldisk where lun=lun_name
CellCLI> list griddisk where celldisk=celldisk_name
```

6. Verify the grid disk was added to the Oracle ASM disk groups.

The following query should return no rows.

```
SQL> SELECT path,header_status FROM v$asm_disk WHERE group_number=0;
```

The following query shows whether all the failure groups have the same number of disks:

```
SQL> SELECT group_number, failgroup, mode_status, count(*) FROM
v$asm_disk
GROUP BY group_number, failgroup, mode_status;
```

### 3.3.7 Moving All Drives to Another Exadata Storage Server

It may necessary to move all drives from one Exadata Storage Server to another Exadata Storage Server.

This need may occur when there is a chassis-level component failure, such as a motherboard or ILOM failure, or when troubleshooting a hardware problem.

1. Back up the files in the following directories:

- /etc/hosts
- /etc/modprobe.conf
- /etc/sysconfig/network
- /etc/sysconfig/network-scripts

2. Safely inactivate all grid disks and shut down Exadata Storage Server.

Refer to "[Shutting Down Exadata Storage Server](#)". Make sure the Oracle ASM `disk_repair_time` attribute is set to a sufficiently large enough value so Oracle ASM does not drop the disks before the grid disks can be activated in another Exadata Storage Server.

3. Move the hard disks, flash disks, disk controller and USB flash drive from the original Exadata Storage Server to the new Exadata Storage Server.

 **Caution:**

- Ensure the first two disks, which are the system disks, are in the same first two slots. Failure to do so causes the Exadata Storage Server to function improperly.
- Ensure the flash cards are installed in the same PCIe slots as the original Exadata Storage Server.

4. Power on the new Exadata Storage Server using either the service processor interface or by pressing the power button.

5. Log in to the console using the service processor or the KVM switch.

6. Check the files in the following directories. If they are corrupted, then restore them from the backups.

- /etc/hosts

- /etc/modprobe.conf
- /etc/sysconfig/network
- /etc/sysconfig/network-scripts

7. Use the `ifconfig` command to retrieve the new MAC address for `eth0`, `eth1`, `eth2`, and `eth3`. For example:

```
ifconfig eth0
eth0 Link encap:Ethernet HWaddr 00:14:4F:CA:D9:AE
 inet addr:10.204.74.184 Bcast:10.204.75.255 Mask:255.255.252.0
 inet6 addr: fe80::214:4fff:feca:d9ae/64 Scope:Link
 UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
 RX packets:141455 errors:0 dropped:0 overruns:0 frame:0
 TX packets:6340 errors:0 dropped:0 overruns:0 carrier:0
 collisions:0 txqueuelen:1000
 RX bytes:9578692 (9.1 MiB) TX bytes:1042156 (1017.7 KiB)
 Memory:f8c60000-f8c80000
```

8. Edit the `ifcfg-eth0` file, `ifcfg-eth1` file, `ifcfg-eth2` file, and `ifcfg-eth3` file in the `/etc/sysconfig/network-scripts` directory to change the `HWADDR` value based on the output from step 7. The following is an example of the `ifcfg-eth0` file:

```
DO NOT REMOVE THESE LINES
%GENERATED BY CELL%
DEVICE=eth0
BOOTPROTO=static
ONBOOT=yes
IPADDR=10.204.74.184
NETMASK=255.255.252.0
NETWORK=10.204.72.0
BROADCAST=10.204.75.255
GATEWAY=10.204.72.1
HOTPLUG=no
IPV6INIT=no
HWADDR=00:14:4F:CA:D9:AE
```

9. Restart Exadata Storage Server.
10. Activate the grid disks using the following command:

```
CellCLI> ALTER GRIDDISK ALL ACTIVE
```

If the Oracle ASM disk on the disks on the cell have not been dropped, then they change to `ONLINE` automatically, and start getting used.

11. Validate the configuration using the following command:

```
CellCLI> ALTER CELL VALIDATE CONFIGURATION
```

12. Activate the ILOM for ASR.

 **See Also:**

- *Oracle Auto Service Request Quick Installation Guide for Oracle Exadata Database Machine*
- *Oracle Exadata System Software User's Guide*  
for information about the CellCLI utility
- *Oracle Automatic Storage Management Administrator's Guide* for information about disk repair time

### 3.3.8 Repurposing a Hard Disk

You may want to delete all data on a disk, and then use the disk for another purpose.

Before repurposing a hard disk, ensure that you have copies of the data that is on the disk.

If you use this procedure for the system disks (disk 0 and disk1), then only the data partitions are erased, not the system partitions.

1. Use the CellCLI `LIST` command to display the Exadata Storage Server objects. You must identify the grid disks and cell disks on the hard drive. For example:

```
CellCLI> LIST PHYSICALDISK
 20:0 D174LX normal
 20:1 D149R0 normal
 ...
```

2. Determine the cell disks and grid disks on the LUN, using a command similar to the following:

```
CellCLI> LIST LUN WHERE physicalDrives='20:0' DETAIL
name: 0_0
deviceName: /dev/sda
diskType: HardDisk
id: 0_0
isSystemLun: TRUE
lunSize: 557.861328125G
lunUID: 0_0
physicalDrives: 20:0
raidLevel: 0
lunWriteCacheMode: "WriteBack, ReadAheadNone, Direct, No Write
Cache if Bad BBU"
status: normal
```

To get the celldisks and grid disks, use a command similar to the following:

```
#cellcli -e "list diskmap" | grep 20:0

 20:0 K68DWJ 0 normal
559G
```

```
CD_00_burd01celadm01 /dev/sda3
"DATA1_CD_00_burd01celadm01, RECO1_CD_00_burd01celadm01"
```

3. From Oracle ASM, drop the Oracle ASM disks on the hard disk using the following command:

```
SQL> ALTER DISKGROUP diskgroup_name DROP DISK asm_disk_name;
```

4. From Exadata Storage Server, drop the cell disks and grid disks on the hard disk using the following command:

```
CellCLI> DROP CELLDISK celldisk_on_this_lun FORCE
```

 **Note:**

To overwrite all data on the cell disk, use the `ERASE` option with the `DROP CELLDISK` command. The following is an example of the command:

```
CellCLI> DROP CELLDISK CD_03_cell01 ERASE=1pass NOWAIT
```

```
CellDisk CD_03_cell01 erase is in progress
```

5. Drop the drive for hot removal. For example:

```
CellCli> ALTER PHYSICALDISK 20:0 DROP FOR REPLACEMENT
```

6. Ensure the blue OK to Remove LED on the disk is lit before removing the disk.

 **Caution:**

Ensure the disk blue LED is turned on before removing the drive. Do not remove the drive if the disk blue LED is unlit, or it may cause your system to crash.

7. Remove the disk to be repurposed, and insert a new disk.
8. Wait for the new hard disk to be added as a LUN.

```
CellCLI> LIST LUN
```

The cell disks and grid disks are automatically be created on the new hard disk, and the grid disks are added to the Oracle ASM group.

**Related Topics**

- *Oracle Exadata System Software User's Guide*



### 3.3.9 Removing and Replacing the Same Hard Disk

What happens if you accidentally remove the wrong hard disk?

If you inadvertently remove the wrong hard disk, then put the disk back. It will automatically be added back in the Oracle ASM disk group, and its data is resynchronized.

**Note:**

When replacing disk due to disk failure or disk problems, the LED is lit on the disk for identification.

### 3.3.10 Re-Enabling a Hard Disk That Was Rejected

If a physical disk was rejected because it was inserted into the wrong slot, you can re-enable the disk.

Run the following command:

**Caution:**

The following command removes all data on the physical disk.

```
CellCLI> ALTER PHYSICALDISK hard_disk_name reenable force
```

The following is an example of the output from the command:

```
Physical disk 20:0 was reenabled.
```

## 3.4 Maintaining Flash Disks on Oracle Exadata Storage Servers

Data is mirrored across Exadata Cells, and write operations are sent to at least two storage cells. If a flash card in one Oracle Exadata Storage Server has problems, then the read and write operations are serviced by the mirrored data in another Oracle Exadata Storage Server. No interruption of service occurs for the application.

If a flash card fails while in write-back mode, then Oracle Exadata System Software determines the data in the flash cache by reading the data from the surviving mirror. The data is then written to the cell that had the failed flash card. The location of the data lost in the failed flash cache is saved by Oracle Exadata System Software at the time of the flash failure. Resilvering then starts by replacing the lost data with the mirrored copy. During resilvering, the grid disk status is `ACTIVE -- RESILVERING WORKING`. If the PMEM cache is in write-through mode, then the data in the failed

PMEM device is already present on the data grid disk, so there is no need for resilvering.

- [Replacing a Flash Disk Due to Flash Disk Failure](#)  
Each Oracle Exadata Storage Server is equipped with flash devices.
- [About Flash Disk Degraded Performance Statuses](#)  
If a flash disk has degraded performance, you might need to replace the disk.
- [Replacing a Flash Disk Due to Flash Disk Problems](#)
- [Performing a Hot Pluggable Replacement of a Flash Disk](#)  
Starting with Oracle Exadata Database Machine X7, flash disks are hot-pluggable on both Extreme Flash (EF) and High Capacity (HC) storage servers.
- [Enabling Write Back Flash Cache](#)  
Write operations serviced by flash instead of by disk are referred to as **write-back flash cache**.
- [Disabling Write Back Flash Cache](#)  
You can disable the Write-Back Flash Cache by enabling Write-Through Flash Cache.
- [Enabling Flash Cache Compression](#)
- [Monitoring Exadata Smart Flash Cache Usage Statistics](#)
- [Disabling Flash Cache Compression](#)

### 3.4.1 Replacing a Flash Disk Due to Flash Disk Failure

Each Oracle Exadata Storage Server is equipped with flash devices.

Starting with Oracle Exadata Database Machine X7, the flash devices are hot-pluggable on the Oracle Exadata Storage Servers. When performing a hot-pluggable replacement of a flash device on Oracle Exadata Storage Servers for X7 or later, the disk status should be `Dropped` for replacement, and the power LED on the flash card should be off, which indicates the flash disk is ready for online replacement.

#### **Caution:**

Removing a card with power LED on could result in a system crash. If a failed disk has a status of `Failed - dropped for replacement` but the power LED is still on, contact Oracle Support Services.

For Oracle Exadata Database Machine X6 and earlier, the flash devices are hot-pluggable on Extreme Flash (EF) storage servers, but not on High Capacity (HC) storage servers. On HC storage servers, you need to power down the storage servers before replacing them.

To identify a failed flash disk, use the following command:

```
CellCLI> LIST PHYSICALDISK WHERE disktype=flashdisk AND status=failed DETAIL
```

The following is an example of the output from an Extreme Flash storage server:

```
name: NVME_10
deviceName: /dev/nvme7n1
diskType: FlashDisk
```

```

luns: 0_10
makeModel: "Oracle NVMe SSD"
physicalFirmware: 8DV1RA13
physicalInsertTime: 2016-09-28T11:29:13-07:00
physicalSerial: CVMD426500E21P6LGN
physicalSize: 1.4554837569594383T
slotNumber: 10
status: failed

```

The following is an example of the output from an Oracle Flash Accelerator F160 PCIe Card:

```
CellCLI> LIST PHYSICALDISK WHERE DISKTYPE=flashdisk AND STATUS=failed
DETAIL
```

```

name: FLASH_5_1
deviceName: /dev/nvme1n1
diskType: FlashDisk
luns: 5_1
makeModel: "Oracle Flash Accelerator F160 PCIe
Card"
physicalFirmware: 8DV1RA13
physicalInsertTime: 2016-11-30T21:24:45-08:00
physicalSerial: 1030M03UYM
physicalSize: 1.4554837569594383T
slotNumber: "PCI Slot: 5; FDOM: 1"
status: failed

```

The following is an example of the output from a Sun Flash Accelerator F40 PCIe card:

```

name: FLASH_5_3
diskType: FlashDisk
luns: 5_3
makeModel: "Sun Flash Accelerator F40 PCIe Card"
physicalFirmware: TI35
physicalInsertTime: 2012-07-13T15:40:59-07:00
physicalSerial: 5L002X4P
physicalSize: 93.13225793838501G
slotNumber: "PCI Slot: 5; FDOM: 3"
status: failed

```

For the PCIe cards, the `name` and `slotNumber` attributes show the PCI slot and the FDOM number. For Extreme Flash storage servers, the `slotNumber` attribute shows the NVMe slot on the front panel.

On Oracle Exadata Database Machine X7 and later systems, all flash disks are in the form of an Add-in-Card (AIC), which is inserted into a PCIe slot on the motherboard. The `slotNumber` attribute shows the PCI number and FDOM number, regardless of whether it is an EF or HC storage server.

If a flash disk is detected to have failed, then an alert is generated indicating that the flash disk, as well as the LUN on it, has failed. The alert message includes either the PCI slot number and FDOM number or the NVMe slot number. These numbers

uniquely identify the field replaceable unit (FRU). If you have configured the system for alert notification, then an alert is sent by e-mail message to the designated address.

A flash disk outage can cause reduction in performance and data redundancy. The failed disk should be replaced with a new flash disk at the earliest opportunity. If the flash disk is used for flash cache, then the effective cache size for the storage server is reduced. If the flash disk is used for flash log, then flash log is disabled on the disk thus reducing the effective flash log size. If the flash disk is used for grid disks, then the Oracle Automatic Storage Management (Oracle ASM) disks associated with these grid disks are automatically dropped with the `FORCE` option from the Oracle ASM disk group, and a rebalance operation starts to restore the data redundancy.

The following procedure describes how to replace an FDOM due to disk failure on High Capacity storage servers that do not support online flash replacement. Replacing an NVMe drive on Extreme Flash storage servers is the same as replacing a physical disk: you can just remove the NVMe drive from the front panel and insert a new one. You do not need to shut down the storage server.

1. Shut down the storage server. See "[Shutting Down Exadata Storage Server](#)"
2. Replace the failed flash disk based on the PCI number and FDOM number. A white Locator LED is lit to help locate the affected storage server.
3. Power up the storage server. The cell services are started automatically. As part of the storage server startup, all grid disks are automatically `ONLINE` in Oracle ASM.
4. Verify that all grid disks have been successfully put online using the following command:

```
CellCLI> LIST GRIDDISK ATTRIBUTES name, asmmodestatus
 data_CD_00_testceladm10 ONLINE
 data_CD_01_testceladm10 ONLINE
 data_CD_02_testceladm10 ONLINE
 ...
```

Wait until `asmmodestatus` shows `ONLINE` or `UNUSED` for all grid disks.

The new flash disk is automatically used by the system. If the flash disk is used for flash cache, then the effective cache size increases. If the flash disk is used for grid disks, then the grid disks are re-created on the new flash disk. If those grid disks were part of an Oracle ASM disk group, then they are added back to the disk group, and the data is rebalanced on them based on the disk group redundancy and `ASM_POWER_LIMIT` parameter.

#### See Also:

- [Performing a Hot Pluggable Replacement of a Flash Disk](#)
- `V$ASM_OPERATION` in *Oracle Database Reference*
- `ASM_POWER_LIMIT` in *Oracle Automatic Storage Management Administrator's Guide*
- The appropriate PCIe Card User's Guide for your system, which is listed in "[Related Documentation](#)"

## 3.4.2 About Flash Disk Degraded Performance Statuses

If a flash disk has degraded performance, you might need to replace the disk.

You may need to replace a flash disk because the disk is has one of the following statuses:

- warning - predictive failure
- warning - poor performance
- warning - write-through caching
- warning - peer failure



### Note:

For Oracle Exadata System Software releases earlier than release 11.2.3.2.2, the status is not present.

An alert is generated when a flash disk is in `predictive failure`, `poor performance`, `write-through caching` or `peer failure` status. The alert includes specific instructions for replacing the flash disk. If you have configured the system for alert notifications, then the alerts are sent by e-mail message to the designated address.

### predictive failure

Flash disk `predictive failure` status indicates that the flash disk will fail soon, and should be replaced at the earliest opportunity. If the flash disk is used for flash cache, then it continues to be used as flash cache. If the flash disk is used for grid disks, then the Oracle ASM disks associated with these grid disks are automatically dropped, and Oracle ASM rebalance relocates the data from the predictively failed disk to other disks.

When a flash disk has `predictive failure` due to one flash disk, then the data is copied. If the flash disk is used for grid disks, then Oracle ASM re-partners the associated partner, and does a rebalance. If the flash disk is used for write back flash cache, then the data is flushed from the flash disks to the grid disks.

To identify a `predictive failure` flash disk, use the following command:

```
CellCLI> LIST PHYSICALDISK WHERE DISKTYPE=flashdisk AND STATUS= \
'warning - predictive failure' DETAIL
```

```
name: FLASH_1_1
deviceName: /dev/nvme3n1
diskType: FlashDisk
luns: 1_1
makeModel: "Oracle Flash Accelerator F160 PCIe Card"
physicalFirmware: 8DV1RA13
physicalInsertTime: 2016-11-30T21:24:45-08:00
physicalSerial: CVMD519000251P6KGN
physicalSize: 1.4554837569594383T
```

```
slotNumber: "PCI Slot: 1; FDOM: 1"
status: warning - predictive failure
```

### poor performance

Flash disk `poor performance` status indicates that the flash disk demonstrates extremely poor performance, and should be replaced at the earliest opportunity. Starting with Oracle Exadata System Software release 11.2.3.2, an under-performing disk is automatically identified and removed from active configuration. If the flash disk is used for flash cache, then flash cache is dropped from this disk thus reducing the effective flash cache size for the storage server. If the flash disk is used for grid disks, then the Oracle ASM disks associated with the grid disks on this flash disk are automatically dropped with `FORCE` option, if possible. If `DROP...FORCE` cannot succeed due to offline partners, then the grid disks are automatically dropped normally, and Oracle ASM rebalance relocates the data from the poor performance disk to other disks.

Oracle Exadata Database Machine then runs a set of performance tests. When poor disk performance is detected by CELLSRV, the cell disk status changes to `normal - confinedOnline`, and the physical disk status changes to `warning - confinedOnline`. The following conditions trigger disk confinement:

- Disk stopped responding. The cause code in the storage alert log is `CD_PERF_HANG`.
- Slow cell disk such as the following:
  - High service time threshold (cause code `CD_PERF_SLOW_ABS`)
  - High relative service time threshold (cause code `CD_PERF_SLOW_RLTV`)
- High read or write latency such as the following:
  - High latency on writes (cause code `CD_PERF_SLOW_LAT_WT`)
  - High latency on reads (cause code `CD_PERF_SLOW_LAT_RD`)
  - High latency on reads and writes (cause code `CD_PERF_SLOW_LAT_RW`)
  - Very high absolute latency on individual I/Os happening frequently (cause code `CD_PERF_SLOW_LAT_ERR`)
- Errors such as I/O errors (cause code `CD_PERF_IOERR`).

If the disk problem is temporary and passes the tests, then it is brought back into the configuration. If the disk does not pass the tests, then it is marked as `poor performance`, and Oracle Auto Service Request (ASR) submits a service request to replace the disk. If possible, Oracle ASM takes the grid disks offline for testing. If Oracle ASM cannot take the disks offline, then the cell disk status stays at `normal - confinedOnline` until the disks can be taken offline safely.

To identify a `poor performance` flash disk, use the following command:

```
CellCLI> LIST PHYSICALDISK WHERE DISKTYPE=flashdisk AND STATUS= \
'warning - poor performance' DETAIL
```

```
name: FLASH_1_4
diskType: FlashDisk
luns: 1_4
makeModel: "Sun Flash Accelerator F20 PCIe Card"
physicalFirmware: D20Y
physicalInsertTime: 2012-09-27T13:11:16-07:00
```

```
physicalSerial: 508002000092e70FMOD2
physicalSize: 22.8880615234375G
slotNumber: "PCI Slot: 1; FDOM: 3"
status: warning - poor performance
```

The disk status change is associated with the following entry in the cell alert history:

```
MESSAGE ID date_time info "Hard disk entered confinement status. The
LUN
 n_m changed status to warning - confinedOnline. CellDisk changed
status to normal
 - confinedOnline. Status: WARNING - CONFINEDONLINE Manufacturer:
name Model
 Number: model Size: size Serial Number: serial_number Firmware:
fw_release
 Slot Number: m Cell Disk: cell_disk_name Grid Disk: grid disk 1,
grid disk 2
 ... Reason for confinement: threshold for service time exceeded"
```

The following would be logged in the storage server alert log:

```
CDHS: Mark cd health state change cell_disk_name with newState
HEALTH_BAD_
ONLINE pending HEALTH_BAD_ONLINE ongoing INVALID cur HEALTH_GOOD
Celldisk entering CONFINE ACTIVE state with cause CD_PERF_SLOW_ABS
activeForced: 0
inactiveForced: 0 trigger HistoryFail: 0, forceTestOutcome: 0
testFail: 0
global conf related state: numHDsConf: 1 numFDsConf: 0 numHDsHung: 0
numFDsHung: 0
...
```

#### Note:

In Oracle Exadata System Software releases earlier than release 11.2.3.2, use the `CALIBRATE` command to identify a bad flash disk, and look for very low throughput and IOPS for each flash disk.

If a flash disk exhibits extremely poor performance, then it is marked as `poor` performance. The flash cache on that flash disk is automatically disabled, and the grid disks on that flash disk are automatically dropped from the Oracle ASM disk group.

### write-through caching

Flash disk `write-through caching` status indicates the capacitors used to support data cache on the PCIe card have failed, and the card should be replaced as soon as possible.

### peer failure

Flash disk `peer failure` status indicates one of the flash disks on the same Sun Flash Accelerator PCIe card has failed or has a problem. For example, if `FLASH_5_3` fails, then `FLASH_5_0`, `FLASH_5_1`, and `FLASH_5_2` have peer failure status. The following is an example:

```
CellCLI> LIST PHYSICALDISK
 36:0 L45F3A normal
 36:1 L45WAE normal
 36:2 L45WQW normal
 ...
 FLASH_5_0 5L0034XM warning - peer failure
 FLASH_5_1 5L0034JE warning - peer failure
 FLASH_5_2 5L002WJH warning - peer failure
 FLASH_5_3 5L002X4P failed
```

When `CellSRV` detects a predictive or peer failure in any flash disk used for write back flash cache and only one `FDOM` is bad, then the data on the bad `FDOM` is resilvered, and the data on the other three `FDOMs` is flushed. `CellSRV` then initiates an Oracle ASM rebalance for the disks if there are valid grid disks. The bad disk cannot be replaced until the tasks are completed. `MS` sends an alert when the disk can be replaced.

## 3.4.3 Replacing a Flash Disk Due to Flash Disk Problems

Oracle Exadata Storage Server is equipped with four PCIe cards. Each card has four flash disks (`FDOMs`) for a total of 16 flash disks. The four PCIe cards are present on PCI slot numbers 1, 2, 4, and 5. Starting with Oracle Exadata Database Machine X7, you can replace the PCIe cards without powering down the storage server. See [Performing a Hot Pluggable Replacement of a Flash Disk](#).

In Oracle Exadata Database Machine X6 and earlier systems, the PCIe cards are not hot-pluggable. The Oracle Exadata Storage Server must be powered down before replacing the flash disks or cards.

Starting with Oracle Exadata Database Machine X7, each flash card on both High Capacity and Extreme Flash storage servers is a field-replaceable unit (`FRU`). The flash cards are also hot-pluggable, so you do not have to shut down the storage server before removing the flash card.

On Oracle Exadata Database Machine X5 and X6 systems, each flash card on High Capacity and each flash drive on Extreme Flash are `FRUs`. This means that there is no peer failure for these systems.

On Oracle Exadata Database Machine X3 and X4 systems, because the flash card itself is a `FRU`, if any `FDOMs` were to fail, the Oracle Exadata System Software would automatically put the rest of `FDOMs` on that card to peer failure so that the data can be moved out to prepare for the flash card replacement.

On Oracle Exadata Database Machine V2 and X2 systems, each `FDOM` is a `FRU`. There is no peer failure for flash for these systems.

Determining when to proceed with disk replacement depends on the release, as described in the following:

- For Oracle Exadata System Software releases earlier than 11.2.3.2:



Wait until the Oracle ASM disks have been successfully dropped by querying the `V$ASM_DISK_STAT` view before proceeding with the flash disk replacement. If the normal drop did not complete before the flash disk fails, then the Oracle ASM disks are automatically dropped with the `FORCE` option from the Oracle ASM disk group. If the `DROP` command did not complete before the flash disk fails, then refer to [Replacing a Flash Disk Due to Flash Disk Failure](#).

- For Oracle Exadata System Software releases 11.2.3.2 and later:

An alert is sent when the Oracle ASM disks have been dropped, and the flash disk can be safely replaced. If the flash disk is used for write-back flash cache, then wait until none of the grid disks are cached by the flash disk. Use the following command to check the `cachedBy` attribute of all the grid disks. The cell disk on the flash disk should not appear in any grid disk's `cachedBy` attribute.

```
CellCLI> LIST GRIDDISK ATTRIBUTES name, cachedBy
```

If the flash disk is used for both grid disks and flash cache, then wait until receiving the alert, and the cell disk is not shown in any grid disk's `cachedBy` attribute.

The following procedure describes how to replace a flash disk on High Capacity storage servers for Oracle Exadata Database Machine X6 and earlier due to disk problems.

#### Note:

On Extreme Flash storage servers for Oracle Exadata Database Machine X6 and all storage servers for Oracle Exadata Database Machine X7 and later, you can just remove the flash disk from the front panel and insert a new one. You do not need to shut down the storage server.

1. Stop the cell services using the following command:

```
CellCLI> ALTER CELL SHUTDOWN SERVICES ALL
```

The preceding command checks if any disks are offline, in `predictive failure` status or need to be copied to its mirror. If Oracle ASM redundancy is intact, then the command takes the grid disks offline in Oracle ASM, and then stops the cell services. If the following error is displayed, then it may not be safe to stop the cell services because a disk group may be forced to dismount due to redundancy.

```
Stopping the RS, CELLSRV, and MS services...
The SHUTDOWN of ALL services was not successful.
CELL-01548: Unable to shut down CELLSRV because disk group DATA,
RECO may be
forced to dismount due to reduced redundancy.
Getting the state of CELLSRV services... running
Getting the state of MS services... running
Getting the state of RS services... running
```

If the error occurs, then restore Oracle ASM disk group redundancy and retry the command when disk status is back to normal for all the disks.

2. Shut down the storage server.
3. Replace the failed flash disk based on the PCI number and FDOM number. A white Locator LED is lit to help locate the affected storage server.
4. Power up the storage server. The cell services are started automatically. As part of the storage server startup, all grid disks are automatically `ONLINE` in Oracle ASM.
5. Verify that all grid disks have been successfully put online using the following command:

```
CellCLI> LIST GRIDDISK ATTRIBUTES name, asmmodestatus
```

Wait until `asmmodestatus` shows `ONLINE` or `UNUSED` for all grid disks.

The new flash disk is automatically used by the system. If the flash disk is used for flash cache, then the effective cache size increases. If the flash disk is used for grid disks, then the grid disks are re-created on the new flash disk. If those grid disks were part of an Oracle ASM disk group, then they are added back to the disk group, and the data is rebalanced on them based on the disk group redundancy and the `ASM_POWER_LIMIT` parameter.

### Related Topics

- [Replacing a Flash Disk Due to Flash Disk Failure](#)  
Each Oracle Exadata Storage Server is equipped with flash devices.
- [Shutting Down Exadata Storage Server](#)  
When performing maintenance on Exadata Storage Servers, it may be necessary to power down or restart the cell.
- `V$ASM_OPERATION`

#### See Also:

- [Oracle Automatic Storage Management Administrator's Guide](#) for information about the `ASM_POWER_LIMIT` parameter
- The appropriate PCIe Card User's Guide for your system, which is listed in [Related Documentation](#)

## 3.4.4 Performing a Hot Pluggable Replacement of a Flash Disk

Starting with Oracle Exadata Database Machine X7, flash disks are hot-pluggable on both Extreme Flash (EF) and High Capacity (HC) storage servers.

Also, for Oracle Exadata Database Machine X6 and earlier, the flash devices on EF storage servers are hot-pluggable. However, for HC storage servers on Oracle Exadata Database Machine X6 and earlier systems, you must power down the storage servers before replacing the flash disks.

To replace a hot-pluggable flash disk device:

1. If necessary, prepare the disk for hot-pluggable replacement.  
Typically, you will replace a flash drive only after Oracle Exadata System Software identifies a problem and sets the device status to `failed - dropped` for replacement, which indicates the flash disk is ready for online replacement.

If you need to replace a flash disk that is in another state, then you must first use the CellCLI `ALTER PHYSICALDISK` command with the `DROP FOR REPLACEMENT` clause to prepare the disk for hot-pluggable replacement.

2. Verify that the flash disk is ready for hot-pluggable replacement.

Verify that the device status is failed - dropped for replacement.

You can check the device status by using the CellCLI `LIST PHYSICALDISK` command. For example:

```
CellCLI> LIST PHYSICALDISK WHERE DISKTYPE=flashdisk AND STATUS LIKE
'.*dropped
for replacement.*' DETAIL
```

```

name: FLASH_6_1
deviceName: /dev/nvme0n1
diskType: FlashDisk
luns: 6_0
makeModel: "Oracle Flash Accelerator F640 PCIe
Card"
physicalFirmware: QDV1RD09
physicalInsertTime: 2017-08-11T12:25:00-07:00
physicalSerial: PHLE6514003R6P4BGN-1
physicalSize: 2.910957656800747T
slotNumber: "PCI Slot: 6; FDOM: 1"
status: failed - dropped for replacement
```

Note that when one FDOM fails, the affected flash disk PCI card is considered failed and the whole card must be replaced.

3. Physically locate the flash disk device.

Use the `slotNumber` information from the CellCLI `LIST PHYSICALDISK` command output to help identify the PCI slot containing the affected flash device.

Also, a white Locator LED is lit to help locate the affected storage server. An amber Fault-Service Required LED is lit to identify the affected flash device.

4. Make sure that the flash disk power LED is off.

#### Caution:

Removing a flash device when the power LED is on could result in a system crash. If the status is failed - dropped for replacement but the power LED is still on, contact Oracle Support Services.

5. Remove and replace the flash disk device.

#### Related Topics

- ALTER PHYSICALDISK

## 3.4.5 Enabling Write Back Flash Cache

Write operations serviced by flash instead of by disk are referred to as **write-back flash cache**.

Starting with Oracle Exadata System Software release 11.2.3.2.1, Exadata Smart Flash Cache can transparently cache frequently-accessed data to fast solid-state storage, improving query response times and throughput.

- [Enable Write Back Flash Cache for 11.2.3.3.1 or Higher](#)  
Enable write back Flash Cache on the storage servers to improve query response times and throughput.
- [Enabling Write Back Flash Cache on a Rolling Basis for Software Versions Lower Than 11.2.3.3.1](#)  
You can enable Write Back Flash Cache on a rolling basis.
- [Enabling Write Back Flash Cache in a Non-Rolling Basis for Software Versions Lower Than 11.2.3.3.1](#)  
You can enable Write Back Flash Cache on a non-rolling basis.

### 3.4.5.1 Enable Write Back Flash Cache for 11.2.3.3.1 or Higher

Enable write back Flash Cache on the storage servers to improve query response times and throughput.

For Oracle Exadata System Software release 11.2.3.3.1 or higher, you do not have to stop cell services or inactivate grid disks when changing the Flash Cache from Write Through mode to Write Back mode.



#### Note:

Any time the Flash Cache is dropped and re-created, there is a performance impact for database operations. While the Flash Cache is being repopulated, there are more cache misses, which impacts database performance.

1. Validate all the physical disks are in NORMAL state before modifying Exadata Smart Flash Cache.

The following command should return no rows:

```
dcli -l root -g cell_group cellcli -e "list physicaldisk attributes
name,status"|grep -v NORMAL
```

2. Drop the Flash Cache.

```
dcli -l root -g cell_group cellcli -e drop flashcache
```

3. Set the `flashCacheMode` attribute to `writeback`.

```
dcli -l root -g cell_group cellcli -e "alter cell
flashCacheMode=writeback"
```

4. Re-create the Flash Cache.

```
dcli -l root -g cell_group cellcli -e create flashcache all
```

5. Verify the `flashCacheMode` has been set to `writeback`.

```
dcli -l root -g cell_group cellcli -e list cell detail | grep flashCacheMode
```

6. Validate the grid disk attributes `cachingPolicy` and `cachedby`.

```
cellcli -e list griddisk attributes name,cachingpolicy,cachedby
```

### 3.4.5.2 Enabling Write Back Flash Cache on a Rolling Basis for Software Versions Lower Than 11.2.3.3.1

You can enable Write Back Flash Cache on a rolling basis.

To modify the Flash Cache attribute from `writethrough` to `writeback`, the Flash Cache must be dropped first. For Oracle Exadata System Software releases prior to 11.2.3.3.1, you must stop cell services or inactivate grid disks when enabling Write Back Flash Cache.

There is a shell script to automate enabling and disabling Write Back Flash Cache. Refer to My Oracle Support note 1500257.1 for the script and additional information.

#### Note:

Any time the Flash Cache is dropped and re-created, there is a performance impact for database operations. While the Flash Cache is being repopulated, there are more cache misses, which impacts database performance.

Oracle Grid Infrastructure homes and Oracle Database homes must at release 11.2.0.3 BP9 or higher to use write-back Flash Cache. Refer to My Oracle Support note 888828.1 for the minimum release requirements for Oracle Exadata System Software, Oracle Grid Infrastructure home, and Oracle Database home.

1. Log in as the `root` user to the first cell to be enabled for Write Back Flash Cache.
2. Check that the Flash Cache is in normal state and no flash disks are degraded or in a critical state.

```
cellcli -e LIST FLASHCACHE detail
```

3. Drop the Flash Cache on the cell.

```
cellcli -e DROP FLASHCACHE
```

4. Inactivate the grid disks on the cell.

```
cellcli -e ALTER GRIDDISK ALL INACTIVE
```

5. Shut down `CELLSRV` services.

```
cellcli -e ALTER CELL SHUTDOWN SERVICES CELLSRV
```

6. Set the `flashCacheMode` attribute to `writeback`.

```
cellcli -e "ALTER CELL FLASHCACHEMODE=writeback"
```

7. Restart cell services.

```
cellcli -e ALTER CELL STARTUP SERVICES CELLSRV
```

8. Reactivate the grid disks on the cell.

```
cellcli -e ALTER GRIDDISK ALL ACTIVE
```

9. Re-create the Flash Cache.

```
cellcli -e CREATE FLASHCACHE ALL
```

10. Check the status of the cell.

```
cellcli -e LIST CELL DETAIL | grep flashCacheMode
```

The `flashCacheMode` attribute should be set to `writeback`.

11. Check the grid disk attributes `asmDeactivationOutcome` and `asmModeStatus` before moving to the next cell using the following command:

```
CellCLI> LIST GRIDDISK ATTRIBUTES
name,asmdeactivationoutcome,asmmodestatus
```

The `asmDeactivationOutcome` attribute should be `yes`, and the `asmModeStatus` attribute should be `online`.

12. Repeat the preceding steps on the next cell.

#### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

### 3.4.5.3 Enabling Write Back Flash Cache in a Non-Rolling Basis for Software Versions Lower Than 11.2.3.3.1

You can enable Write Back Flash Cache on a non-rolling basis.

The cell services must be shut down before changing the `flashCacheMode` attribute. For Oracle Exadata System Software releases prior to 11.2.3.3.1, you must stop cell services when disabling Write Back Flash Cache in a non-rolling manner.

There is a shell script to automate enabling and disabling write back Flash Cache. Refer to My Oracle Support note 1500257.1 for the script and additional information.

Oracle Grid Infrastructure homes and Oracle Database homes must at release 11.2.0.3 BP9 or higher to use write-back Flash Cache. Refer to My Oracle Support note 888828.1 for the minimum release requirements for Oracle Exadata System Software, Oracle Grid Infrastructure home, and Oracle Database home.

1. Log in as the `root` user to a database node.
2. Shut down the entire cluster.

```
cd $Grid_home/bin
./crsctl stop cluster -all
```

3. Drop the Flash Cache for all cells.

```
dcli -g cell_group -l root cellcli -e DROP FLASHCACHE
```

4. Shut down CELLSRV services.

```
dcli -g cell_group -l root cellcli -e ALTER CELL SHUTDOWN
SERVICES CELLSRV
```

5. Confirm that the Flash Cache is in `writethrough` mode.

```
dcli -g cell_group -l root "cellcli -e list cell detail | grep -i
flashcachemode"
```

6. Set the `flashCacheMode` attribute to `writeback`.

```
dcli -g cell_group -l root cellcli -e "ALTER CELL
FLASHCACHEMODE=writeback"
```

7. Restart cell services.

```
dcli -g cell_group -l root cellcli -e ALTER CELL STARTUP SERVICES
CELLSRV
```

8. Re-create the Flash Cache.

```
dcli -g cell_group -l root cellcli -e CREATE FLASHCACHE ALL
```

9. Restart the cluster:

```
cd $Grid_home/bin
./crsctl start cluster -all
```

## 3.4.6 Disabling Write Back Flash Cache

You can disable the Write-Back Flash Cache by enabling Write-Through Flash Cache.

Starting with Oracle Exadata System Software release 11.2.3.2.1, Exadata Smart Flash Cache can transparently cache frequently-accessed data to fast solid-state storage, improving query response times and throughput.

Write operations serviced by flash instead of by disk are referred to as **write back flash cache**.

- [Disable Write Back Flash Cache for Exadata X8M or Later Servers](#)  
To change the FlashCache mode from WriteBack to WriteThrough on an Exadata Database Machine X8M or later system with PMEM cache, if the PMEM cache is in WriteBack mode, you first need to modify the PMEM cache.
- [Disable Write Back Flash Cache for 11.2.3.3.1 or Higher](#)  
You can disable write back Flash Cache on the storage servers by changing the mode to Write Through.
- [Disabling Write Back Flash Cache on a Rolling Basis for Software Versions Lower Than 11.2.3.3.1](#)  
You can enable Write Back Flash Cache on each storage server using a rolling method.
- [Disabling Write Back Flash Cache on a Non-Rolling Basis for Software Versions Lower Than 11.2.3.3.1](#)  
You can enable Write Back Flash Cache on a non-rolling basis.

### 3.4.6.1 Disable Write Back Flash Cache for Exadata X8M or Later Servers

To change the FlashCache mode from WriteBack to WriteThrough on an Exadata Database Machine X8M or later system with PMEM cache, if the PMEM cache is in WriteBack mode, you first need to modify the PMEM cache.



#### Note:

To reduce the performance impact on the application, disable the write back flash cache during a period of reduced workload.

1. For X8M or later systems with PMEM cache, if the PMEM Cache is in WriteBack mode:

- a. Flush the PMEM cache.

If the PMEM cache utilizes all available PMEM cell disks, you can use the `ALL` keyword as shown here.

```
dcli -l root -g cell_group cellcli ALTER PMEMCACHE ALL FLUSH
```

Otherwise, list the specific disks using the `CELLDISK="cdisk1 [,cdisk2] ..."` clause.

- b. Drop the PMEM cache.

```
dcli -l root -g cell_group cellcli DROP PMEMCACHE
```

- c. Modify the PMEM cache to use WriteThrough mode.

```
dcli -l root -g cell_group cellcli ALTER CELL
pmemCacheMode=WriteThrough
```

- d. Re-create the PMEM cache.

If the PMEM cache utilizes all available PMEM cell disks, you can use the `ALL` keyword as shown here. Otherwise, list the specific disks using the `CELLDISK="cdisk1 [,cdisk2] ..."` clause. If the `size` attribute is not specified, then



the maximum size is allocated. All available space on each cell disk in the list is used for PMEM Cache.

```
dcli -l root -g cell_group cellcli -e CREATE PMEMCACHE ALL
```

- e. Verify the `pmemCacheMode` has been set to `writethrough`.

```
dcli -l root -g cell_group cellcli -e list cell detail | grep
pmemCacheMode
```

2. Validate all the physical disks are in `NORMAL` state before modifying the `FlashCache`.

```
dcli -l root -g cell_group cellcli -e "LIST PHYSICALDISK
ATTRIBUTES name,status"
|grep -v NORMAL
```

The command should return no rows.

3. Determine amount of dirty data in the flash cache.

```
cellcli -e "LIST METRICCURRENT ATTRIBUTES name,metricvalue WHERE
name LIKE '\FC_BY_DIRTY.*\ ' "
```

4. Flush the Flash cache.

If the Flash cache utilizes all available Flash cell disks, you can use the `ALL` keyword instead of listing the flash disks.

```
dcli -g cell_group -l root cellcli -e "ALTER FLASHCACHE
CELLDISK='\FD_02_dm01celadm12,
FD_03_dm01celadm12,FD_00_dm01celadm12,FD_01_dm01celadm12\' FLUSH"
```

5. Check the progress of the flushing of flash cache.

The flushing process is complete when `FC_BY_DIRTY` is 0 MB.

```
dcli -g cell_group -l root cellcli -e "LIST METRICCURRENT
ATTRIBUTES name,metricvalue
WHERE name LIKE '\FC_BY_DIRTY.*\ ' "
```

Or, you can check to see if the attribute `flushstatus` has been set to `Completed`.

```
dcli -g cell_group -l root cellcli -e "LIST CELLDISK ATTRIBUTES
name, flushstatus,
flusherror" | grep FD
```

6. After flushing of the flash cache completes, drop the flash cache.

```
dcli -g cell_group -l root cellcli -e drop flashcache
```

7. Modify the Flash cache to use WriteThrough mode.

```
dcli -g cell_group -l root cellcli -e "ALTER CELL
flashCacheMode=writethrough"
```

8. Re-create the Flash cache.

If the Flash cache utilizes all available Flash cell disks, you can use the `ALL` keyword instead of listing the cell disks.

If you do not include the `size` attribute, then all available space on each cell disk in the list is used for Exadata Smart Flash Cache.

```
dcli -l root -g cell_group cellcli -e "create flashcache
celldisk='\FD_02_dm01celadm12,
FD_03_dm01celadm12,FD_00_dm01celadm12,FD_01_dm01celadm12\'
```

9. Verify the `flashCacheMode` has been set to `writethrough`.

```
dcli -l root -g cell_group cellcli -e list cell detail | grep
flashCacheMode
```

### 3.4.6.2 Disable Write Back Flash Cache for 11.2.3.3.1 or Higher

You can disable write back Flash Cache on the storage servers by changing the mode to Write Through.

With release 11.2.3.3.1 or higher, you do not have to stop the `CELLSRV` process or inactivate grid disks.



#### Note:

To reduce the performance impact on the application, disable the write back flash cache during a period of reduced workload.

1. Validate all the Physical Disks are in `NORMAL` state before modifying FlashCache.

The following command should return no rows:

```
dcli -l root -g cell_group cellcli -e "LIST PHYSICALDISK ATTRIBUTES
name,status"|grep -v NORMAL
```

2. Determine amount of dirty data in the flash cache.

```
cellcli -e "LIST METRICCURRENT ATTRIBUTES name,metricvalue WHERE name
LIKE '\FC_BY_DIRTY.*\'
```

3. Flush the flash cache .

```
dcli -g cell_group -l root cellcli -e "ALTER FLASHCACHE ALL FLUSH"
```

4. Check the progress of the flushing of flash cache.

The flushing process is complete when FC\_BY\_DIRTY is 0 MB.

```
dcli -g cell_group -l root cellcli -e "LIST METRICCURRENT
ATTRIBUTES name,metricvalue WHERE name LIKE \'FC_BY_DIRTY.*\' "
```

Or, you can check to see if the attribute `flushstatus` has been set to `Completed`.

```
dcli -g cell_group -l root cellcli -e "LIST CELLDISK ATTRIBUTES
name, flushstatus, flusherror" | grep FD
```

5. After flushing of the flash cache completes, drop the flash cache.

```
dcli -g cell_group -l root cellcli -e drop flashcache
```

6. Set the `flashCacheMode` attribute to `writethrough`.

```
dcli -g cell_group -l root cellcli -e "ALTER CELL
flashCacheMode=writethrough"
```

7. Re-create the flash cache.

```
dcli -l root -g cell_group cellcli -e create flashcache all
```

8. Verify the `flashCacheMode` has been set to `writethrough`.

```
dcli -l root -g cell_group cellcli -e list cell detail | grep
flashCacheMode
```

### 3.4.6.3 Disabling Write Back Flash Cache on a Rolling Basis for Software Versions Lower Than 11.2.3.3.1

You can enable Write Back Flash Cache on each storage server using a rolling method.

The cell services must be shut down before changing the `flashCacheMode` attribute. The cell services can shut down on rolling basis. The flash cache must be flushed and dropped before changing the attribute to `writethrough`. After the flush operation begins, all caching to the flash cache stops. Ensure cell services resynchronization is complete on the current storage server before changing the next storage server.

There is a shell script to automate enabling and disabling write back Flash Cache. Refer to My Oracle Support note 1500257.1 for the script and additional information.

1. Log in as the `root` user to the first cell to be disabled for write back flash cache.
2. Verify the `asmDeactivationOutcome` attribute is `yes` for all grid disks on the cell.

```
dcli -g cell_group -l root cellcli -e "LIST GRIDDISK WHERE \
asmdeactivationoutcome != 'Yes' attributes name, \
asmdeactivationoutcome, \
asmmodestatus"
```

The grid disk attribute `asmDeactivationOutcome` must be `yes`, and the `asmModeStatus` attribute must be `online` for all grid disks on the current cell before moving to the next cell. If a grid disk does not have an `asmDeactivationOutcome` attribute value of `yes`, then you must resolve this issue before proceeding.

3. Check the amount of dirty data in the flash cache.

```
dcli -g cell_group -l root cellcli -e "LIST METRICCURRENT ATTRIBUTES \
name,metricvalue WHERE name LIKE \'FC_BY_DIRTY.*\'"
```

4. Flush the flash cache.

```
dcli -g cell_group -l root cellcli -e ALTER FLASHCACHE ALL FLUSH
```

5. Check the status of the flash cache.

```
dcli -g cell_group -l root cellcli -e LIST CELLDISK ATTRIBUTES name, \
flushstatus, flusherror | grep FD
```

The status shows `completed` when the flush is done.

6. Perform the following set of steps on for all cells, one cell at a time.

In other words, perform steps (a) through (i) on one cell, then perform the same steps on another cell until all the cells are done.

- a. Drop the flash cache.

```
cellcli -e DROP FLASHCACHE
```

- b. Inactivate all grid disks on the cell.

```
cellcli -e ALTER GRIDDISK ALL INACTIVE
```

- c. Shut down CELLSRV services.

```
cellcli -e ALTER CELL SHUTDOWN SERVICES CELLSRV
```

- d. Set the `flashCacheMode` attribute to `writethrough`.

```
cellcli -e "ALTER CELL FLASHCACHEMODE=writethrough"
```

- e. Restart cell services.

```
cellcli -e ALTER CELL STARTUP SERVICES CELLSRV
```

- f. Reactivate the grid disks on the cell.

```
cellcli -e ALTER GRIDDISK ALL ACTIVE
```

- g. Re-create the flash cache.

```
cellcli -e CREATE FLASHCACHE ALL
```

- h. Check the status of the cell.

```
cellcli -e LIST CELL DETAIL | grep flashCacheMode
```

- i. Check the grid disk attributes `asmDeactivationOutcome` and `asmModeStatus`.

```
cellcli -e LIST GRIDDISK ATTRIBUTES
name,status,asmdeactivationoutcome,asmmodestatus
```

The `asmDeactivationOutcome` attribute should be `yes`, and the `asmModeStatus` attribute should be `online`.

If the disk status is `SYNCING`, wait until it is `ACTIVE` before proceeding.

### 3.4.6.4 Disabling Write Back Flash Cache on a Non-Rolling Basis for Software Versions Lower Than 11.2.3.3.1

You can enable Write Back Flash Cache on a non-rolling basis.

When changing the Flash Cache mode on a non-rolling basis, ensure the entire cluster is shut down, including the Oracle Clusterware stack and all the databases. The cell services must be shut down before changing the `flashCacheMode` attribute. The Flash Cache must be flushed and dropped before changing the attribute to `writethrough`. The Flash Cache flush operation can be performed prior to shutting down the entire cluster. After the flush operation begins, all caching to the Flash Cache stops.

1. Log in as the `root` user to the first database node to be disabled for write back Flash Cache.
2. Check the amount of dirty data in the Flash Cache using the following command:

```
dcli -g cell_group -l root cellcli -e "LIST METRICCURRENT
ATTRIBUTES \
 name,metricvalue WHERE name LIKE \'FC_BY_DIRTY.*\'"
```

3. Flush the Flash Cache using the following command:

```
dcli -g cell_group -l root cellcli -e ALTER FLASHCACHE ALL FLUSH
```

4. Check the status as the blocks are moved to disk using the following command. The count reduces to zero.

```
dcli -g cell_group -l root cellcli -e "LIST METRICCURRENT
ATTRIBUTES name, \
 metricvalue WHERE NAME LIKE \'FC_BY_DIRTY.*\'"
```

5. Check the status of the flash disks using the following command:

```
dcli -g cell_group -l root cellcli -e LIST CELLDISK ATTRIBUTES
name, flushstatus, flusherror | grep FD
```

The status shows `completed` when the flush is done.

6. Shut down the database and the entire cluster using the following commands:

```
cd Grid_home/bin
./crsctl stop cluster -all
```

7. Drop the Flash Cache across all cells using the following command:

```
dcli -g cell_group -l root cellcli -e DROP FLASHCACHE
```

8. Shut down CELLSRV services using the following command:

```
dcli -g cell_group -l root cellcli -e ALTER CELL SHUTDOWN SERVICES
CELLSRV
```

9. Set the `flashCacheMode` attribute to `writethrough` using the following command:

```
dcli -g cell_group -l root cellcli -e "ALTER CELL
FLASHCACHEMODE=writethrough"
```

10. Restart cell services using the following command:

```
dcli -g cell_group -l root cellcli -e ALTER CELL STARTUP SERVICES
CELLSRV
```

11. Re-create the Flash Cache using the following command:

```
dcli -g cell_group -l root cellcli -e CREATE FLASHCACHE ALL
```

12. Check the Flash Cache mode of the cells using the following command:

```
dcli -g cell_group -l root cellcli -e LIST CELL DETAIL | grep
flashCacheMode
```

13. Restart the cluster and database using the following commands:

```
cd Grid_home/bin
./crsctl start cluster -all
```

### 3.4.7 Enabling Flash Cache Compression

Flash cache compression can be enabled on Oracle Exadata Database Machine X4-2, Oracle Exadata Database Machine X3-2, and Oracle Exadata Database Machine X3-8 Full Rack systems. Oracle Exadata Database Machine X5-2, X5-8, and later systems do not have flash cache compression. Flash cache compression dynamically increases the logical capacity of the flash cache by transparently compressing user data as it is loaded into the flash cache.

 **Note:**

- Oracle Advanced Compression Option is required to enable flash cache compression.
- User data is not retained when enabling flash cache compression.

The following procedure describes how to enable flash cache compression:

1. Perform this step only if writeback flash cache is enabled. If writeback flash cache is not enabled, skip this step. Performing this step when writeback flash cache is not enabled can result in error messages. You can check the flash cache mode by running the following command:

```
cellcli -e LIST CELL DETAIL | grep flashCacheMode
```

If writeback flash cache is enabled, then save the user data on the flash cell disks.

```
cellcli -e ALTER FLASHCACHE ALL FLUSH
```

During the flash operation, the `flushstatus` attribute has a value of `working`. When the flush operation completes successfully, the value is changed to `complete`. For grid disks, the attribute `cachedby` should be `null`. Also, the number of dirty buffers (unflushed) will be 0 after flush is complete.

```
cellcli -e LIST METRICCURRENT FC_BY_DIRTY
FC_BY_DIRTY FLASHCACHE 0.000 MB
```

2. Remove the flash cache from the cell.

```
cellcli -e DROP FLASHCACHE ALL
```

3. Remove the flash log from the cell.

```
cellcli -e DROP FLASHLOG ALL
```

4. Drop the cell disks on the flash disks.

```
cellcli -e DROP CELLDISK ALL FLASHDISK
```

5. Enable flash cache compression using the following commands, based on the system:

- For Oracle Exadata Database Machine X3-2, Oracle Exadata Database Machine X3-8, and Oracle Exadata Database Machine X4-2 storage servers with Oracle Exadata System Software release 11.2.3.3.1 or higher:

```
cellcli -e ALTER CELL flashcachecompress=true
```

- For Oracle Exadata Database Machine X3-2 storage servers with Oracle Exadata System Software release 11.2.3.3.0:

```
cellcli -e ALTER CELL flashCacheCompX3Support=true
cellcli -e ALTER CELL flashCacheCompress=true
```

6. Verify the size of the physical disks has increased.

```
cellcli -e LIST PHYSICALDISK attributes name,physicalSize,status WHERE
disktype=flashdisk
```

The status should be normal. Use the following information to validate the expected size when Compression is ON:

- Aura 2.0/F40/X3:
  - Physical Disk Size: 93.13 G (OFF) or 186.26 G (ON)
  - Flash Cache Size: 1489 G (OFF) or 2979 G (ON)
- Aura 2.1/F80/X4:
  - Physical Disk Size: 186.26 G (OFF) or 372.53 G (ON)
  - Flash Cache Size: 2979 G (OFF) or 5959 G (ON)

7. Create the cell disks on the flash disks.

```
cellcli -e CREATE CELLDISK ALL FLASHDISK
CellDisk FD_00_exampleceladm18 successfully created
...
CellDisk FD_15_exampleceladm18 successfully created
```

8. Create the flash log.

```
cellcli -e CREATE FLASHLOG ALL
Flash log RaNdOmceladm18_FLASHLOG successfully created
```

9. Create the flash cache on the cell.

```
cellcli -e CREATE FLASHCACHE ALL
Flash cache exampleceladm18_FLASHCACHE successfully created
```

### 3.4.8 Monitoring Exadata Smart Flash Cache Usage Statistics

Use the following methods to monitor Exadata Smart Flash Cache usage:

- AWR report, in the **Exadata Statistics** section.
  - Under **Performance Summary** you can find various statistics related to Flash Cache and its benefits.
  - Under **Exadata Smart Statistics** there is a section for **Flash Cache** with several different reports on Exadata Smart Flash Cache statistics.
- ExaWatcher reports

Flash Cache size and read, write, and population operation related stats are exposed in the Cell Server Charts and in the FlashCache related stats section.



- Use the CellCLI `LIST` command to display and monitor metrics for the flash cache.

### Related Topics

- [Monitoring Smart Flash Cache](#)
- [Cell Server Charts](#)  
Cell server statistics are useful for tracking features that are specific to Exadata storage servers. This page displays statistics related to Smart Flash Cache and Smart IOs.

## 3.4.9 Disabling Flash Cache Compression

Flash cache compression can be disabled on Oracle Exadata Database Machine X4-2, Oracle Exadata Database Machine X3-2, and Oracle Exadata Database Machine X3-8 Full Rack systems. Oracle Exadata Database Machine X5-2, X5-8, and later systems do not have flash cache compression.

### Note:

- User data is not retained when disabling flash cache compression.

The following procedure describes how to disable flash cache compression:

1. Save the user data on the flash cell disks.

```
cellcli -e ALTER FLASHCACHE ALL FLUSH
```

For grid disks, the attribute `cachedby` should be `null`. Also, the number of dirty buffers (unflushed) will be 0 after flush is complete.

```
cellcli -e LIST METRICCURRENT FC_BY_DIRTY
FC_BY_DIRTY FLASHCACHE 0.000 MB
```

2. Remove the flash cache from the cell.

```
cellcli -e DROP FLASHCACHE ALL
```

3. Remove the flash log from the cell.

```
cellcli -e DROP FLASHLOG ALL
```

4. Drop the cell disks on the flash disks.

```
cellcli -e DROP CELLDISK ALL FLASHDISK
```

5. Disable Flash Cache Compression using the following commands, based on the system:

- If Exadata Storage Cell Server image is 11.2.3.3.1 or higher and the Exadata Storage Cell is X3-2 or X4-2:

```
cellcli -e ALTER CELL flashcachecompress=false
```

- If Exadata Storage Cell Server image is 11.2.3.3.0 and the Exadata Storage Cell is X3-2:

```
cellcli -e ALTER CELL flashCacheCompX3Support=true
cellcli -e ALTER CELL flashCacheCompress=false
```

 **Note:**

Note that `flashCacheCompress` is set to `false`, but `flashCacheCompX3Support` has to be set to `true`.

You can verify that Flash Cache Compress has been disabled by viewing the cell attributes:

```
cellcli -e LIST CELL attributes name,flashCacheCompress
```

Correct values are `FALSE` or a null string.

6. Verify the size of the physical disks has decreased.

```
cellcli -e LIST PHYSICALDISK attributes name,physicalSize,status WHERE
disktype=flashdisk
```

The status should be normal. Use the following information to validate the expected size when Compression is OFF:

- Aura 2.0/F40/X3:
  - Physical Disk Size: 93.13 G (OFF) or 186.26 G (ON)
  - Flash Cache Size: 1489 G (OFF) or 2979 G (ON)
- Aura 2.1/F80/X4:
  - Physical Disk Size: 186.26 G (OFF) or 372.53 G (ON)
  - Flash Cache Size: 2979 G (OFF) or 5959 G (ON)

7. Create the cell disks on the flash disks.

```
cellcli -e CREATE CELLDISK ALL FLASHDISK
CellDisk FD_00_exampleceladm18 successfully created
...
CellDisk FD_15_exampleceladm18 successfully created
```

8. Create the flash log.

```
cellcli -e CREATE FLASHLOG ALL
Flash log RaNdOmceladm18_FLASHLOG successfully created
```

Verify the flash log is in normal mode.

```
cellcli -e LIST FLASHLOG DETAIL
```

**9.** Create the flash cache on the cell.

```
cellcli -e CREATE FLASHCACHE ALL
Flash cache exampleceladm18_FLASHCACHE successfully created
```

Verify the flash cache is in normal mode.

```
cellcli -e LIST FLASHCACHE DETAIL
```

**10.** Verify that flash cache compression is disabled.

```
cellcli -e LIST CELL
```

The value of the `flashCacheCompress` attribute should be `false`.

## 3.5 Maintaining PMEM Devices on Oracle Exadata Storage Servers

Starting with Oracle Exadata X8M, High Capacity (HC) and Extreme Flash (EF) storage servers contain persistent memory (PMEM) devices.

If a PMEM device fails, Oracle Exadata System Software isolates the failed device and automatically recovers the PMEM cache if required.

If the PMEM cache is in write-back mode, the recovery operation, also known as resilvering, restores the lost data by reading a surviving mirrored copy. During resilvering, the grid disk status is `ACTIVE -- RESILVERING WORKING`. If the PMEM cache is in write-through mode, then the data in the failed PMEM device is already stored in the data grid disk, and no recovery is required.

- [Replacing a PMEM Device Due to Device Failure](#)  
If the PMEM device has a status of `Failed`, you should replace the PMEM device on the Oracle Exadata Storage Server.
- [Replacing a PMEM Device Due to Degraded Performance](#)  
If a PMEM device has degraded performance, you might need to replace the module.
- [Enabling and Disabling Write-Back PMEM Cache](#)  
Starting with Oracle Exadata System Software release 19.3, the PMEM cache can transparently cache frequently-accessed data to fast solid-state storage, improving query response times and throughput.

## 3.5.1 Replacing a PMEM Device Due to Device Failure

If the PMEM device has a status of `Failed`, you should replace the PMEM device on the Oracle Exadata Storage Server.

A PMEM fault could cause server to reboot. The failed device should be replaced with a new PMEM device at the earliest opportunity. Until the PMEM device is replaced, the effective PMEM cache size for the storage server is reduced. If the PMEM device is used for PMEM log, then PMEM log is disabled on the disk thus reducing the effective PMEM log size.

An alert is generated when a PMEM device failure is detected. The alert message includes the slot number and cell disk name. If you have configured the system for alert notification, then an alert is sent by e-mail message to the designated address.

To identify a failed PMEM device, you can also use the following command:

```
CellCLI> LIST PHYSICALDISK WHERE disktype=PMEM AND status=failed DETAIL
```

```
name: PMEM_0_1
diskType: PMEM
luns: P0_D1
makeModel: "Intel NMA1XBD128GQS"
physicalFirmware: 1.02.00.5365
physicalInsertTime: 2019-09-28T11:29:13-07:00
physicalSerial: 8089-A2-1838-00001234
physicalSize: 126.375G
slotNumber: "CPU: 0; DIMM: 1"
status: failed
```

In the above output, the `slotNumber` shows the socket number and DIMM slot number.

1. Locate the storage server that contains the failed PMEM device.

A white Locator LED is lit to help locate the affected storage server. When you have located the server, you can use the Fault Remind button to locate the failed DIMM.

### Caution:

Do not attempt to remove a faulty DCPMM DIMM when the Do Not Service LED indicator is illuminated.

2. Power down the storage server with the failed PMEM device and unplug the power cable for the server.
3. Replace the failed PMEM device.
  - X9-2: See "Servicing the DIMMs" in *Oracle Exadata Storage Server X9-2 EF, HC, and XT Service Manual* at <https://docs.oracle.com/en/servers/x86/x9-2/ixa-storage/servicing-dimms.html>
  - X8-2: See "Servicing the DIMMs" in *Oracle Exadata Storage Server X8-2 EF, HC, and XT Service Manual* at <https://docs.oracle.com/en/servers/x86/x8-2/exadata-storage-service-manual/gqtc.html>
4. Restart the storage server.

 **Note:**

During the restart, the storage server will shut down a second time to complete the initialization of the new PMEM device.

The new PMEM device is automatically used by the system. If the PMEM device is used for PMEM cache, then the effective PMEM cache size increases. If the PMEM device is used for PMEM log, then PMEM log is enabled on the device.

## 3.5.2 Replacing a PMEM Device Due to Degraded Performance

If a PMEM device has degraded performance, you might need to replace the module.

If degraded performance is detected on a PMEM device, the module status is set to `warning - predictive failure` and an alert is generated. The alert includes specific instructions for replacing the PMEM device. If you have configured the system for alert notifications, then the alerts are sent by e-mail message to the designated address.

The `predictive failure` status indicates that the PMEM device will fail soon, and should be replaced at the earliest opportunity. No new data is cached in the PMEM device until it is replaced.

To identify a PMEM device with the status `predictive failure`, you can also use the following command:

```
CellCLI> LIST PHYSICALDISK WHERE disktype=PMEM AND status='warning -
predictive failure' DETAIL
```

```
name: PMEM_0_6
diskType: PMEM
luns: P0_D6
makeModel: "Intel NMA1XBD128GQS"
physicalFirmware: 1.02.00.5365
physicalInsertTime: 2019-11-30T21:24:45-08:00
physicalSerial: 8089-A2-1838-00001234
physicalSize: 126.375G
slotNumber: "CPU: 0; DIMM: 6"
status: warning - predictive failure
```

You can also locate the PMEM device using the information in the `LIST DISKMAP` command:

```
CellCLI> LIST DISKMAP
```

Name	PhysicalSerial	SlotNumber	Status	PhysicalSize
CellDisk	DevicePartition	GridDisks		
PMEM_0_1	8089-a2-0000-00000460	"CPU: 0; DIMM: 1"	normal	126G
PM_00_cel01	/dev/dax5.0	PMEMCACHE_PM_00_cel01		
PMEM_0_3	8089-a2-0000-000004c2	"CPU: 0; DIMM: 3"	normal	126G
PM_02_cel01	/dev/dax4.0	PMEMCACHE_PM_02_cel01		
PMEM_0_5	8089-a2-0000-00000a77	"CPU: 0; DIMM: 5"	normal	126G
PM_03_cel01	/dev/dax3.0	PMEMCACHE_PM_03_cel01		
PMEM_0_6	8089-a2-0000-000006ff	"CPU: 0; DIMM: 6"	warning -	126G

```

PM_04_cel01 /dev/dax0.0 PMEMCACHE_PM_04_cel01
PMEM_0_8 8089-a2-0000-00000750 "CPU: 0; DIMM: 8" normal 126G
PM_05_cel01 /dev/dax1.0 PMEMCACHE_PM_05_cel01
PMEM_0_10 8089-a2-0000-00000103 "CPU: 0; DIMM: 10" normal 126G
PM_01_cel01 /dev/dax2.0 PMEMCACHE_PM_01_cel01
PMEM_1_1 8089-a2-0000-000008f6 "CPU: 1; DIMM: 1" normal 126G
PM_06_cel01 /dev/dax11.0 PMEMCACHE_PM_06_cel01
PMEM_1_3 8089-a2-0000-000003bb "CPU: 1; DIMM: 3" normal 126G
PM_08_cel01 /dev/dax10.0 PMEMCACHE_PM_08_cel01
PMEM_1_5 8089-a2-0000-00000708 "CPU: 1; DIMM: 5" normal 126G
PM_09_cel01 /dev/dax9.0 PMEMCACHE_PM_09_cel01
PMEM_1_6 8089-a2-0000-00000811 "CPU: 1; DIMM: 6" normal 126G
PM_10_cel01 /dev/dax6.0 PMEMCACHE_PM_10_cel01
PMEM_1_8 8089-a2-0000-00000829 "CPU: 1; DIMM: 8" normal 126G
PM_11_cel01 /dev/dax7.0 PMEMCACHE_PM_11_cel01
PMEM_1_10 8089-a2-0000-00000435 "CPU: 1; DIMM: 10" normal 126G
PM_07_cel01 /dev/dax8.0 PMEMCACHE_PM_07_cel01

```

If the PMEM device is used for write-back PMEM cache, then the data is flushed from the PMEM device to the flash cache. To ensure that data is flushed from the PMEM device, check the `cachedBy` attribute of all the grid disks and ensure that the affected PMEM device is not listed.

1. Locate the storage server that contains the failing PMEM device.

A white Locator LED is lit to help locate the affected storage server. When you have located the server, you can use the Fault Remind button to locate the failed DIMM.

### Caution:

Do not attempt to remove a faulty DCPMM DIMM when the Do Not Service LED indicator is illuminated.

2. Power down the storage server with the failing PMEM device and unplug the power cable for the server.
3. Replace the failing PMEM device.
  - X9-2: See "Servicing the DIMMs" in *Oracle Exadata Storage Server X9-2 EF, HC, and XT Service Manual* at <https://docs.oracle.com/en/servers/x86/x9-2/exa-storage/servicing-dimms.html>
  - X8-2: See "Servicing the DIMMs" in *Oracle Exadata Storage Server X8-2 EF, HC, and XT Service Manual* at <https://docs.oracle.com/en/servers/x86/x8-2/exadata-storage-service-manual/gqtc.html>
4. Restart the storage server.

### Note:

During the restart, the storage server will shut down a second time to complete the initialization of the new PMEM device.

The new PMEM device is automatically used by the system. If the PMEM device is used for PMEM cache, then the effective PMEM cache size increases. If the PMEM device is used for PMEM log, then PMEM log is enabled on the device.

### 3.5.3 Enabling and Disabling Write-Back PMEM Cache

Starting with Oracle Exadata System Software release 19.3, the PMEM cache can transparently cache frequently-accessed data to fast solid-state storage, improving query response times and throughput.

Write operations serviced by PMEM cache instead of by disk are referred to as **Write-Back PMEM cache**. This feature can be enabled or disabled, as needed.



#### Note:

Any time the PMEM cache is dropped and re-created, there is a performance impact for database operations. While the PMEM cache is being repopulated, there are more cache misses, which impacts database performance.

- [Enable Write Back PMEM Cache](#)  
To change the PMEM Cache mode from Write Through to Write Back on an Exadata Database Machine X8M or later system, if the Flash Cache is in Write Through mode, you first need to modify the Flash Cache.
- [Disable Write-Back PMEM Cache](#)  
Use these steps if you need to disable Write-Back PMEM cache on the storage servers.

#### 3.5.3.1 Enable Write Back PMEM Cache

To change the PMEM Cache mode from Write Through to Write Back on an Exadata Database Machine X8M or later system, if the Flash Cache is in Write Through mode, you first need to modify the Flash Cache.



#### Note:

- The best practice recommendation is to configure PMEM Cache in Write Through mode. This configuration provides the best performance and availability.
- If you change the PMEM Cache mode from Write Through to Write Back, then modify the configuration during a period of reduced workload to reduce the performance impact on your applications.

1. If the Flash Cache is currently set to Write Though mode, you must first change the Flash Cache mode to Write Back.

Use the following command to determine the current setting for the Flash Cache, where `cell_group` is a text file that lists the storage servers to query:

```
dcli -l root -g cell_group cellcli -e list cell detail | grep
flashCacheMode
```

- a.** Validate all the physical disks are in `NORMAL` state before modifying the Flash Cache.

```
dcli -l root -g cell_group cellcli -e "LIST PHYSICALDISK ATTRIBUTES
name,status"
|grep -v NORMAL
```

- b.** Validate all the physical disks are in `NORMAL` state before modifying the Flash Cache.

The command should return no rows.

- c.** Determine amount of dirty data in the Flash Cache.

```
cellcli -e "LIST METRICCURRENT ATTRIBUTES name,metricvalue WHERE
name LIKE \'FC_BY_DIRTY.*\' "
```

- d.** Flush the Flash Cache.

If the Flash Cache utilizes all available Flash cell disks, you can use the `ALL` keyword instead of listing the flash disks.

```
dcli -g cell_group -l root cellcli -e "ALTER FLASHCACHE
CELLDISK=\'FD_02_dm01celadm12,
FD_03_dm01celadm12,FD_00_dm01celadm12,FD_01_dm01celadm12\' FLUSH"
```

- e.** Check the progress of the flushing the Flash Cache.

The flushing process is complete when `FC_BY_DIRTY` is 0 MB.

```
dcli -g cell_group -l root cellcli -e "LIST METRICCURRENT
ATTRIBUTES name,metricvalue
WHERE name LIKE \'FC_BY_DIRTY.*\' "
```

Or, you can check to see if the attribute `flushstatus` has been set to `Completed`.

```
dcli -g cell_group -l root cellcli -e "LIST CELLDISK ATTRIBUTES
name, flushstatus,
flusherror" | grep FD
```

- f.** After flushing of the Flash Cache completes, drop the Flash Cache.

```
dcli -g cell_group -l root cellcli -e drop flashcache
```

- g.** Modify the Flash cache to use Write Back mode.

```
dcli -g cell_group -l root cellcli -e "ALTER CELL
flashCacheMode=writeback"
```

- h.** Re-create the Flash Cache.



If the Flash Cache utilizes all available Flash cell disks, you can use the `ALL` keyword instead of listing the cell disks.

If you do not include the `size` attribute, then all available space on each cell disk in the list is used for the Flash Cache.

```
dcli -l root -g cell_group cellcli -e "create flashcache
celldisk='FD_02_dm01celadm12,
FD_03_dm01celadm12,FD_00_dm01celadm12,FD_01_dm01celadm12\'
```

- i. Verify the `flashCacheMode` has been set to `writeback`.

```
dcli -l root -g cell_group cellcli -e list cell detail | grep
flashCacheMode
```

2. Flush the PMEM cache.

If the PMEM cache utilizes all available PMEM cell disks, you can use the `ALL` keyword as shown here.

```
dcli -l root -g cell_group cellcli ALTER PMEMCACHE ALL FLUSH
```

Otherwise, list the specific disks using the `CELDDISK="cdisk1 [,cdisk2] ..."` clause.

3. Drop the PMEM cache.

```
dcli -l root -g cell_group cellcli DROP PMEMCACHE
```

4. Modify the PMEM cache to use Write Back mode.

```
dcli -l root -g cell_group cellcli ALTER CELL
pmemCacheMode=WriteBack
```

Starting with Oracle Exadata System Software release 20.1.0, this command warns about the best practice recommendation to use PMEM Cache in Write Through mode and prompts for confirmation of the change.

5. Re-create the PMEM cache.

If the PMEM cache utilizes all available PMEM cell disks, you can use the `ALL` keyword as shown here. Otherwise, list the specific disks using the `CELDDISK="cdisk1 [,cdisk2] ..."` clause. If the `size` attribute is not specified, then the maximum size is allocated. All available space on each cell disk in the list is used for PMEM Cache.

```
dcli -l root -g cell_group cellcli -e CREATE PMEMCACHE ALL
```

6. Verify the `pmemCacheMode` has been set to `writeback`.

```
dcli -l root -g cell_group cellcli -e list cell detail | grep
pmemCacheMode
```

### 3.5.3.2 Disable Write-Back PMEM Cache

Use these steps if you need to disable Write-Back PMEM cache on the storage servers.

You do not have to stop the cellsrv process or inactivate grid disks when disabling Write-Back PMEM cache. However, to reduce the performance impact on the application, disable the Write-Back PMEM cache during a period of reduced workload.

1. Validate all the Physical Disks are in NORMAL state before modifying PMEM cache.

The following command should return no rows:

```
dcli -l root -g cell_group cellcli -e "LIST PHYSICALDISK ATTRIBUTES
name,status"|grep -v NORMAL
```

2. Flush the PMEM cache.

```
dcli -g cell_group -l root cellcli -e "ALTER PMEMCACHE ALL FLUSH"
```

The PMEM cache flushes the dirty data to the lower layer Write-Back Flash Cache.

3. Check that the flushing operation for the PMEM cache has completed.

The flushing process is complete when the PMEM devices do not show up in the `cachedBy` attribute for the grid disks.

```
CellCLI> LIST GRIDDISK ATTRIBUTES name, cachedBy
DATA_CD_00_cel01 FD_00_cel01
DATA_CD_01_cel01 FD_01_cel01
DATA_CD_02_cel01 FD_03_cel01
DATA_CD_03_cel01 FD_02_cel01
DATA_CD_04_cel01 FD_00_cel01
DATA_CD_05_cel01 FD_02_cel01
...
```

4. Drop the PMEM cache.

```
dcli -g cell_group -l root cellcli -e drop pmemcache all
```

5. Set the `pmemCacheMode` attribute to `writethrough`.

```
dcli -g cell_group -l root cellcli -e "ALTER CELL
pmemCacheMode=writethrough"
```

6. Re-create the PMEM cache.

```
dcli -l root -g cell_group cellcli -e create pmemcache all
```

7. Verify the `pmemCacheMode` has been set to `writethrough`.

```
CellCLI> LIST CELL ATTRIBUTES pmemcachemode
WriteThrough
```

## 3.6 Maintaining the M.2 Disks of Oracle Exadata Storage Server

Exadata Database Machine X7 and later systems come with two internal M.2 devices that contain the system area.

In all previous systems, the first two disks of the Oracle Exadata Storage Server are system disks and the portions on these system disks are referred to as the system area.



### Note:

Oracle Exadata Rack and Oracle Exadata Storage Servers can remain online and available while replacing an M.2 disk.

- [Monitoring the Status of M.2 Disks](#)  
You can monitor the status of a M.2 disk by checking its attributes with the CellCLI `LIST PHYSICALDISK` command.
- [Replacing a M.2 Disk Due to Failure or Other Problems](#)

### 3.6.1 Monitoring the Status of M.2 Disks

You can monitor the status of a M.2 disk by checking its attributes with the CellCLI `LIST PHYSICALDISK` command.

The disk firmware maintains the error counters, and marks a drive with Predictive Failure when the disk is about to fail. The drive, not the cell software, determines if it needs replacement.

- Use the CellCLI command `LIST PHYSICALDISK` to determine the status of a M.2 disk:

```
CellCLI> LIST PHYSICALDISK WHERE disktype='M2Disk' DETAIL
name: M2_SYS_0
deviceName: /dev/sdm
diskType: M2Disk
makeModel: "INTEL SSDSCKJB150G7"
physicalFirmware: N2010112
physicalInsertTime: 2017-07-14T08:42:24-07:00
physicalSerial: PHDW7082000M150A
physicalSize: 139.73558807373047G
slotNumber: "M.2 Slot: 0"
status: failed

name: M2_SYS_1
deviceName: /dev/sdn
diskType: M2Disk
makeModel: "INTEL SSDSCKJB150G7"
physicalFirmware: N2010112
physicalInsertTime: 2017-07-14T12:25:05-07:00
```

```

physicalSerial: PHDW708200SZ150A
physicalSize: 139.73558807373047G
slotNumber: "M.2 Slot: 1"
status: normal

```

The Exadata Storage Server M.2 disk statuses are:

- normal
- not present
- failed
- warning - predictive failure

## 3.6.2 Replacing a M.2 Disk Due to Failure or Other Problems

Failure of a M.2 disks reduces redundancy of the system area, and can impact patching, imaging, and system rescue. Therefore, the disk should be replaced with a new disk as soon as possible. When a M.2 disk fails, the storage server automatically and transparently switches to using the software stored on the inactive system disk, making it the active system disk.

An Exadata alert is generated when an M.2 disk fails. The alert includes specific instructions for replacing the disk. If you have configured the system for alert notifications, then the alert is sent by e-mail to the designated address.

M.2 disk is hot-pluggable and can be replaced when the power is on.

After the M.2 disk is replaced, Oracle Exadata System Software automatically adds the new device to the system partition and starts the rebuilding process.

1. Identify the failed M.2 disk.

```

CellCLI> LIST PHYSICALDISK WHERE diskType=M2Disk AND status!=normal DETAIL
name: M2_SYS_0
deviceName: /dev/sda
diskType: M2Disk
makeModel: "INTEL SSDSCKJB150G7"
physicalFirmware: N2010112
physicalInsertTime: 2017-07-14T08:42:24-07:00
physicalSerial: PHDW7082000M150A
physicalSize: 139.73558807373047G
slotNumber: "M.2 Slot: 0"
status: failed

```

2. Locate the cell that has the white LED lit.
3. Open the chassis and identify the M.2 disk by the slot number in Step 1. The amber LED for this disk should be lit to indicate service is needed.

M.2 disks are hot pluggable, so you do not need to power down the cell before replacing the disk.

4. Remove the M.2 disk:
  - a. Rotate both riser board socket ejectors up and outward as far as they will go.

The green power LED on the riser board turns off when you open the socket ejectors.

- b. Carefully lift the riser board straight up to the remove it from the sockets.
5. Insert the replacement M.2 disk:
  - a. Unpack the replacement flash riser board and place it on an antistatic mat.
  - b. Align the notch in the replacement riser board with the connector key in the connector socket.
  - c. Push the riser board into the connector socket until the riser board is securely seated in the socket.

**▲ Caution:**

If the riser board does not easily seat into the connector socket, verify that the notch in the riser board is aligned with the connector key in the connector socket. If the notch is not aligned, damage to the riser board might occur.

- d. Rotate both riser board socket ejectors inward until the ejector tabs lock the riser board in place.  
 The green power LED on the riser board turns on when you close the socket ejectors.
6. Confirm the M.2 disk has been replaced.

```
CellCLI> LIST PHYSICALDISK WHERE DISKTYPE=M2Disk DETAIL
```

```

name: M2_SYS_0
deviceName: /dev/sdm
diskType: M2Disk
makeModel: "INTEL SSDSCKJB150G7"
physicalFirmware: N2010112
physicalInsertTime: 2017-08-24T18:55:13-07:00
physicalSerial: PHDW708201G0150A
physicalSize: 139.73558807373047G
slotNumber: "M.2 Slot: 0"
status: normal

name: M2_SYS_1
deviceName: /dev/sdn
diskType: M2Disk
makeModel: "INTEL SSDSCKJB150G7"
physicalFirmware: N2010112
physicalInsertTime: 2017-08-24T18:55:13-07:00
physicalSerial: PHDW708200SZ150A
physicalSize: 139.73558807373047G
slotNumber: "M.2 Slot: 1"
status: normal

```

7. Confirm the system disk arrays are have an active sync status, or are being rebuilt.

```

mdadm --detail /dev/md[2-3] [4-5]
/dev/md24:
 Container : /dev/md/ism0, member 0

```

```

 Raid Level : raid1
 Array Size : 104857600 (100.00 GiB 107.37 GB)
 Used Dev Size : 104857600 (100.00 GiB 107.37 GB)
 Raid Devices : 2
 Total Devices : 2

 State : active
 Active Devices : 2
 Working Devices : 2
 Failed Devices : 0
 Spare Devices : 0

 UUID : 152f728a:6d294098:5177b2e5:8e0d766c
 Number Major Minor RaidDevice State
 1 8 16 0 active sync /dev/sdb
 0 8 0 1 active sync /dev/sda
/dev/md25:
 Container : /dev/md/ims0, member 1
 Raid Level : raid1
 Array Size : 41660416 (39.73 GiB 42.66 GB)
 Used Dev Size : 41660544 (39.73 GiB 42.66 GB)
 Raid Devices : 2
 Total Devices : 2

 State : clean
 Active Devices : 2
 Working Devices : 2
 Failed Devices : 0
 Spare Devices : 0

 UUID : 466173ba:507008c7:6d65ed89:3c40cf23
 Number Major Minor RaidDevice State
 1 8 16 0 active sync /dev/sdb
 0 8 0 1 active sync /dev/sda

```

## 3.7 Managing the RAM Cache on the Storage Servers

Cell RAM Cache is a cache in front of the Flash Cache and is an extension of the database cache. It is faster than the Flash Cache, but has smaller capacity.

The Cell RAM Cache feature was introduced in Oracle Exadata System Software release 18c (18.1.0.0.0). Cell RAM Cache is disabled by default (`ramCacheMode` is set to `auto`).



### Note:

Commencing with the September 2022 Oracle Exadata System Software release updates (versions 22.1.3, 21.2.16, and later), the storage server RAM Cache feature is deprecated.

- [About the Cell RAM Cache](#)  
The Cell RAM Cache provides much lower IO latency for online transaction processing (OLTP) reads.
- [Sizing Recommendations for the Cell RAM Cache](#)  
Use the buffer pool advisory section in the Automatic Workload Repository (AWR) reports during their peak OLTP workloads to determine the recommended size of the Cell RAM Cache.
- [Enabling the Cell RAM Cache](#)  
To enable the Cell RAM Cache feature, on each Oracle Exadata Storage Server set the `ramCacheMode` cell attribute to `on`.
- [Viewing the Cell RAM Cache Size](#)  
After the `ramCacheMode` attribute is set to `on`, the storage server automatically uses as much free memory as is available on the storage server to create the Cell RAM Cache.
- [Changing the Size of the Cell RAM Cache](#)  
Cell RAM Cache automatically uses as much free memory as is available on the storage server to create the Cell RAM Cache.
- [Monitoring Cell RAM Cache Usage Statistics](#)
- [Disabling the Cell RAM Cache](#)  
To disable the Cell RAM Cache feature, set the `ramCacheMode` cell attribute to `off` on each Oracle Exadata Storage Server.

### 3.7.1 About the Cell RAM Cache

The Cell RAM Cache provides much lower IO latency for online transaction processing (OLTP) reads.

In an OLTP workload, the `cell single block physical read wait` statistic typically shows up as a top consumer in database processing time. If these reads could be procured from the Cell RAM Cache, you can greatly reduce the read latency and reduce the IO wait time for these reads, improving the performance of OLTP applications.

Alternatively, you can view the Cell RAM Cache as an extension of the database buffer cache. The buffer cache misses become Cell RAM Cache hits, which accelerates OLTP performance because you are getting more cache hits from the combined power of the buffer cache and the Cell RAM Cache.

### 3.7.2 Sizing Recommendations for the Cell RAM Cache

Use the buffer pool advisory section in the Automatic Workload Repository (AWR) reports during their peak OLTP workloads to determine the recommended size of the Cell RAM Cache.

The default size for the Cell RAM Cache on Exadata Storage Servers without the memory expansion kit is rather limited, so this feature is not enabled by default. To get the acceleration benefits, you should first install memory expansion kits on the storage servers. Then, you can enable the Cell RAM Cache and the Oracle Exadata System Software automatically sizes the Cell RAM Cache based on the free memory on Exadata Storage Servers.

The buffer pool advisory section in the AWR reports during peak OLTP workloads can help determine how you should configure the Cell RAM Cache. For example:

**Buffer Pool Advisory**

- Only rows with estimated physical reads >0 are displayed
- ordered by Block Size, Buffers For Estimate

P	Size for Est (M)	Size Factor	Buffers (thousands)	Est Phys Read Factor	Estimated Phys Reads (thousands)	Est Phys Read Time	Est %DBTime for Rds
D	4,992	0.10	584	3.30	1,115,122	1	116408.00
D	9,984	0.20	1,167	2.84	960,889	1	100280.00
D	14,976	0.30	1,751	2.45	830,534	1	86649.00
D	19,968	0.39	2,334	2.13	719,317	1	75020.00
D	24,960	0.49	2,918	1.85	626,155	1	65278.00
D	29,952	0.59	3,501	1.62	549,365	1	57248.00
D	34,944	0.69	4,085	1.44	486,275	1	50651.00
D	39,936	0.79	4,668	1.28	433,335	1	45115.00
D	44,928	0.89	5,252	1.14	387,107	1	40281.00
D	49,920	0.98	5,835	1.02	344,729	1	35850.00
D	50,688	1.00	5,925	1.00	338,431	1	35191.00
D	54,912	1.08	6,419	0.90	304,551	1	31648.00
D	59,904	1.18	7,002	0.78	265,520	1	27567.00
D	64,896	1.28	7,586	0.67	227,121	1	23552.00
D	69,888	1.38	8,169	0.56	189,077	1	19573.00
D	74,880	1.48	8,753	0.45	151,446	1	15638.00
D	79,872	1.58	9,336	0.34	114,191	1	11743.00
D	84,864	1.67	9,920	0.23	77,520	1	7908.00
D	89,856	1.77	10,503	0.12	41,621	1	4154.00
D	94,848	1.87	11,087	0.04	11,971	1	1054.00
D	99,840	1.97	11,670	0.03	11,059	1	958.00

In the above report, with the current buffer cache size (size factor of 1.00), the database performs approximately 338 million physical reads. If you increased the buffer cache by 87% (size factor of 1.87), you would reduce the physical reads to around 12 million.

If you created a Cell RAM Cache that is 87% of the buffer cache size, you could have 338,431,000 reads - 11,971,000 reads, or approximately 326 million reads being satisfied by the Cell RAM Cache. You can size the Cell RAM Cache based on the number of OLTP reads that you would like to benefit from the Cell RAM Cache. The total size of the Cell RAM Cache should be divided by the number of available storage servers to get the target RAM size to use for each storage server.

If you have an Oracle Real Application Clusters (Oracle RAC) database with multiple database instances, then each database instance has an independent buffer pool advisory in the respective AWR report. The physical read savings may vary from instance to instance. The report has already taken into account the case where a buffer cache miss can be satisfied by a cache fusion block transfer from another instance. So the buffer pool advisory report only estimates the actual storage reads after discounting the cache fusion block transfers. To size the Cell RAM Cache for Oracle RAC is very simple: decide how much additional space you need for each instance and add up the values. The sum is the amount of Cell RAM Cache that you need to provision.

Similarly, if you have multiple databases sharing the same set of cells, you can add up the additional buffer cache sizes for each database. The sum is the total amount of Cell RAM Cache you need to provision on the cells.

After you know what size Cell RAM Cache you want, you can then make a decision on which memory expansion kit suits your need best. You can add extra memory for both the database server and the storage server, and you can upgrade the servers in any order.

- Expanding database server memory gives you more memory to use on a given server, and potentially have a bigger buffer cache in addition to larger SGA, PGA, and so on.



- If you expand storage server memory, then the Cell RAM Cache is shared across all database servers. The additional memory can be leveraged by all database servers across all Oracle Virtual Machines (Oracle VM) and database instances.

If you run different workloads at different times across different database servers or Oracle VMs, or if you are running a consolidated environment with multiple databases, then expanding the storage server memory might provide more flexibility for maximizing the usage of the additional memory because that memory can be shared across all databases.

### 3.7.3 Enabling the Cell RAM Cache

To enable the Cell RAM Cache feature, on each Oracle Exadata Storage Server set the `ramCacheMode` cell attribute to `on`.

The Cell RAM Cache feature is disabled by default (`ramCacheMode` is set to `auto`). On each storage server, when you set `ramCacheMode=on`, the storage server automatically uses all available free memory to create the Cell RAM Cache.

1. Change the value of the `ramCacheMode` attribute on each cell.

You can use `dcli` to modify multiple cells with a single command, or you can run the following command on each Oracle Exadata Storage Server.

```
CellCLI> ALTER CELL ramCacheMode=on
Cell host03celadm10 successfully altered
```

2. Restart CellSrv.

```
CellCLI> ALTER CELL RESTART SERVICES CELLSRV
```

### 3.7.4 Viewing the Cell RAM Cache Size

After the `ramCacheMode` attribute is set to `on`, the storage server automatically uses as much free memory as is available on the storage server to create the Cell RAM Cache.

The creation of RAM cache takes place asynchronously in the background. If you query the size of the Cell RAM Cache immediately after enabling this feature, you will not see an accurate size. You can monitor the cell `alert.log` to follow the progress of Cell RAM Cache creation.

- To view the current status of the Cell RAM Cache, retrieve the `ramCacheMode`, `ramCacheSize` and `ramCacheMaxSize` attributes for the cell.

```
CellCLI> LIST CELL ATTRIBUTES ramCacheMaxSize,ramCacheMode,
ramCacheSize
 18.875G On 18.875G
```

## 3.7.5 Changing the Size of the Cell RAM Cache

Cell RAM Cache automatically uses as much free memory as is available on the storage server to create the Cell RAM Cache.

After you enable the Cell RAM Cache feature, the storage server automatically uses as much free memory as is available on the Oracle Exadata Storage Server to create the Cell RAM Cache. The size of the Cell RAM Cache can be determined by using the CellCLI `LIST CELL ramCacheSize` command.

The `ramCacheMaxSize` attribute determines the maximum amount of memory that can be used for the Cell RAM Cache.

To limit the size of the Cell RAM Cache, modify the value of the `ramCacheMaxSize` attribute on each storage server.

You can use `exadcli` to modify multiple storage servers with a single command, or you can run the following command on each Oracle Exadata Storage Server.

```
CellCLI> ALTER CELL ramCacheMaxSize=1G
Cell host03celadm10 successfully altered
```

### Example 3-1 Limit the Size of the Cell RAM Cache

To limit the maximum size of the Cell RAM Cache to 1 GB, use the following command on to modify multiple Oracle Exadata Storage Servers:

```
exadcli -c cell01,cell02,cell03 -l celladministrator alter cell
ramCacheMaxSize=1G
```

### Related Topics

- [Sizing Recommendations for the Cell RAM Cache](#)

## 3.7.6 Monitoring Cell RAM Cache Usage Statistics

Use the following methods to monitor Cell RAM Cache usage:

- ExaWatcher reports  
RamCache size and read, write, and population operation related stats are exposed via `cellsrvstat` in the “RamCache related stats” section.
- AWR report, in the Memory Cache section.

There are 3 subsections:

- Memory Cache Space Usage: For each cell, this table shows the space allocated for the Cell RAM Cache and the percentage of space used in the Cell RAM Cache for OLTP activity.
- Memory Cache User Reads: For each cell, this table shows the read statistics for the Cell RAM Cache based on the number of reads and the amount of data read (MB).
- Memory Cache Internal Writes: For each cell, this table shows the write statistics for the Cell RAM Cache based on the number of write requests and the amount of data written (MB).

## 3.7.7 Disabling the Cell RAM Cache

To disable the Cell RAM Cache feature, set the `ramCacheMode` cell attribute to `off` on each Oracle Exadata Storage Server.



### Note:

The Cell RAM Cache feature is disabled by default (`ramCacheMode` is set to `auto`).

1. Change the value of the `ramCacheMode` attribute on each cell.

You can use `dcli` to modify multiple cells with a single command, or you can run the following command on each Oracle Exadata Storage Server.

```
CellCLI> ALTER CELL ramCacheMode=off
Cell host03celadm10 successfully altered
```

2. Restart CellSrv.

```
CellCLI> ALTER CELL RESTART SERVICES CELLSRV
```

## 3.8 Resizing Grid Disks

You can resize grid disks and Oracle ASM disk groups to shrink one with excess free space and increase the size of another that is near capacity.

Initial configuration of Exadata Database Machine disk group sizes is based on Oracle best practices and the location of the backup files.

- For internal backups: allocation of available space is 40% for the DATA disk groups, and 60% for the RECO disk groups.
- For external backups: allocation of available space is 80% for the DATA disk group, and 20% for the RECO disk group.

The disk group allocations can be changed after deployment. For example, the DATA disk group allocation may be too small at 60%, and need to be resized to 80%.

If your system has no free space available on the cell disks and one disk group, for example RECO, has plenty of free space, then you can resize the RECO disk group to a smaller size and reallocate the free space to the DATA disk group. The free space available after shrinking the RECO disk group is at a non-contiguous offset from the existing space allocations for the DATA disk group. Grid disks can use space anywhere on the cell disks and do not have to be contiguous.

If you are expanding the grid disks and the cell disks already have sufficient space to expand the existing grid disks, then you do not need to first resize an existing disk group. You would skip steps 2 and 3 below where the example shows the RECO disk group and grid disks are shrunk (you should still verify the cell disks have enough free space before growing the DATA grid disks). The amount of free space the administrator should reserve depends on the level of failure coverage.

If you are shrinking the size of the grid disks, you should understand how space is reserved for mirroring. Data is protected by Oracle ASM using normal or high redundancy to create one or two copies of data, which are stored as file extents. These copies are stored in separate failure groups. A failure in one failure group does not affect the mirror copies, so data is still accessible.

When a failure occurs, Oracle ASM re-mirrors, or **rebalances**, any extents that are not accessible so that redundancy is reestablished. For the re-mirroring process to succeed, sufficient free space must exist in the disk group to allow creation of the new file extent mirror copies. If there is not enough free space, then some extents will not be re-mirrored and the subsequent failure of the other data copies will require the disk group to be restored from backup. Oracle ASM sends an error when a re-mirror process fails due to lack of space.

You must be using Oracle Exadata System Software release 12.1.2.1.0 or higher, or have the patch for bug 19695225 applied to your software.

This procedure for resizing grid disks applies to bare metal and virtual machine (VM) deployments.

- [Determine the Amount of Available Space](#)  
To increase the size of the disks in a disk group you must either have unallocated disk space available, or you have to reallocate space currently used by a different disk group.
- [Shrink the Oracle ASM Disks in the Donor Disk Group](#)  
If there is no free space available on the cell disks, you can reduce the space used by one disk group to provide additional disk space for a different disk group.
- [Shrink the Grid Disks in the Donor Disk Group](#)  
After shrinking the disks in the Oracle ASM disk group, you then shrink the size of the grid disks on each cell.
- [Increase the Size of the Grid Disks Using Available Space](#)  
You can increase the size used by the grid disks if there is unallocated disk space either already available, or made available by shrinking the space used by a different Oracle ASM disk group.
- [Increase the Size of the Oracle ASM Disks](#)  
You can increase the size used by the Oracle ASM disks after increasing the space allocated to the associated grid disks.

#### Related Topics

- [Understanding ASM Capacity and Reservation of Free Space in Exadata \(My Oracle Support Doc ID 1551288.1\)](#)
- [Bug 19695225 - Running Many Create or Alter Griddisk Commands Over Time Causes Cell Disk Metadata Corruption \(ORA-600 \[addNewSegmentsToGDisk\\_2\]\) and Loss of Cell Disk Content \(My Oracle Support Doc ID 1991445.1\)](#)

## 3.8.1 Determine the Amount of Available Space

To increase the size of the disks in a disk group you must either have unallocated disk space available, or you have to reallocate space currently used by a different disk group.

You can also use a script available in "[Script to Calculate New Grid Disk and Disk Group Sizes in Exadata \(My Oracle Support Doc ID 1464809.1\)](#)" to assist in determining how much free space is available to shrink a disk group.

## 1. View the space currently used by the disk groups.

```
SELECT name, total_mb, free_mb, total_mb - free_mb used_mb,
 round(100*free_mb/total_mb,2) pct_free
FROM v$asm_diskgroup
ORDER BY 1;
```

NAME	TOTAL_MB	FREE_MB	USED_MB	PCT_FREE
DATA1	68812800	9985076		14.51
RECO1	94980480	82594920		86.96

The example above shows that the DATA1 disk group has only about 15% of free space available while the RECO1 disk group has about 87% free disk space. The PCT\_FREE displayed here is raw free space, not usable free space. Additional space is needed for rebalancing operations.

## 2. For the disk groups you plan to resize, view the count and status of the failure groups used by the disk groups.

```
SELECT dg.name, d.failgroup, d.state, d.header_status,
 d.mount_mode,
 d.mode_status, count(1) num_disks
FROM V$ASM_DISK d, V$ASM_DISKGROUP dg
WHERE d.group_number = dg.group_number
AND dg.name IN ('RECO1', 'DATA1')
GROUP BY dg.name, d.failgroup, d.state, d.header_status,
 d.mount_status,
 d.mode_status
ORDER BY 1, 2, 3;
```

NAME	FAILGROUP	STATE	HEADER_STATU	MOUNT_S	MODE_ST	NUM_DISKS
DATA1	EXA01CELADM01	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM02	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM03	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM04	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM05	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM06	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM07	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM08	NORMAL	MEMBER	CACHED	ONLINE	12

DATA1	EXA01CELADM09	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM10	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM11	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM12	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM13	NORMAL	MEMBER	CACHED	ONLINE	12
DATA1	EXA01CELADM14	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM01	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM02	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM03	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM04	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM05	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM06	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM07	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM08	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM09	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM10	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM11	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM12	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM13	NORMAL	MEMBER	CACHED	ONLINE	12
RECOC1	EXA01CELADM14	NORMAL	MEMBER	CACHED	ONLINE	12

The above example is for a full rack, which has 14 cells and 14 failure groups for DATA1 and RECOC1. Verify that each failure group has at least 12 disks in the NORMAL state (`num_disks`). If you see disks listed as MISSING, or you see an unexpected number of disks for your configuration, then do not proceed until you resolve the problem.

Extreme Flash systems should see a disk count of 8 instead of 12 for `num_disks`.

- List the corresponding grid disks associated with each cell and each failure group, so you know which grid disks to resize.

```
SELECT dg.name, d.failgroup, d.path
FROM V$ASM_DISK d, V$ASM_DISKGROUP dg
WHERE d.group_number = dg.group_number
AND dg.name IN ('RECOC1', 'DATA1')
ORDER BY 1, 2, 3;
```

NAME	FAILGROUP	PATH
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_00_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_01_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_02_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_03_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_04_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_05_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_06_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_07_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_08_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_09_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_10_exa01celadm01
DATA1	EXA01CELADM01	o/192.168.74.43/DATA1_CD_11_exa01celadm01
DATA1	EXA01CELADM02	o/192.168.74.44/DATA1_CD_00_exa01celadm01
DATA1	EXA01CELADM02	o/192.168.74.44/DATA1_CD_01_exa01celadm01
DATA1	EXA01CELADM02	o/192.168.74.44/DATA1_CD_02_exa01celadm01
...		

```

RECOC1 EXA01CELADM13 o/192.168.74.55/
RECOC1_CD_00_exa01celadm13
RECOC1 EXA01CELADM13 o/192.168.74.55/
RECOC1_CD_01_exa01celadm13
RECOC1 EXA01CELADM13 o/192.168.74.55/
RECOC1_CD_02_exa01celadm13
...
RECOC1 EXA01CELADM14 o/192.168.74.56/
RECOC1_CD_09_exa01celadm14
RECOC1 EXA01CELADM14 o/192.168.74.56/
RECOC1_CD_10_exa01celadm14
RECOC1 EXA01CELADM14 o/192.168.74.56/
RECOC1_CD_11_exa01celadm14

```

168 rows returned.

#### 4. Check the cell disks for available free space.

Free space on the cell disks can be used to increase the size of the DATA1 grid disks. If there is not enough available free space to expand the DATA1 grid disks, then you must shrink the RECOC1 grid disks to provide the additional space for the desired new size of DATA1 grid disks.

```

[root@exa01adm01 tmp]# dcli -g ~/cell_group -l root "cellcli -e
list celldisk \
 attributes name,freespace"
exa01celadm01: CD_00_exa01celadm01 0
exa01celadm01: CD_01_exa01celadm01 0
exa01celadm01: CD_02_exa01celadm01 0
exa01celadm01: CD_03_exa01celadm01 0
exa01celadm01: CD_04_exa01celadm01 0
exa01celadm01: CD_05_exa01celadm01 0
exa01celadm01: CD_06_exa01celadm01 0
exa01celadm01: CD_07_exa01celadm01 0
exa01celadm01: CD_08_exa01celadm01 0
exa01celadm01: CD_09_exa01celadm01 0
exa01celadm01: CD_10_exa01celadm01 0
exa01celadm01: CD_11_exa01celadm01 0
...

```

In this example, there is no free space available, so you must shrink the RECOC1 grid disks first to provide space for the DATA1 grid disks. In your configuration there might be plenty of free space available and you can use that free space instead of shrinking the RECOC1 grid disks.

#### 5. Calculate the amount of space to shrink from the RECOC1 disk group and from each grid disk.

The minimum size to safely shrink a disk group and its grid disks must take into account the following:

- Space currently in use (`USED_MB`)
- Space expected for growth (`GROWTH_MB`)
- Space needed to rebalance in case of disk failure (`DFC_MB`), typically 15% of total disk group size

The minimum size calculation taking the above factors into account is:

$$\text{Minimum DG size (MB)} = (\text{USED\_MB} + \text{GROWTH\_MB}) * 1.15$$

- `USED_MB` can be derived from `V$ASM_DISKGROUP` by calculating `TOTAL_MB - FREE_MB`
- `GROWTH_MB` is an estimate specific to how the disk group will be used in the future and should be based on historical patterns of growth

For the RECOC1 disk group space usage shown in step 1, we see the minimum size it can shrink to assuming no growth estimates is:

$$\begin{aligned} \text{Minimum RECOC1 size} &= (\text{TOTAL\_MB} - \text{FREE\_MB} + \text{GROWTH\_MB}) * 1.15 \\ &= (94980480 - 82594920 + 0) * 1.15 = 14243394 \text{ MB} = 13,910 \text{ GB} \end{aligned}$$

In the example output shown in Step 1, RECOC1 has plenty of free space and DATA1 has less than 15% free. So, you could shrink RECOC1 and give the freed disk space to DATA1. If you decide to reduce RECOC1 to half of its current size, the new size is  $94980480 / 2 = 47490240$  MB. This size is significantly above the minimum size we calculated for the RECOC1 disk group above, so it is safe to shrink it down to this value.

The query in Step 2 shows that there are 168 grid disks for RECOC1, because there are 14 cells and 12 disks per cell ( $14 * 12 = 168$ ). The estimated new size of each grid disk for the RECOC1 disk group is  $47490240 / 168$ , or 282,680 MB.

Find the closest 16 MB boundary for the new grid disk size. If you do not perform this check, then the cell will round down the grid disk size to the nearest 16 MB boundary automatically, and you could end up with a mismatch in size between the Oracle ASM disks and the grid disks.

```
SQL> SELECT 16*TRUNC(&new_disk_size/16) new_disk_size FROM dual;
Enter value for new_disk_size: 282680
```

```
NEW_DISK_SIZE

 282672
```

Based on the above result, you should choose 282672 MB as the new size for the grid disks in the RECOC1 disk group. After resizing the grid disks, the size of the RECOC1 disk group will be 47488896 MB.

6. Calculate how much to increase the size of each grid disk in the DATA1 disk group.

Ensure the Oracle ASM disk size and the grid disk sizes match across the entire disk group. The following query shows the combinations of disk sizes in each disk group. Ideally, there is only one size found for all disks and the sizes of both the Oracle ASM (`total_mb`) disks and the grid disks (`os_mb`) match.

```
SELECT dg.name, d.total_mb, d.os_mb, count(1) num_disks
FROM v$asm_diskgroup dg, v$asm_disk d
WHERE dg.group_number = d.group_number
GROUP BY dg.name, d.total_mb, d.os_mb;
```

```
NAME TOTAL_MB OS_MB NUM_DISKS

```



DATA1	409600	409600	168
RECO1	565360	565360	168

After shrinking RECO1's grid disks, the following space is left per disk for DATA1:

Additional space for DATA1 disks =  $RECO1\_current\_size - RECO1\_new\_size$   
=  $565360 - 282672 = 282688$  MB

To calculate the new size of the grid disks for the DATA1 disk group, use the following:

DATA1's disks new size =  $DATA1\_disks\_current\_size +$   
 $new\_free\_space\_from\_RECO1$   
=  $409600 + 282688 = 692288$  MB

Find the closest 16 MB boundary for the new grid disk size. If you do not perform this check, then the cell will round down the grid disk size to the nearest 16 MB boundary automatically, and you could end up with a mismatch in size between the Oracle ASM disks and the grid disks.

```
SQL> SELECT 16*TRUNC(&new_disk_size/16) new_disk_size FROM dual;
Enter value for new_disk_size: 692288
```

```
NEW_DISK_SIZE

 692288
```

Based on the query result, you can use the calculated size of 692288 MB for the disks in the DATA1 disk groups because the size is on a 16 MB boundary. If the result of the query is different from the value you supplied, then you must use the value returned by the query because that is the value to which the cell will round the grid disk size.

The calculated value of the new grid disk size will result in the DATA1 disk group having a total size of 116304384 MB (168 disks \* 692288 MB).

## 3.8.2 Shrink the Oracle ASM Disks in the Donor Disk Group

If there is no free space available on the cell disks, you can reduce the space used by one disk group to provide additional disk space for a different disk group.

This task is a continuation of an example where space in the RECO1 disk group is being reallocated to the DATA1 disk group.

Before resizing the disk group, make sure the disk group you are taking space from has sufficient free space.

1. Shrink the Oracle ASM disks for the RECO disk group down to the new desired size for all disks.

Use the new size for the disks in the RECO disk group that was calculated in Step 5 of [Determine the Amount of Available Space](#).

```
SQL> ALTER DISKGROUP recoc1 RESIZE ALL SIZE 282672M REBALANCE POWER
64;
```

 **Note:**

The `ALTER DISKGROUP` command may take several minutes to complete. The SQL prompt will not return until this operation has completed.

If the specified disk group has quorum disks configured within the disk group, then the `ALTER DISKGROUP ... RESIZE ALL` command could fail with error `ORA-15277`. Quorum disks are configured if the requirements specified in *Managing Quorum Disks for High Redundancy Disk Groups* are met. As a workaround, you can specify the storage server failure group names (for the ones with a `FAILURE_TYPE` of `REGULAR`, not `QUORUM`) explicitly in the SQL command, for example:

```
SQL> ALTER DISKGROUP recoc1 RESIZE DISKS IN FAILGROUP exacell01
SIZE 282672M,
exacell02 SIZE 282672M, exacell03 SIZE 282672M REBALANCE POWER 64;
```

Wait for rebalance to finish by checking the view `GV$ASM_OPERATION`.

```
SQL> set lines 250 pages 1000
SQL> col error_code form a10
SQL> SELECT dg.name, o.*
2 FROM gv$asm_operation o, v$asm_diskgroup dg
3 WHERE o.group_number = dg.group_number;
```

Proceed to the next step **ONLY** when the query against `GV$ASM_OPERATION` shows no rows for the disk group being altered.

2. Verify the new size of the ASM disks using the following queries:

```
SQL> SELECT name, total_mb, free_mb, total_mb - free_mb used_mb,
2 ROUND(100*free_mb/total_mb,2) pct_free
3 FROM v$asm_diskgroup
4 ORDER BY 1;
```

NAME	TOTAL_MB	FREE_MB	USED_MB	PCT_FREE
DATA1	68812800	9985076	58827724	14.51
RECO1	47488896	35103336	12385560	73.92

```
SQL> SELECT dg.name, d.total_mb, d.os_mb, COUNT(1) num_disks
2 FROM v$asm_diskgroup dg, v$asm_disk d
3 WHERE dg.group_number = d.group_number
4 GROUP BY dg.name, d.total_mb, d.os_mb;
```

NAME	TOTAL_MB	OS_MB	NUM_DISKS
DATA1	409600	409600	168
RECO1	282672	565360	168

The above query example shows that the disks in the RECOC1 disk group have been resized to a size of 282672 MG each, and the total disk group size is 47488896 MB.

### 3.8.3 Shrink the Grid Disks in the Donor Disk Group

After shrinking the disks in the Oracle ASM disk group, you then shrink the size of the grid disks on each cell.

This task is a continuation of an example where space in the RECOC1 disk group is being reallocated to the DATA1 disk group.

You must have first completed the task [Shrink the Oracle ASM Disks in the Donor Disk Group](#).

1. Shrink the grid disks associated with the RECO disk group on all cells down to the new, smaller size.

For each storage cell identified in [Determine the Amount of Available Space in Step 3](#), shrink the grid disks to match the size of the Oracle ASM disks that were shrunk in the previous task. Use commands similar to the following:

```
dcli -c exa01celadm01 -l root "cellcli -e alter griddisk
RECOC1_CD_00_exa01celadm01 \
,RECOC1_CD_01_exa01celadm01 \
,RECOC1_CD_02_exa01celadm01 \
,RECOC1_CD_03_exa01celadm01 \
,RECOC1_CD_04_exa01celadm01 \
,RECOC1_CD_05_exa01celadm01 \
,RECOC1_CD_06_exa01celadm01 \
,RECOC1_CD_07_exa01celadm01 \
,RECOC1_CD_08_exa01celadm01 \
,RECOC1_CD_09_exa01celadm01 \
,RECOC1_CD_10_exa01celadm01 \
,RECOC1_CD_11_exa01celadm01 \
size=282672M "
```

```
dcli -c exa01celadm02 -l root "cellcli -e alter griddisk
RECOC1_CD_00_exa01celadm02 \
,RECOC1_CD_01_exa01celadm02 \
,RECOC1_CD_02_exa01celadm02 \
,RECOC1_CD_03_exa01celadm02 \
,RECOC1_CD_04_exa01celadm02 \
,RECOC1_CD_05_exa01celadm02 \
,RECOC1_CD_06_exa01celadm02 \
,RECOC1_CD_07_exa01celadm02 \
,RECOC1_CD_08_exa01celadm02 \
,RECOC1_CD_09_exa01celadm02 \
,RECOC1_CD_10_exa01celadm02 \
,RECOC1_CD_11_exa01celadm02 \
size=282672M "
```

...

```
dcli -c exa01celadm14 -l root "cellcli -e alter griddisk
RECOC1_CD_00_exa01celadm14 \
```

```
,RECOC1_CD_01_exa01celadm14 \
,RECOC1_CD_02_exa01celadm14 \
,RECOC1_CD_03_exa01celadm14 \
,RECOC1_CD_04_exa01celadm14 \
,RECOC1_CD_05_exa01celadm14 \
,RECOC1_CD_06_exa01celadm14 \
,RECOC1_CD_07_exa01celadm14 \
,RECOC1_CD_08_exa01celadm14 \
,RECOC1_CD_09_exa01celadm14 \
,RECOC1_CD_10_exa01celadm14 \
,RECOC1_CD_11_exa01celadm14 \
size=282672M "
```

**2. Verify the new size of the grid disks using the following command:**

```
[root@exa01adm01 tmp]# dcli -g cell_group -l root "cellcli -e list griddisk
attributes name,size where name like \'RECOC1.*\' "
```

```
exa01celadm01: RECOC1_CD_00_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_01_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_02_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_03_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_04_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_05_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_06_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_07_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_08_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_09_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_10_exa01celadm01 276.046875G
exa01celadm01: RECOC1_CD_11_exa01celadm01 276.046875G
...
```

The above example shows that the disks in the RECOC1 disk group have been resized to a size of 282672 MB each (276.046875 \* 1024).

## 3.8.4 Increase the Size of the Grid Disks Using Available Space

You can increase the size used by the grid disks if there is unallocated disk space either already available, or made available by shrinking the space used by a different Oracle ASM disk group.

This task is a continuation of an example where space in the RECOC1 disk group is being reallocated to the DATAC1 disk group. If you already have sufficient space to expand an existing disk group, then you do not need to reallocate space from a different disk group.

**1. Check that the cell disks have the expected amount of free space.**

After completing the tasks to shrink the Oracle ASM disks and the grid disks, you would expect to see the following free space on the cell disks:

```
[root@exa01adm01 tmp]# dcli -g ~/cell_group -l root "cellcli -e list
celldisk \
attributes name,freespace"

exa01celadm01: CD_00_exa01celadm01 276.0625G
```

```

exa01celadm01: CD_01_exa01celadm01 276.0625G
exa01celadm01: CD_02_exa01celadm01 276.0625G
exa01celadm01: CD_03_exa01celadm01 276.0625G
exa01celadm01: CD_04_exa01celadm01 276.0625G
exa01celadm01: CD_05_exa01celadm01 276.0625G
exa01celadm01: CD_06_exa01celadm01 276.0625G
exa01celadm01: CD_07_exa01celadm01 276.0625G
exa01celadm01: CD_08_exa01celadm01 276.0625G
exa01celadm01: CD_09_exa01celadm01 276.0625G
exa01celadm01: CD_10_exa01celadm01 276.0625G
exa01celadm01: CD_11_exa01celadm01 276.0625G
...

```

- For each storage cell, increase the size of the DATA grid disks to the desired new size.

Use the size calculated in [Determine the Amount of Available Space](#).

```

dcli -c exa01celadm01 -l root "cellcli -e alter griddisk
DATAAC1_CD_00_exa01celadm01 \
,DATAAC1_CD_01_exa01celadm01 \
,DATAAC1_CD_02_exa01celadm01 \
,DATAAC1_CD_03_exa01celadm01 \
,DATAAC1_CD_04_exa01celadm01 \
,DATAAC1_CD_05_exa01celadm01 \
,DATAAC1_CD_06_exa01celadm01 \
,DATAAC1_CD_07_exa01celadm01 \
,DATAAC1_CD_08_exa01celadm01 \
,DATAAC1_CD_09_exa01celadm01 \
,DATAAC1_CD_10_exa01celadm01 \
,DATAAC1_CD_11_exa01celadm01 \
size=692288M "
...

```

```

dcli -c exa01celadm14 -l root "cellcli -e alter griddisk
DATAAC1_CD_00_exa01celadm14 \
,DATAAC1_CD_01_exa01celadm14 \
,DATAAC1_CD_02_exa01celadm14 \
,DATAAC1_CD_03_exa01celadm14 \
,DATAAC1_CD_04_exa01celadm14 \
,DATAAC1_CD_05_exa01celadm14 \
,DATAAC1_CD_06_exa01celadm14 \
,DATAAC1_CD_07_exa01celadm14 \
,DATAAC1_CD_08_exa01celadm14 \
,DATAAC1_CD_09_exa01celadm14 \
,DATAAC1_CD_10_exa01celadm14 \
,DATAAC1_CD_11_exa01celadm14 \
size=692288M "

```

- Verify the new size of the grid disks associated with the DATAAC1 disk group using the following command:

```

dcli -g cell_group -l root "cellcli -e list griddisk attributes
name,size \
where name like \'DATAAC1.*\' "

```

```
exa01celadm01: DATA1_CD_00_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_01_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_02_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_03_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_04_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_05_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_06_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_07_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_08_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_09_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_10_exa01celadm01 676.0625G
exa01celadm01: DATA1_CD_11_exa01celadm01 676.0625G
```

Instead of increasing the size of the DATA disk group, you could instead create new disk groups with the new free space or keep it free for future use. In general, Oracle recommends using the smallest number of disk groups needed (typically DATA, RECO, and DBFS\_DG) to give the greatest flexibility and ease of administration. However, there may be cases, perhaps when using virtual machines or consolidating many databases, where additional disk groups or available free space for future use may be desired.

If you decide to leave free space on the grid disks in reserve for future use, please see the My Oracle Support Note 1684112.1 for the steps on how to allocate free space to an existing disk group at a later time.

#### Related Topics

- [How to resize ASM disks in Exadata \(My Oracle Support Doc ID 1684112.1\)](#)
- [Determine the Amount of Available Space](#)  
To increase the size of the disks in a disk group you must either have unallocated disk space available, or you have to reallocate space currently used by a different disk group.

## 3.8.5 Increase the Size of the Oracle ASM Disks

You can increase the size used by the Oracle ASM disks after increasing the space allocated to the associated grid disks.

This task is a continuation of an example where space in the RECO1 disk group is being reallocated to the DATA1 disk group.

You must have completed the task of resizing the grid disks before you can resize the corresponding Oracle ASM disk group.

1. Increase the Oracle ASM disks for DATA1 disk group to the new size of the grid disks on the storage cells.

```
SQL> ALTER DISKGROUP data1 RESIZE ALL;
```

This command resizes the Oracle ASM disks to match the size of the grid disks.

 **Note:**

If the specified disk group has quorum disks configured within the disk group, then the `ALTER DISKGROUP ... RESIZE ALL` command could fail with error `ORA-15277`. Quorum disks are configured if the requirements specified in *Oracle Exadata Database Machine Maintenance Guide* are met.

As a workaround, you can specify the storage server failure group names (for the ones of `FAILURE_TYPE "REGULAR"`, not `"QUORUM"`) explicitly in the SQL command, for example:

```
SQL> ALTER DISKGROUP datac1 RESIZE DISKS IN FAILGROUP
exace1101, exace1102, exace1103;
```

2. Wait for the rebalance operation to finish.

```
SQL> set lines 250 pages 1000
SQL> col error_code form a10
SQL> SELECT dg.name, o.* FROM gv$asm_operation o, v$asm_diskgroup
dg
WHERE o.group_number = dg.group_number;
```

Do not continue to the next step until the query returns zero rows for the disk group that was altered.

3. Verify that the new sizes for the Oracle ASM disks and disk group is at the desired sizes.

```
SQL> SELECT name, total_mb, free_mb, total_mb - free_mb used_mb,
ROUND(100*free_mb/total_mb,2) pct_free
FROM v$asm_diskgroup
ORDER BY 1;
```

NAME		TOTAL_MB	FREE_MB	USED_MB
PCT_FREE				
DATA1		116304384	57439796	
58864588	49.39			
RECO1		47488896	34542516	
12946380	72.74			

```
SQL> SELECT dg.name, d.total_mb, d.os_mb, COUNT(1) num_disks
FROM v$asm_diskgroup dg, v$asm_disk d
WHERE dg.group_number = d.group_number
GROUP BY dg.name, d.total_mb, d.os_mb;
```

NAME	TOTAL_MB	OS_MB	NUM_DISKS
DATA1	692288	692288	168

RECOC1	282672	282672	168
--------	--------	--------	-----

The results of the queries show that the RECOC1 and DATAC1 disk groups and disk have been resized.

#### Related Topics

- [Determine the Amount of Available Space](#)  
To increase the size of the disks in a disk group you must either have unallocated disk space available, or you have to reallocate space currently used by a different disk group.

## 3.9 Using the Oracle Exadata System Software Rescue Procedure

In the rare event that the system disks fail simultaneously, you must use the Oracle Exadata Storage Server rescue procedure to recover the system.

- [About the Oracle Exadata System Software Rescue Procedure](#)  
The rescue procedure is necessary when system disks fail, the operating system has a corrupt file system, or there was damage to the boot area.
- [Performing the Rescue Procedure](#)  
You can use the rescue procedure to recover the Oracle Exadata Storage Server system software.
- [Configuring Exadata Database Machine Storage Servers After Rescue](#)  
After a successful rescue, you must configure the cell. If the data partitions were preserved, then the cell disks were imported automatically during the rescue procedure.
- [Configuring Oracle Exadata Database Machine Eighth Rack Storage Servers After Rescue](#)  
For storage servers that are part of an Eighth Rack system, after a successful rescue, you must configure the cell using these steps.
- [Re-creating a Damaged CELLBOOT USB Flash Drive](#)

### 3.9.1 About the Oracle Exadata System Software Rescue Procedure

The rescue procedure is necessary when system disks fail, the operating system has a corrupt file system, or there was damage to the boot area.

If only one system disk fails, then use CellCLI commands to recover.

If you are using normal redundancy, then there is only one mirror copy for the cell being rescued. The data may be irrecoverably lost if that single mirror also fails during the rescue procedure. Oracle recommends that you take a complete backup of the data on the mirror copy, and immediately take the mirror copy cell offline to prevent any new data changes to it prior to attempting a rescue. This ensures that all data residing on the grid disks on the failed cell and its mirror copy is inaccessible during rescue procedure.

The Oracle Automatic Storage Management (Oracle ASM) disk repair timer has a default repair time of 3.6 hours. If you know that you cannot perform the rescue procedure within that time frame, then you should use the Oracle ASM rebalance procedure to rebalance the disk until you can do the rescue procedure.



When using high redundancy disk groups, such as having more than one mirror copy in Oracle ASM for all the grid disks of the failed cell, then take the failed cell offline. Oracle ASM automatically drops the grid disks on the failed cell after the configured Oracle ASM time out, and starts rebalancing data using mirror copies. The default timeout is two hours. If the cell rescue takes more than two hours, then you must re-create the grid disks on the rescued cells in Oracle ASM.

 **Caution:**

Use the rescue procedure with extreme caution. Incorrectly using the procedure can cause data loss.

It is important to note the following when using the rescue procedure:

- The rescue procedure can potentially rewrite some or all of the disks in the cell. If this happens, then you can lose all the content on those disks without possibility of recovery.

Use extreme caution when using this procedure, and pay attention to the prompts. Ideally, you should use the rescue procedure only with assistance from Oracle Support Services, and when you have decided that you can afford the loss of data on some or all of the disks.

- The rescue procedure does not destroy the contents of the data disks or the contents of the data partitions on the system disks unless you explicitly choose to do so during the rescue procedure.
- Starting in Oracle Exadata System Software release 11.2, the rescue procedure restores the Oracle Exadata System Software to the same release. This includes any patches that existed on the cell as of the last successful boot. Note the following about using the rescue procedure:
  - Cell configuration information, such as alert configurations, SMTP information, administrator e-mail address, and so on is not restored.
  - The network configuration that existed at the end of last successful run of `/usr/local/bin/ipconf` utility is restored.
  - The SSH identities for the cell, and the `root`, `celladmin` and `cellmonitor` users are restored.
  - Integrated Lights Out Manager (ILOM) configurations for Oracle Exadata Storage Servers are not restored. Typically, ILOM configurations remain undamaged even in case of Oracle Exadata System Software failures.
- The rescue procedure does not examine or reconstruct data disks or data partitions on the system disks. If there is data corruption on the grid disks, then do not use the rescue procedure. Instead use the rescue procedure for Oracle Database and Oracle ASM.

After a successful rescue, you must reconfigure the cell, and if you had chosen to preserve the data, then import the cell disks. If you chose not to preserve the data, then you should create new cell disks, and grid disks.

#### Related Topics

- About Oracle ASM Disk Repair Timer
- ALTER CELL

- Configuring Cells, Cell Disks, and Grid Disks with CellCLI

## 3.9.2 Performing the Rescue Procedure

You can use the rescue procedure to recover the Oracle Exadata Storage Server system software.

1. Connect to the Oracle Exadata Storage Server using the console.
2. Start the Oracle Exadata Storage Server and select the option to boot into rescue mode.
  - On X7 and newer servers, during the initial boot sequence you will see something like the following menu of boot options:

```
Exadata_DBM_0: CELL_BOOT_trying_HD0
Exadata_DBM_0: CELL_BOOT_trying_CELLBOOT
Exadata_DBM_1: CELL_BOOT_in_rescue_mode
```

Use the ^ and v keys to change the selection.  
Press 'e' to edit the selected item, or 'c' for a command prompt.

### Note:

The above menu appears for only a short time if there are no inputs. Therefore, to retain the menu, press the up-arrow or down-arrow key immediately when it appears.

In the menu, select **Exadata\_DBM\_1: CELL\_BOOT\_in\_rescue\_mode**, and then press **Enter**.

- On X6 and older servers, during the initial boot sequence, you will see something like the following:

```
Press any key to enter the menu
Booting Exadata_DBM_0: CELL_USB_BOOT_trying_C0D0_as_HD1 in 4
seconds...
Booting Exadata_DBM_0: CELL_USB_BOOT_trying_C0D0_as_HD1 in 3
seconds...
Press any key to see the menu.
```

When you see the above, press any key to enter the boot options menu.

### Note:

For older versions of Oracle Exadata System Software, you may see the "Oracle Exadata" splash screen. If the splash screen appears, press any key on the keyboard. The splash screen remains visible for only 5 seconds.

In the boot options menu, select **CELL\_USB\_BOOT\_CELLBOOT\_usb\_in\_rescue\_mode**, and then press **Enter**.

3. When prompted, select the option to reinstall the Oracle Exadata System Software. Then, confirm your selection.

For example:

```

Choose from the following by typing letter in '()':
(e)nter interactive diagnostics shell.
Use diagnostics shell password to login as root user
(reboot or power cycle to exit the shell),
(r)einstall or try to recover damaged system,
Select: r
[INFO] Reinstall or try to recover damaged system
Continue (y/n) [n]: y

```

4. If prompted, specify the rescue root password.
 

If you do not have the required password, then contact Oracle Support Services.
5. When prompted, specify whether you want to erase the data partitions and data disks.

Specify **n** to preserve existing data on the storage server.

If you specify **y**, you will permanently erase all of the data on the storage server. Do not specify this option unless you are sure that it is safe.

For example:

```

Do you want to erase data partitions and data disks (y/n) [n]: n

```

6. When prompted, specify the root password to enter the rescue shell.
 

If you do not have the required password, then contact Oracle Support Services.

For example:

```

[INFO] You are in the rescue mode.
[INFO] Imaging pre-boot phase finished with success.
[INFO] Installation will continue after reboot.
[INFO] Log in to the rescue shell as root with the rescue
(Diagnostic shell) root password.

```

...

```

Welcome to Exadata Shell!
Give root password for maintenance
(or press Control-D to continue):

```

7. Using the rescue prompt, reboot the storage server to complete the rescue process.

For example:

```

sh-4.2# shutdown -r now

```

The rescue process typically takes between 45 and 90 minutes to complete. The storage server may reboot a few times during the rescue process. An on-screen message indicates when the rescue process is completed. For example:

```
Run validation checkconfigs - PASSED
2020-08-17 18:14:01 -0600 The first boot completed with SUCCESS
```

Finally, the login prompt is also displayed.

### 3.9.3 Configuring Exadata Database Machine Storage Servers After Rescue

After a successful rescue, you must configure the cell. If the data partitions were preserved, then the cell disks were imported automatically during the rescue procedure.

1. Re-create the cell disks and grid disks for any disks that were replaced during the rescue procedure.

- a. Create the cell disks on only the replaced disks using the following command:

```
cellcli -e create celldisk all harddisk
```

- b. Get the name of the new cell disks created.
- c. Get the mapping of the grid disks.

```
cellcli -e list griddisk attributes name,offset,size
```

Get the grid disk attributes from an existing disk. If you replaced a system disk (slot 0 or slot 1 on X6 and earlier servers), then you need to retrieve the values from the other system disk. If any of the grid disks are SPARSE grid disks, then also get the virtualsize attribute from another sparse disk.

For example, if the new grid disks are CD\_01\* and CD\_08\*, then you would use a command such as this:

```
cellcli -e list griddisk attributes
name,cachingpolicy,offset,size,virtualsize |
egrep '_CD_00|_CD_07'
DATA1_CD_00_dbm01celadm04 default 32M 779G
DATA1_CD_07_dbm01celadm04 default 32M 779G
DBFSC1_CD_07_dbm01celadm04 default 1.0575714111328125T 33.6875G
RECOC1_CD_00_dbm01celadm04 none 887.046875G
195.90625G
RECOC1_CD_07_dbm01celadm04 none 887.046875G
195.90625G
SPARSE1_CD_00_dbm01celadm04 default 779.046875G
108G 1.0546875T
SPARSE1_CD_07_dbm01celadm04 default 779.046875G
108G 1.0546875T
```

- d. Create the grid disks on the new cell disks using the retrieved attributes.

For example, using the attributes retrieved in the previous step for CD00, you would create the grid disks on CD01 as follows:

```
cellcli -e create griddisk DATA1_CD_01_dbm01celadm04
celldisk=CD_01_dbm01celadm04,
size=779G, cachingpolicy=default

cellcli -e create griddisk SPARSE1_CD_01_dbm01celadm04
celldisk=CD_01_dbm01celadm04,
size=108G, virtualsize=1.0546875T,cachingpolicy=default

cellcli -e create griddisk RECO1_CD_01_dbm01celadm04
celldisk=CD_01_dbm01celadm04 ,
size=195.90625G, cachingpolicy=none
```

Using the attributes retrieved in the previous step for CD07, you would create the grid disks on CD08 as follows:

```
cellcli -e create griddisk DATA1_CD_08_dbm01celadm04
celldisk=CD_08_dbm01celadm04,
size=779G, cachingpolicy=default

cellcli -e create griddisk SPARSE1_CD_08_dbm01celadm04
celldisk=CD_08_dbm01celadm04,
size=108G, virtualsize=1.0546875T,cachingpolicy=default

cellcli -e create griddisk RECO1_CD_08_dbm01celadm04
celldisk=CD_08_dbm01celadm04,
size=195.90625G, cachingpolicy=none

cellcli -e create griddisk DBFSC1_CD_08_dbm01celadm04
celldisk=CD_08_dbm01celadm04,
size=33.6875G, cachingpolicy=default
```

**2. Check the status of the grid disk.**

If any grid disk is inactive, alter its status to active.

```
CellCLI> ALTER GRIDDISK ALL ACTIVE
```

**3. Log in to the Oracle Automatic Storage Management (Oracle ASM) instance, and set the disks to ONLINE for each disk group:**

```
SQL> ALTER DISKGROUP disk_group_name ONLINE DISKS IN FAILGROUP \
cell_name WAIT;
```

 **Note:**

- If the command fails because the disks were already force-dropped, then you need to force-add the disks back to the Oracle ASM disk groups.
- The grid disk attributes `asmmodestatus` and `asmdeactivationoutcome` will not report correctly until the `ALTER DISKGROUP` statement is complete.

4. Reconfigure the cell using the `ALTER CELL` command.

In the following example, e-mail notification is configured to send notification messages to the storage server administrator according to the specified notification policy:

```
CellCLI> ALTER CELL -
 mailServer='mail_relay.example.com', -
 smtpFromAddr='john.doe@example.com', -
 smtpToAddr='jane.smith@example.com', -
 notificationPolicy='critical,warning,clear', -
 notificationMethod='mail,snmp'
```

## 5. Re-create the I/O Resource Management (IORM) plan.

## 6. Re-create the metric thresholds.

**Related Topics**

- Administering Database Resource Management
- `ALTER THRESHOLD`

## 3.9.4 Configuring Oracle Exadata Database Machine Eighth Rack Storage Servers After Rescue

For storage servers that are part of an Eighth Rack system, after a successful rescue, you must configure the cell using these steps.

In Oracle Exadata System Software release 11.2.3.3 and later, no extra steps are needed after cell rescue.

1. Copy the `/opt/oracle.SupportTools/resourcecontrol` utility from another storage server to the `/opt/oracle.SupportTools/resourcecontrol` directory on the recovered server.
2. Ensure proper permissions are set on the utility.

```
chmod 740 /opt/oracle.SupportTools/resourcecontrol
```

3. Verify the current configuration.

```
/opt/oracle.SuppprtTools/resourcecontrol -show
```

```
Validated hardware and OS. Proceed.
Number of cores active: 6
```

```
Number of harddisks active: 6
Number of flashdisks active: 8
```

For an Eighth Rack configuration, the output depends on the hardware model:

- X3 storage server: 6 active CPU, 6 hard disks, and 8 flash disks should be enabled
  - X4 storage server: 6 active CPU cores, 6 hard disks, and 8 flash disks should be enabled
  - X5 HC storage server: 8 active CPU cores, 6 hard disks, and 2 flash disks should be enabled
  - X5 EF storage server: 8 active CPU cores and 4 flash disks should be enabled
  - X6 HC storage server: 10 active CPU cores, 6 hard disks, and 2 flash disks should be enabled
  - X6 EF storage server: 10 active CPU cores and 4 flash disks should be enabled
  - X7 HC storage server: 10 active CPU cores, 6 hard disks, and 2 flash disks should be enabled
  - X7 EF storage server: 10 active CPU cores and 4 flash disks should be enabled
  - X8 HC storage server: 16 active CPU cores, 6 hard disks, and 2 flash disks should be enabled
  - X8 EF storage server: 16 active CPU cores and 4 flash disks should be enabled
4. If the configuration shows all the cores and disks enabled, then enable the Eighth Rack configuration.

```
CellCLI> ALTER CELL eighthRack=true
```

## 3.9.5 Re-creating a Damaged CELLBOOT USB Flash Drive

If the CELLBOOT USB flash drive is lost or damaged, then you can create a new one using the following procedure:



### Note:

To create a USB flash drive for a machine running Oracle Exadata Storage Server Software release 12.1.2.1.0 or later requires a machine running Oracle Linux 6.

1. Log in to the cell as the `root` user.
2. On X6 and older servers only, which do not contain M.2 system devices:
  - a. Attach a new USB flash drive.

This flash drive should have a capacity of at least 1 GB, and up to 8 GB.

- b. Remove any other USB flash drives from the system.
3. Run the following commands:

```
cd /opt/oracle.SupportTools
./make_cellboot_usb -rebuild -verbose
```

## 3.10 Changing Existing Elastic Configurations for Storage Cells

You can modify the capacity of your Exadata Database Machine using elastic configuration.

- [Adding a Cell Node](#)  
In this scenario, you want to add a new storage server (or cell) to an existing Exadata Database Machine that includes disk groups.
- [Adding a New Storage Server to an Eighth Rack Cluster](#)  
Perform the following steps to add a new Exadata Database Machine X7 or later storage server to an existing Exadata Database Machine X7 or later Eighth Rack.
- [Adding Storage Cells using OEDACLI](#)  
OEDACLI provides the interface to execute elastic storage expansion for different configurations such as Bare Metal, single Oracle VM or multiple Oracle VMs.
- [Expanding an Existing Exadata Storage Grid](#)
- [Dropping a Storage Server from an Existing Disk Group or Storage Grid](#)  
You can remove a storage server from an existing Oracle Exadata Rack.
- [Dropping Storage Servers using OEDACLI](#)  
OEDACLI provides the interface to drop storage servers for different configuration such as Bare Metal, single Oracle VM or multiple Oracle VMs.

### Related Topics

- [Changing Existing Elastic Configurations for Database Servers](#)  
Elastic configurations provide a flexible and efficient mechanism to change the server configuration of your Exadata Database Machine.

### 3.10.1 Adding a Cell Node

In this scenario, you want to add a new storage server (or cell) to an existing Exadata Database Machine that includes disk groups.

1. If this is a brand new storage server, perform these steps:
  - a. Complete all necessary cabling requirements to make the new storage server available to the desired storage grid.  
*Refer to the [Oracle Exadata Database Machine Installation and Configuration Guide](#).*
  - b. Image the storage server with the appropriate Oracle Exadata System Software image and provide appropriate input when prompted for the IP addresses.
  - c. Skip to step 2.
2. If this is an existing storage server in the rack and you are allocating it to another cluster within the RDMA Network Fabric network, note the IP addresses assigned to the RDMA Network Fabric interfaces (such as `ib0` and `ib1` or `re0` and `re1`) of the storage server being added.



Add the IP addresses to the `/etc/oracle/cell/network-config/cellip.ora` file on every Oracle RAC node.

- a. `cd /etc/oracle/cell/network-config`
- b. `cp cellip.ora cellip.ora.orig`
- c. `cp cellip.ora cellip.ora-bak`
- d. Add the new entries to `/etc/oracle/cell/network-config/cellip.ora-bak`.
- e. Copy the edited file to the `cellip.ora` file on all database nodes using the following command, where `database_nodes` refers to a file containing the names of each database server in the cluster, with each name on a separate line:

```
/usr/local/bin/dcli -g database_nodes -l root -f cellip.ora-bak -d
/etc/oracle/cell/network-config/cellip.ora
```

3. If Oracle Auto Service Request (ASR) alerting was set up on the existing storage servers, configure cell Oracle ASR alerting for the storage server being added.

- a. From any existing storage server, list the cell `snmpsubscriber` attribute.

```
cellcli -e LIST CELL ATTRIBUTES snmpsubscriber
```

- b. Apply the same `snmpsubscriber` attribute value to the new storage server by running the command below as the `celladmin` user, replacing `snmpsubscriber` with the value from the previous command.

```
cellcli -e "ALTER CELL snmpsubscriber=snmpsubscriber"
```

- c. From any existing storage server, list the cell attributes required for configuring cell alerting.

```
cellcli -e LIST CELL ATTRIBUTES
```

```
notificationMethod,notificationPolicy,mailServer,smtpToAddr,smtpFrom,
smtpFromAddr,smtpUseSSL,smtpPort
```

- d. Apply the same values to the new storage server by running the command below as the `celladmin` user, substituting the placeholders with the values found from the existing storage server.

```
cellcli -e "ALTER CELL
notificationMethod='notificationMethod',
notificationPolicy='notificationPolicy',
mailServer='mailServer',
smtpToAddr='smtpToAddr',
smtpFrom='smtpFrom',
smtpFromAddr='smtpFromAddr',
smtpUseSSL=smtpUseSSL,
smtpPort=smtpPort"
```

4. If needed, create cell disks on the cell being added.

- a. Log in to the cell as `celladmin` and list the cell disks.

```
cellcli -e LIST CELLDISK
```

- b. If the cell disks are not present on the media types, then create the cell disks.

```
cellcli -e CREATE CELLDISK ALL
```

- c. If your system has PMEM devices, then check that the PMEM log was created by default.

```
cellcli -e LIST PMEMLOG
```

You should see the name of the PMEM log. It should look like `cellnodename_PMEMLOG`, and its status should be `normal`.

If the PMEM log does not exist, create it.

```
cellcli -e CREATE PMEMLOG ALL
```

- d. Check that the flash log was created by default.

```
cellcli -e LIST FLASHLOG
```

You should see the name of the flash log. It should look like `cellnodename_FLASHLOG`, and its status should be `normal`.

If the flash log does not exist, create it.

```
cellcli -e CREATE FLASHLOG ALL
```

- e. If your system has PMEM devices, then check the current PMEM cache mode and compare it to the PMEM cache mode on existing cells.

```
cellcli -e LIST CELL ATTRIBUTES pmemcachemode
```

To change the PMEM cache mode to match the PMEM cache mode of existing cells, do the following:

- i. If the PMEM cache exists and the cell is in `WriteBack` PMEM cache mode, you must first flush the PMEM cache.

```
cellcli -e ALTER PMEMCACHE ALL FLUSH
```

Wait for the command to return.

If the PMEM cache mode is `WriteThrough`, then you do not need to flush the cache first.

- ii. Drop the PMEM cache.

```
cellcli -e DROP PMEMCACHE ALL
```

**iii.** Change the PMEM cache mode.

The value of the `pmemCacheMode` attribute is either `writeback` or `writethrough`. The value has to match the PMEM cache mode of the other storage cells in the cluster.

```
cellcli -e "ALTER CELL
PMEMCacheMode=writeback_or_writethrough"
```

**iv.** Re-create the PMEM cache.

```
cellcli -e CREATE PMEMCACHE ALL
```

**f.** Check the current flash cache mode and compare it to the flash cache mode on existing cells.

```
cellcli -e LIST CELL ATTRIBUTES flashcachemode
```

To change the flash cache mode to match the flash cache mode of existing cells, do the following:

**i.** If the flash cache exists and the cell is in `WriteBack` flash cache mode, you must first flush the flash cache.

```
cellcli -e ALTER FLASHCACHE ALL FLUSH
```

Wait for the command to return.

**ii.** Drop the flash cache.

```
cellcli -e DROP FLASHCACHE ALL
```

**iii.** Change the flash cache mode.

The value of the `flashCacheMode` attribute is either `writeback` or `writethrough`. The value has to match the flash cache mode of the other storage cells in the cluster.

```
cellcli -e "ALTER CELL
flashCacheMode=writeback_or_writethrough"
```

**iv.** Create the flash cache.

```
cellcli -e CREATE FLASHCACHE ALL
```

**5.** Create the grid disks on the cell being added.**a.** Query the `size` and `cachingpolicy` attributes of the existing grid disks from an existing cell.

```
CellCLI> LIST GRIDDISK ATTRIBUTES
name,asmDiskGroupName,cachingpolicy,size,offset
```

**b.** For each disk group found by the above command, create grid disks on the new cell that is being added to the cluster.

Match the `size` and the `cachingpolicy` attributes of the existing grid disks for the particular disk group reported by the command above. Grid disks should be created in the order of increasing offset to ensure similar layout and performance characteristics as the existing cells. For example, the `LIST GRIDDISK` command could return something like this:

DATA1	default	2.15625T	32M
DBFS_DG	default	33.796875G	2.695465087890625T
RECO1	none	552.109375G	2.1562957763671875T

When creating grid disks, begin with `DATA1`, then `RECO1`, and finally `DBFS_DG` using the following command:

```
cellcli -e CREATE GRIDDISK ALL HARDDISK
 PREFIX=matching_prefix_of_the_corresponding_existing_diskgroup,
 size=size_followed_by_G_or_T,
 cachingPolicy='value_from_command_above_for_this_disk_group',
 comment ="Cluster cluster_name diskgroup diskgroup_name"
```

### Caution:

Be sure to specify the EXACT size shown along with the unit (either T or G)

6. Log in to each Oracle RAC node and verify that the newly created grid disks are visible from the Oracle RAC nodes.

In the following example, *Grid\_home* refers to the directory in which the Oracle Grid Infrastructure software is installed.

```
$Grid_home/bin/kfod op=disks disks=all | grep cellName_being_added
```

The `kfod` command should list all the grid disks created in step 5 above.

7. Add the newly created grid disks to the respective existing Oracle ASM disk groups.

In this example, *comma\_separated\_disk\_names* refers to the disk names from step 5 corresponding to *disk\_group\_name*.

```
SQL> ALTER DISKGROUP disk_group_name ADD DISK
'comma_separated_disk_names';
```

This command kicks off an Oracle ASM rebalance at the default power level.

8. Monitor the progress of the rebalance by querying `GV$ASM_OPERATION`.

```
SQL> SELECT * FROM GV$ASM_OPERATION;
```

When the rebalance completes, the addition of the cell to the Oracle RAC cluster is complete.

9. Download and run the latest version of Oracle EXAchk to ensure that the resulting configuration implements the latest best practices for Exadata Database Machine.

## 3.10.2 Adding a New Storage Server to an Eighth Rack Cluster

Perform the following steps to add a new Exadata Database Machine X7 or later storage server to an existing Exadata Database Machine X7 or later Eighth Rack.

1. If configured, drop the PMEM Cache and PMEM log.

```
$ cellcli -e drop pmemcache all
$ cellcli -e drop pmemlog all
```

2. On the new storage server, drop the flash cache, flash log and cell disks.

```
cellcli -e drop flashcache all
cellcli -e drop flashlog all
cellcli -e drop celldisk all
```

3. On the new storage server, enable the `eighthrack` attribute.

```
cellcli -e alter cell eighthRack=true
```

4. On the new storage server, create the cell disks.

```
cellcli -e create celldisk all
```

5. On the new storage server, create the flash log.

```
cellcli -e create flashlog all
```

6. If applicable, on the new storage server, create the PMEM log.

```
cellcli -e create pmemlog all
```

7. On any of the existing storage servers, retrieve the value of the cell attribute `flashcachemode`.

```
cellcli -e list cell attributes flashcachemode
```

The `flashcachemode` attribute on the new storage server is set to `WriteThrough` by default. All storage servers should have the same `flashcachemode` attribute setting.

If the existing storage servers are using `WriteBack` mode, then you should change the attribute `flashcachemode` on the new storage server, as shown here:

```
cellcli -e alter cell flashcachemode=writeback
```

8. On the new storage server, create the flash cache.

```
cellcli -e create flashcache all
```

9. If the storage servers use PMEM cache, then retrieve the value of the cell attribute `pmemcachemode`.

```
cellcli -e list cell attributes pmemcachemode
```

The `pmemcachemode` attribute on the new storage server is set to `WriteThrough` by default. All storage servers should have the same `pmemcachemode` attribute setting.

If the existing storage servers are using `WriteBack` mode, then you should change the attribute `pmemcachemode` on the new storage server, as shown here:

```
cellcli -e alter cell pmemcachemode=writeback
```

10. If the storage servers use PMEM cache, then, on the new storage server, create the PMEM cache.

```
cellcli -e create pmemcache all
```

11. On any of the existing storage servers, obtain information on the grid disk configuration.

```
cellcli -e list griddisk attributes name,offset,size,cachingpolicy
```

12. On the new storage server, create the grid disks (repeat for each set of grid disks to match the configuration of the existing storage servers).

In the following command, replace the italicized text with the corresponding values obtained in step 11.

```
cellcli -e CREATE GRIDDISK ALL HARDDISK PREFIX=matching_prefix_of_the_
corresponding_existing_diskgroup, size=size_followed_by_G_or_T,
cachingPolicy='value_from_command_above_for_this_disk_group',
comment ="Cluster cluster_name diskgroup diskgroup_name"
```

13. On the new storage server, validate the grid disks have the same configuration of the grid disks as the existing storage servers (by comparing with the information obtained in step 11).

```
cellcli -e list griddisk attributes name,offset,size,cachingpolicy
```

14. (X2 to X8 servers only) If the environment has partition keys (pkeys) implemented, configure pkeys for the RDMA Network Fabric interfaces. Refer to step 6 from [Implementing InfiniBand Partitioning across OVM RAC clusters on Exadata \(My Oracle Support Doc ID 2075398.1\)](#) for this task.

15. On the new storage server, identify the IP address for both ports for either InfiniBand Network Fabric or RoCE Network Fabric.

```
cellcli -e list cell attributes name,ipaddress1,ipaddress2
```

16. Add the IP addresses from step 15 to the `/etc/oracle/cell/network-config/cellip.ora` file on every database server.

Perform these steps on any database server in the cluster:

- a. `cd /etc/oracle/cell/network-config`

- b. `cp cellip.ora cellip.ora.orig`
- c. `cp cellip.ora cellip.ora-bak`
- d. Add the new entries to `/etc/oracle/cell/network-config/cellip.ora-bak`.
- e. Copy the edited file to the `cellip.ora` file on all database s using the following command, where `database_nodes` refers to a file containing the names of each database server in the cluster, with each name on a separate line:

```
/usr/local/bin/dcli -g database_nodes -l root -f cellip.ora-bak -
d /etc/oracle/cell/network-config/cellip.ora
```

17. Connect to any of the Oracle ASM instances and ensure the grid disks from the new storage server are discoverable.

```
SQL> set pagesize 30
SQL> set linesize 132
SQL> col path format a70
SQL> SELECT inst_id,path FROM gv$asm_disk WHERE
header_status='CANDIDATE'
 2> ORDER BY inst_id,path;
```

```
INST_ID PATH

--
1 o/192.168.17.235;192.168.17.236/DATA1_CD_00_celadm11
1 o/192.168.17.235;192.168.17.236/DATA1_CD_01_celadm11
1 o/192.168.17.235;192.168.17.236/DATA1_CD_02_celadm11
1 o/192.168.17.235;192.168.17.236/DATA1_CD_03_celadm11
1 o/192.168.17.235;192.168.17.236/DATA1_CD_04_celadm11
1 o/192.168.17.235;192.168.17.236/DATA1_CD_05_celadm11
1 o/192.168.17.235;192.168.17.236/RECO1_CD_00_celadm11
1 o/192.168.17.235;192.168.17.236/RECO1_CD_01_celadm11
1 o/192.168.17.235;192.168.17.236/RECO1_CD_02_celadm11
1 o/192.168.17.235;192.168.17.236/RECO1_CD_03_celadm11
1 o/192.168.17.235;192.168.17.236/RECO1_CD_04_celadm11
1 o/192.168.17.235;192.168.17.236/RECO1_CD_05_celadm11
2 o/192.168.17.235;192.168.17.236/DATA1_CD_00_celadm11
2 o/192.168.17.235;192.168.17.236/DATA1_CD_01_celadm11
2 o/192.168.17.235;192.168.17.236/DATA1_CD_02_celadm11
2 o/192.168.17.235;192.168.17.236/DATA1_CD_03_celadm11
2 o/192.168.17.235;192.168.17.236/DATA1_CD_04_celadm11
2 o/192.168.17.235;192.168.17.236/DATA1_CD_05_celadm11
2 o/192.168.17.235;192.168.17.236/RECO1_CD_00_celadm11
2 o/192.168.17.235;192.168.17.236/RECO1_CD_01_celadm11
2 o/192.168.17.235;192.168.17.236/RECO1_CD_02_celadm11
2 o/192.168.17.235;192.168.17.236/RECO1_CD_03_celadm11
2 o/192.168.17.235;192.168.17.236/RECO1_CD_04_celadm11
2 o/192.168.17.235;192.168.17.236/RECO1_CD_05_celadm11
```

18. Connect to one of the Oracle ASM instances and add the new disks to the existing disk groups.

```
SQL> ALTER DISKGROUP datacl ADD DISK 'o/192.168.17.235;192.168.17.
236/DATAC1*';
SQL> ALTER DISKGROUP recocl ADD DISK 'o/192.168.17.235;192.168.17.
236/RECOC1*';
```

 **Note:**

The rebalance operation triggered by adding the disks will run at the default Oracle Maximum Availability Architecture (MAA) best practice power (should be 4). If the application service level performance is not a concern, then consider increasing the power for a faster rebalance.

19. Obtain a report of the number of disks per failure group. 6 disks per failure group are expected for High Capacity (HC) Storage Servers and 4 disks per failure group are expected for Extreme Flash (EF) Storage Servers.

```
SQL> SELECT d.group_number,dg.name,failgroup,mode_status,COUNT(*)
2> FROM v$asm_disk d,v$asm_diskgroup dg
3> WHERE d.group_number=dg.group_number
4> AND failgroup_type='REGULAR'
5> GROUP BY d.group_number,dg.name,failgroup,mode_status;
```

GROUP_NUMBER	NAME	FAILGROUP	MODE_ST	COUNT(*)
1	DATAC1	CELADM08	ONLINE	6
1	DATAC1	CELADM09	ONLINE	6
1	DATAC1	CELADM10	ONLINE	6
1	DATAC1	CELADM11	ONLINE	6
2	RECOC1	CELADM08	ONLINE	6
2	RECOC1	CELADM09	ONLINE	6
2	RECOC1	CELADM10	ONLINE	6
2	RECOC1	CELADM11	ONLINE	6

20. If Oracle Auto Service Request (ASR) alerting was set up on the existing storage servers, configure cell Oracle ASR alerting for the storage server being added.

- a. From any existing storage server, list the cell `snmpsubscriber` attribute.

```
cellcli -e LIST CELL ATTRIBUTES snmpsubscriber
```

- b. Apply the same `snmpsubscriber` attribute value to the new storage server by running the command below as the `celladmin` user, replacing `snmpsubscriber` with the value from the previous command.

```
cellcli -e "ALTER CELL snmpsubscriber=snmpsubscriber"
```



- c. From any existing storage server, list the cell attributes required for configuring cell alerting.

```
cellcli -e LIST CELL ATTRIBUTES

notificationMethod,notificationPolicy,mailServer,smtpToAddr,smtpFrom,
smtpFromAddr,smtpUseSSL,smtpPort
```

- d. Apply the same values to the new storage server by running the command below as the `celladmin` user, substituting the placeholders with the values found from the existing storage server.

```
cellcli -e "ALTER CELL
notificationMethod='notificationMethod',
notificationPolicy='notificationPolicy',
mailServer='mailServer',
smtpToAddr='smtpToAddr',
smtpFrom='smtpFrom',
smtpFromAddr='smtpFromAddr',
smtpUseSSL=smtpUseSSL,
smtpPort=smtpPort"
```

### 3.10.3 Adding Storage Cells using OEDACLI

OEDACLI provides the interface to execute elastic storage expansion for different configurations such as Bare Metal, single Oracle VM or multiple Oracle VMs.

OEDACLI creates all the objects required on the storage servers and adds the new grid disks to the disk groups. One of the existing storage servers acts as a guide for the configuration of the new storage servers. There is only one rebalance operation triggered even if multiple storage servers are added. The grid disks of the new servers are added to all the disk groups for all the clusters configured.

#### Prerequisites

- All the new storage servers must be installed in and connected into the physical rack.
  - All the new storage servers must have the management and RDMA Network Fabric networks configured.
  - OEDACLI must be executed from a machine that has network access to the database servers (bare metal or guest domains) and the storage servers.
1. Generate a new OEDA XML configuration file, reflecting an exact list of the current database servers and storage servers used in the environment.

Use the OEDA `DISCOVER` command, where the *dirname* usually is the directory where OEDACLI is installed and *host\_names\_list* is the list of nodes to be discovered, separated by commas or spaces, for example,  
'dbnode1,dbnode2,celadm01,celadm02'.

```
DISCOVER ES hostnames='host_names_list' location='dirname'
```

For an environment with multiple Oracle VMs, the command generates a global XML configuration file containing information of all clusters, and also one XML configuration file for each cluster. In the following commands, you should use the global XML configuration file instead of a cluster-specific configuration file.

2. Create an OEDACLI script to update the XML configuration file using OEDACLI.

For each cell to be added, the script requires:

- The name (hostname) of one cell which is part of the current configuration, to be used as a reference for the creation of the objects (cell disks, grid disks, flash cache, and so on)
- The name (hostname) and IP addresses for the Management and RDMA Network Fabric interfaces.
- The rack number. For non-interconnected environments this is 1.
- ULOC is the location on the physical rack. Although it is not used for the storage expansion, pick the value according the information referenced in *Oracle Exadata Database Machine System Overview, Part 2 - Cabling Diagrams*

Save the following commands in a file named `add_cell_script.cmd`. In this example, two new storage servers are being added: `celadm04` and `celadm05`.

```
CLONE NEWCELL SRCNAME=celadm01 tgtname=celadm04
SET ADMINNET NAME=celadm04, IP=203.0.113.35
SET PRIVNET NAME1=celadm04-priv1, IP1=192.168.216.235, NAME2=celadm04-
priv2, IP2=192.168.216.236
SET ILOMNET NAME=celadm04-ilom, IP=203.0.113.135
SET RACK NUM=1, ULOC=39
SAVE ACTION FORCE
CLONE NEWCELL SRCNAME=celadm01 tgtname=celadm05
SET ADMINNET NAME=celadm05, IP=203.0.113.36
SET PRIVNET NAME1=celadm05-priv1, IP1=192.168.216.221, NAME2=celadm05-
priv2, IP2=192.168.216.222
SET ILOMNET NAME=celadm05-ilom, IP=203.0.113.136
SET RACK NUM=1, ULOC=14
SAVE ACTION FORCE
SAVE FILE
```

3. Run the script `add_cell_script.cmd`.

In this example, the `oeda_xml_file` is the file generated in the first step, and `add_cell_script.cmd` is the script created in the previous step.

```
$ oedacli -c oeda_xml_file -f add_cell_script.cmd
```

When you run the script, it updates the OEDA XML configuration file, adding directives and attributes related to the new storage servers. It does not trigger any action on any component (storage servers or database servers).

4. Create an OEDACLI script to configure the new storage server(s), or cell(s).

The script uses the following settings:

- The Oracle ASM power limit for rebalance is set to 4, which is an Oracle MAA best practice.

- The `WAIT` option is set to `FALSE`, which means the disk rebalance operations are run in parallel for every disk group in the cluster. The number of outstanding rebalances that can be run concurrently is limited to the number of database servers. If you set `WAIT` to `TRUE`, then each rebalance operation is run sequentially.

Save the following commands in a file named `config_cell_script.cmd`. In this example, the cluster name is `q1-vm01`. Replace this with the name of your cluster. Also, replace the example cell names (`celadm04, celadm05`) with your own.

```
ALTER CLUSTER ADDCELLS='celadm04,celadm05' power=4, wait=false
WHERE clustername='q1-vm01'
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
SAVE FILE
```

5. Run the script `config_cell_script.cmd`.

In this example, the `oeda_xml_file` is the file generated in the first step, and `config_cell_script.cmd` is the script created in the previous step.

```
$ oedacli -c oeda_xml_file -f config_cell_script.cmd
```

When you run the script, it creates the flash cache, flash log, cell disks and grid disks. It also adds the disks to the Oracle ASM disk groups and adds the storage servers to the cluster.

### 3.10.4 Expanding an Existing Exadata Storage Grid

In this scenario, you have an Exadata storage cell in an Exadata rack, and you want to add the storage cell to an Exadata storage grid that you most probably want to expand.

1. Decommission the storage cell from its current cluster. To do this, follow the procedure in "[Dropping a Storage Server from an Existing Disk Group or Storage Grid](#)".
2. Add the storage cell to the desired Exadata storage grid. To do this, follow the procedure in "[Adding a Cell Node](#)".
3. Download and run the latest `exachk` to ensure that the resulting configuration implements the latest best practices for Oracle Exadata.

### 3.10.5 Dropping a Storage Server from an Existing Disk Group or Storage Grid

You can remove a storage server from an existing Oracle Exadata Rack.

1. Drop the disks belonging to the storage server to be removed from Oracle Automatic Storage Management (Oracle ASM).

 **Note:**

For Exadata Database Machine VM deployments, the substeps below need to be executed from all the Oracle VM clusters.

- a. Log in to any node in the cluster.
- b. Query the list of grid disks being used by the cluster for the targeted storage server.

```
Grid_home/bin/asmcmd lsdsk --suppressheader | grep
cellName_being_removed | awk -F '/' '{print $NF}'
```

 **Note:**

Make sure the available free space in every disk group that contains disks from the storage server being removed is at least 15% of the allocated storage for that disk group.

- c. Drop the Oracle ASM disks returned by the command above from their respective disk groups.

```
SQL> ALTER DISKGROUP diskgroup_name DROP DISKS IN FAILGROUP
cellName_being_removed;
```

- d. The disk drop operation above kicks off a rebalance operation at the default power level. Monitor for the rebalance using the following command:

```
SQL> SELECT * FROM gv$asm_operation;
```

Wait until the rebalance completes, that is, wait until `gv$asm_operation` returns no rows.

- e. Verify that all the disk groups do not have any references to the disks from the storage server being removed.

```
SQL> SELECT path, name, header_status, mode_status, mount_status,
state,
failgroup FROM v$asm_disk ORDER BY path;
```

The `header_status` column for all the disks belonging to the storage server being removed should show `FORMER`.

 **Reminder:**

For Exadata Oracle VM deployments, the substeps above need to be executed from all the Oracle VM clusters.

2. Clean up the storage server being removed.

Log in to the storage server as `celladmin` and run the following commands. Run the following commands for each set of grid disks:

- a. Drop the grid disks.

```
cellcli -e drop griddisk all prefix=prefix_of_the_grid_disk
```

- b. If flash cache exists and the storage server is in WriteBack flash cache mode, you must first flush the flash cache before dropping it.

```
cellcli -e alter flashcache all flush
```

Wait for the command to return.

- c. Drop the flash cache.

```
cellcli -e drop flashcache all
```

- d. Drop the cell disks.

```
cellcli -e drop celldisk all
```

If you need to erase data securely, you can run the `DROP CELLDISK` command with the `erase` option, or the `DROP CELL` with the `erase` option.

The time required to complete the erase operation is listed in the table under the `DROP CELL` command.

3. Remove the entry of the storage server being removed from `/etc/oracle/cell/network-config/cellip.ora` on all the database server nodes in the cluster.

Run the following steps on any database server node in the cluster:

- a. Make a backup copies of the `cellip.ora` file.

```
cd /etc/oracle/cell/network-config
cp cellip.ora cellip.ora.orig
cp cellip.ora cellip.ora-bak
```

- b. Remove the entries for the storage server being removed from `/etc/oracle/cell/network-config/cellip.ora-bak`.

- c. Use `dcli` to copy the updated `cellip.ora-bak` file to the other database servers.

In the following command `database_nodes` refers to a file containing the names of each database server in the cluster. Each database server name is on a separate line in the file.

```
/usr/local/bin/dcli -g database_nodes -l root -f cellip.ora-bak -
d
/etc/oracle/cell/network-config/cellip.ora
```

4. Download and run the latest version of Oracle EXAchk to ensure that the resulting configuration implements the latest best practices for Exadata Database Machine.

## 3.10.6 Dropping Storage Servers using OEDACLI

OEDACLI provides the interface to drop storage servers for different configuration such as Bare Metal, single Oracle VM or multiple Oracle VMs.

The procedure to remove the grid disks from the disk groups and drop the objects on the storage server is implemented through only a few commands. There is only one rebalance operation triggered, regardless of the number of storage cells dropped.

### Prerequisites

- A valid OEDA XML configuration file, reflecting an exact list of the compute nodes and storage cells used in the environment to be expanded. The first step in this task generates a new OEDA XML configuration file to ensure that the current configuration is used.
  - OEDACLI must be executed from a machine that has network access to the database servers (bare metal or guest domains) and the storage servers.
1. Generate a new OEDA XML configuration file, reflecting an exact list of the current database servers and storage servers used in the environment.

Use the OEDA `DISCOVER` command, where the *dirname* usually is the directory where OEDACLI is installed and *host\_names\_list* is the list of nodes to be discovered, separated by commas or spaces, for example, 'dbnode1,dbnode2,celadm01,celadm02'.

```
DISCOVER ES hostnames='host_names_list' location='dirname'
```

For an environment with multiple Oracle VMs, the command generates a global XML configuration file containing information of all clusters, and also one XML configuration file for each cluster. In the following commands, you should use the global XML configuration file instead of a cluster-specific configuration file.

2. Create an OEDACLI script to update the XML configuration file using OEDACLI.

In this example, the `WAIT` option is set to `TRUE`, which means each rebalance operation is run sequentially. After the rebalance for the last disk group has completed, the storage servers are removed from the configuration.

Save the following commands in a file named `drop_cell_cluster.cmd`. In this example, two storage servers are being removed: `celadm04` and `celadm05`. In this example, the clustername `q1-vm01` is used. You should replace this with the name of your cluster.

```
ALTER CLUSTER DROPCELLS='celadm04,celadm05' power=4, wait=true \
WHERE clustername=q1-vm01
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
SAVE FILE
```

3. Run the script `drop_cell_cluster.cmd`.

In this example, the *oeda\_xml\_file* is the file generated in the first step, and `drop_cell_cluster.cmd` is the script created in the previous step.

```
$ oedacli -c oeda_xml_file -f drop_cell_cluster.cmd
```

When you run the script, it drops the grid disks from all the Oracle ASM disk groups configured in the cluster, de-configures the cell (drops objects like flash cache, flash log, cell disks, grid disks, and so on).

4. Create an OEDACLI script to remove the storage servers, or cells, from the OEDA XML configuration file.

Save the following commands in a file named `drop_cell_xml.cmd`.

```
DELETE NEWCELL WHERE SRCNAME='celadm04 celadm05'
SAVE ACTION FORCE
SAVE FILE
```

5. Run the script `drop_cell_xml.cmd`.

In this example, the `oeda_xml_file` is the file generated in the first step, and `drop_cell_xml.cmd` is the script created in the previous step.

```
$ oedacli -c oeda_xml_file -f drop_cell_xml.cmd
```

When you run the script, it removes the information for the storage servers from the OEDA XML configuration file

## 3.11 Managing Disk Controller Batteries

This section applies only to Exadata systems prior to X6 that use batteries. Newer systems have CVPM02 (Cache Vault), which is a super cap and not a battery.

- [About Disk Controller Batteries](#)  
The disk controllers in Oracle Exadata storage servers and database servers have battery-backed write cache to accelerate write performance.
- [Monitoring Batteries in the Database Servers](#)
- [Replacing Batteries in Disk Controllers](#)

### 3.11.1 About Disk Controller Batteries

The disk controllers in Oracle Exadata storage servers and database servers have battery-backed write cache to accelerate write performance.

#### Note:

This applies only to Exadata systems prior to X6 that use batteries. Newer systems have CVPM02 (Cache Vault), which is a super cap and not a battery.

If the battery charge capacity degrades such that the battery can no longer protect the cached data for a power loss of 48 hours or more, then the write cache is disabled and the disk controller switches to write through mode. This results in reduced write performance, but there is no data loss. Oracle Exadata Storage Servers generate an alert when battery charge capacity is insufficient or the temperature is high, and when the battery should be replaced.

Battery charge capacity degrades over time, and its life expectancy is inversely proportional to the operating temperature. The worst case life expectancy of the battery in Oracle Exadata Rack is as follows:

Inlet Ambient Temperature	Battery Lifetime
< 25 degrees Celsius (77 degrees Fahrenheit)	3 years
< 32 degrees Celsius (89.6 degrees Fahrenheit)	2 years

### 3.11.2 Monitoring Batteries in the Database Servers

 **Note:**

Exadata Storage Servers generate an alert when battery charge capacity is insufficient or the temperature is high, and when the battery should be replaced.

The battery charge capacity and battery temperature in the database servers can be monitored using the following commands:

 **Note:**

If you are running Oracle Exadata System Software 19.1.0 or later, substitute `/opt/MegaRAID/storcli/storcli64` for `/opt/MegaRAID/MegaCli/MegaCli64` in the following commands:

```
/opt/MegaRAID/MegaCli/MegaCli64 -AdpBbuCmd -a0 | grep "Full Charge" -A5 |
sort \
| grep Full -A1
```

The following is an example of the output from the command:

```
Full Charge Capacity: 1357 mAh
Max Error: 2 %
```

Proactive battery replacement should be done on batteries that show capacity less than 800 mAh and have maximum error less than 10%. Immediately replace any battery that has less than 674 mAh or has maximum error more than 10%.

The battery temperature can be monitored using the following command:

```
/opt/MegaRAID/MegaCli/MegaCli64 -AdpBbuCmd -a0 | grep BatteryType; \
/opt/MegaRAID/MegaCli/MegaCli64 -AdpBbuCmd -a0 | grep -i temper
```

The following is an example of the output from the command:

```
BatteryType: iBBU08
Temperature: 38 C
```



```
Temperature : OK
Over Temperature : No
```

If the battery temperature is greater than or equal to 55 degrees Celsius, then determine the cause, and correct the problem.

### 3.11.3 Replacing Batteries in Disk Controllers

If the battery charge capacity in the disk controllers falls below the minimum threshold, then Oracle will replace the failed batteries at no extra charge, if the system is covered either by the Oracle Premier Support for Systems or occurs during the warranty period.

For customers with Premier Support for Systems, Oracle attempts to proactively replace the batteries in Oracle Exadata Rack before the end of the estimated lifetime, on a best efforts basis.

## 3.12 Managing F20 PCIe Energy Storage Modules

Sun Flash Accelerator F20 PCIe cards are used in Exadata Database Machine X3 models.

- [About F20 PCIe Energy Storage Modules](#)  
The Sun Flash Accelerator F20 PCIe card includes an energy storage module (ESM) to ensure data integrity during a power interruption, functioning similar to a battery backup.
- [Replacing Flash ESM](#)

### 3.12.1 About F20 PCIe Energy Storage Modules

The Sun Flash Accelerator F20 PCIe card includes an energy storage module (ESM) to ensure data integrity during a power interruption, functioning similar to a battery backup.

Sun Flash Accelerator F20 PCIe cards accelerate performance in Oracle Exadata Rack by caching frequently-accessed Oracle Database data and avoids the need to do physical I/O to the disk in Exadata Storage Server. Write operations to the flash cards are temporarily staged in volatile local DRAM memory on the card to speed write operations. The data in the DRAM is protected by an Energy Storage Module (ESM) which provides enough electrical power, in the event of a power failure, to move the data in the DRAM to the local flash.

The flash modules used in Oracle Exadata X3 systems have an expected endurance of 10 years or more, even in write intensive applications. Flash endurance is determined primarily by the total data written to flash across many years, as well as the type of data written. No application runs at maximum flash write IOPS for every second of every day for years. Applications also do many reads and have periods of high and low activity, such as day versus night, quarter close, end of a trading day, and so on. A very highly write intensive application might average 25 percent of the maximum flash write IOPS when measured over many months. Each Exadata X3 storage server has a total flash write endurance of over 50 PB for typical database data. In a full rack, if the application writes an average of 250,000 8K flash IOPS (25 percent of maximum writes) for 10 years, then it will write a total of 41 PB of data to each cell. This is less than the 50 PB per cell endurance.

If the ESM does not have sufficient charge, then the F20 PCIe card operates in fail-safe write-through mode, bypassing the DRAM memory and writing all data directly to flash. This results in reduced write performance, but there is no data loss. Exadata Storage Server generates an alert when the ESM capacity is insufficient, and the ESM should be replaced.

The charge capacity of the ESM degrades over time, and its life expectancy is inversely proportional to the operating temperature. The worst case life expectancy of the ESM in Oracle Exadata Rack is as follows:

Type of Exadata Storage Server	Lifetime
Exadata Storage Server with Sun Fire X4275 Servers	3 years
Exadata Storage Server with Sun Fire X4270 M2 Servers	4 years

#### Related Topics

- [Monitoring Hard Disk Controller Write-through Caching Mode](#)  
The hard disk controller on each Oracle Exadata Storage Server periodically performs a discharge and charge of the controller battery. During the operation, the write cache policy changes from write-back caching to write-through caching.

## 3.12.2 Replacing Flash ESM

If the charge capacity in the F20 PCIe ESM falls below the minimum threshold, then Oracle will replace the failed ESM modules at no extra charge, if the system is covered either by the Oracle Premier Support for Systems or occurs during the warranty period.

For customers with Premier Support for Systems, Oracle attempts to proactively replace the F20 PCIe ESM in the Oracle Exadata Rack before the end of the estimated lifetime, on a best efforts basis.

## 3.13 LED Status Descriptions

The LEDs on the Oracle Exadata Rack components help you identify the component that needs servicing.

- [Exadata Storage Server X7-2 and Later LEDs](#)
- [Exadata Storage Server X5-2 Servers](#)
- [Exadata Storage Server X4-2L Servers](#)
- [Exadata Storage Server X3-2 Servers](#)
- [Exadata Storage Server with Sun Fire X4270 M2 Servers](#)
- [Exadata Storage Server with Sun Fire X4275 Servers LEDs](#)
- [Sun Flash Accelerator F20 PCIe Card LEDs](#)

### 3.13.1 Exadata Storage Server X7-2 and Later LEDs

The following table describes the color codes of the LEDs on Exadata Storage Server X7-2 and later servers.

Table 3-1 Exadata Storage Server X7-2 and Later LED Status Descriptions

Component	LED Status
Fan module	<ul style="list-style-type: none"> <li>• Top Fan Fault LED is <b>off</b>: The system is powered on and the fan module is functioning correctly.</li> <li>• Top Fan Fault LED is <b>amber</b>: The fan module is faulty. The front and rear panel Fault-Service Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Power supply	<ul style="list-style-type: none"> <li>• Fault-Service Required LED is <b>amber</b>: The power supply is faulty. The front and rear panel Fault-Service Required LEDs are also lit if the system detects a power supply fault.</li> <li>• AC OK LED is <b>green</b>: Power supply can be removed during a hot-swap procedure.</li> </ul>
Server chassis	<ul style="list-style-type: none"> <li>• OK to Remove LED is <b>blue</b>: The storage device can be removed safely during a hot-swap procedure.</li> <li>• Fault-Service Required LED is <b>amber</b>: The system is running but the storage device is faulty. The front and rear panel Fault-Service Required LEDs are also lit if the system detects a storage device fault.</li> <li>• OK/Activity LED is <b>green</b>: Data is being read from or written to the storage device.</li> <li>• Do Not Service LED is <b>white/on</b>: Indicates that the storage server is required to remain online to preserve data availability. Do not restart or power off the storage server. Otherwise, the Oracle Automatic Storage Management (Oracle ASM) disk group may forcibly dismount and compromise data availability.  Typically, the Do Not Service LED lights in response to an issue with a partner storage server. However, during an Oracle ASM cluster rolling upgrade, the Do Not Service LED lights simultaneously on all participating storage servers. Furthermore, on affected grid disks, the <code>asmDeactivationOutcome</code> attribute contains the value: <code>Cannot deactivate because ASM is in rolling upgrade mode</code>.</li> <li>• Do Not Service LED is <b>off</b>: The storage server can be safely powered off for servicing.</li> </ul>

The following table describes the storage device status based on the LEDs.

**Table 3-2 Storage Device Status of Exadata Storage Server X7-2 and Later Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Fault-Service Required LED (Amber)	On	On	On
OK to Remove LED (Blue)	Off	On	On

### 3.13.2 Exadata Storage Server X5-2 Servers

[Table 3-3](#) describes the color codes of the LEDs on Exadata Storage Server X5-2 Servers.

**Table 3-3 LED Status Descriptions**

Component	LED Status
Exadata Storage Server X5-2 Servers fan module	<ul style="list-style-type: none"> <li>Fan Status LED is off: The system is powered on and the fan module is functioning correctly.</li> <li>Fan Status LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Exadata Storage Server X5-2 Servers power supply	<ul style="list-style-type: none"> <li>OK to Remove LED is green: The power supply can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> </ul>
Exadata Storage Server X5-2 Servers	<ul style="list-style-type: none"> <li>OK to Remove LED is blue: A storage drive can safely be removed during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>

[Table 3-4](#) describes the disk status based on the LEDs.

**Table 3-4 Disk Status of Exadata Storage Server X5-2 Servers Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	On	On	On
OK to Remove (Blue)	Off	On	On

### 3.13.3 Exadata Storage Server X4-2L Servers

[Table 3-5](#) describes the color codes of the LEDs on Exadata Storage Server X3-2 Servers.

**Table 3-5 LED Status Descriptions**

Component	LED Status
Exadata Storage Server X4-2L Servers fan module	<ul style="list-style-type: none"> <li>Fan Status LED is green: The system is powered on and the fan module is functioning correctly.</li> <li>Fan Status LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Exadata Storage Server X4-2L Servers power supply	<ul style="list-style-type: none"> <li>OK to Remove LED is green: The power supply can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> </ul>
Exadata Storage Server X4-2L Servers	<ul style="list-style-type: none"> <li>OK to Remove LED is blue: A storage drive can safely be removed during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>

[Table 3-8](#) describes the disk status based on the LEDs.

**Table 3-6 Disk Status of Exadata Storage Server X4-2L Servers Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	On	On	On
OK to Remove (Blue)	Off	On	On

### 3.13.4 Exadata Storage Server X3-2 Servers

[Table 3-7](#) describes the color codes of the LEDs on Exadata Storage Server X3-2 Servers.

**Table 3-7 LED Status Descriptions**

Component	LED Status
Exadata Storage Server X3-2 Servers fan module	<ul style="list-style-type: none"> <li>Fan Status LED is green: The system is powered on and the fan module is functioning correctly.</li> <li>Fan Status LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Exadata Storage Server X3-2 Servers power supply	<ul style="list-style-type: none"> <li>OK to Remove LED is green: The power supply can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> </ul>
Exadata Storage Server X3-2 Servers	<ul style="list-style-type: none"> <li>OK to Remove LED is blue: A storage drive can safely be removed during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>

[Table 3-8](#) describes the disk status based on the LEDs.

**Table 3-8 Disk Status of Exadata Storage Server X3-2 Servers Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	On	On	On
OK to Remove (Blue)	Off	On	On

### 3.13.5 Exadata Storage Server with Sun Fire X4270 M2 Servers

[Table 3-9](#) describes the color codes of the LEDs on Exadata Storage Server with Sun Fire X4270 M2 Servers.

**Table 3-9 LED Status Descriptions**

Component	LED Status
Exadata Storage Server with Sun Fire X4270 M2 Servers fan module	<ul style="list-style-type: none"> <li>Fan Status LED is green: The system is powered on and the fan module is functioning correctly.</li> <li>Fan Status LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Exadata Storage Server with Sun Fire X4270 M2 Servers power supply	<ul style="list-style-type: none"> <li>OK to Remove LED is green: The power supply can be removed safely during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> </ul>
Exadata Storage Server with Sun Fire X4270 M2 Servers	<ul style="list-style-type: none"> <li>OK to Remove LED is blue: A storage drive can safely be removed during a hot-swap procedure.</li> <li>Service Action Required LED is amber: The storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>

Table 3-10 describes the disk status based on the LEDs.

**Table 3-10 Disk Status of Exadata Storage Server with Sun Fire X4270 M2 Servers Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	On	On	On
OK to Remove (Blue)	Off	On	On

### 3.13.6 Exadata Storage Server with Sun Fire X4275 Servers LEDs

Table 3-11 describes the color codes of the LEDs on Exadata Storage Server with Sun Fire X4275 Servers.

**Table 3-11 Exadata Storage Server with Sun Fire X4275 Servers LED Status Descriptions**

Component	LED Status
Exadata Storage Server with Sun Fire X4275 Servers	<ul style="list-style-type: none"> <li>• OK to Remove LED is blue: A storage drive can safely be removed during a hot-swap procedure.</li> <li>• Service Action Required LED is amber: The storage drive is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a storage drive fault.</li> <li>• OK/Activity LED is green: Data is being read from or written to the storage drive.</li> </ul>
Exadata Storage Server with Sun Fire X4275 Servers fan module	<ul style="list-style-type: none"> <li>• Power/OK LED is green: The system is powered on and the fan module is functioning correctly.</li> <li>• Service Action Required LED is amber: The fan module is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a fan module fault.</li> </ul>
Exadata Storage Server with Sun Fire X4275 Servers power supply	<ul style="list-style-type: none"> <li>• AC Present LED is green: Power supply can be removed during a hot-swap procedure.</li> <li>• Service Action Required LED is amber: The power supply is faulty. The front and rear panel Service Action Required LEDs are also lit if the system detects a power supply fault.</li> <li>• OK to Remove LED is blue: The power supply can safely be removed during a hot-swap procedure.</li> </ul>

Table 3-12 describes the disk status based on the LEDs.

**Table 3-12 Disk Status of Exadata Storage Server with Sun Fire X4275 Servers Based on LEDs**

LED	Predictive Failure	Poor Performance	Critical
Service Action Required (Amber)	Flashing	Flashing	On
OK to Remove (Blue)	Off	Flashing	Flashing

### 3.13.7 Sun Flash Accelerator F20 PCIe Card LEDs

Table 3-13 describes the color codes of the LEDs on Sun Flash Accelerator F20 PCIe cards.

**Table 3-13 Sun Flash Accelerator F20 PCIe Card LED Status Descriptions**

Component	LED Status
FMod 3, 2, 1, 0	LED is green: It flashes when there is read/write activity in the corresponding FMod. It is on when the corresponding FMod is present.



**Table 3-13 (Cont.) Sun Flash Accelerator F20 PCIe Card LED Status Descriptions**

Component	LED Status
ESM OK LED	LED is green: It is on when the ESM is adequately charged and capable of providing enough emergency power to retain data integrity in the event of a power interruption. It is off when the system is down, and it flashes when the ESM is charging.
ESM Service Required LED	LED is amber indicates the ESM is not retaining enough reserve power to support write-back mode. The card is functioning in write-through (caching) mode. Service is required.
Power LED	LED is on: Indicates the card is on.

## 3.14 Exadata Storage Server Images

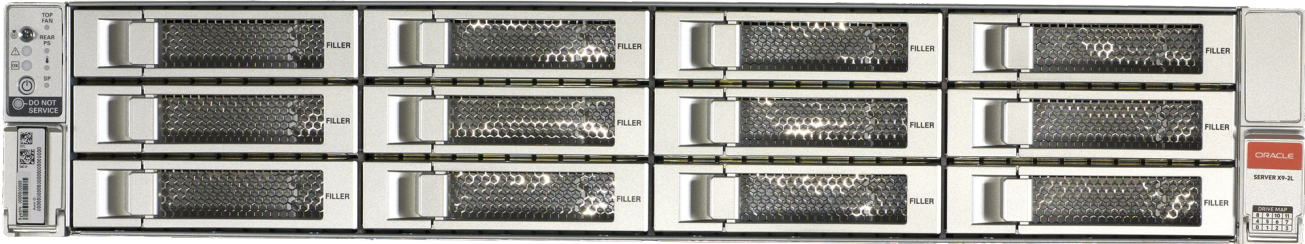
The Exadata Storage Server models have different external layouts and physical appearance.

- [Oracle Exadata Storage Server X9M-2 Extreme Flash Server Images](#)
- [Oracle Exadata Storage Server X9M-2 High Capacity Server Images](#)
- [Oracle Exadata Storage Server X9M-2 Extended Server Images](#)
- [Oracle Exadata Storage Server X8M-2 and X8-2 High Capacity and Extended \(XT\) Server Images](#)
- [Oracle Exadata Storage Server X8M-2 and X8-2 Extreme Flash Server Images](#)
- [Oracle Exadata Storage Server X7-2 High Capacity Server Images](#)
- [Oracle Exadata Storage Server X7-2 Extreme Flash Server Images](#)
- [High Capacity Exadata Storage Server X6-2 Images](#)
- [Extreme Flash Exadata Storage Server X6-2 Images](#)
- [High Capacity Exadata Storage Server X5-2 Images](#)
- [Extreme Flash Exadata Storage Server X5-2 Images](#)
- [Exadata Storage Server X4-2L Images](#)
- [Exadata Storage Server X3-2L Images](#)
- [Exadata Storage Server with Sun Fire X4270 M2 Images](#)
- [Exadata Storage Server with Sun Fire X4275 Images](#)

### 3.14.1 Oracle Exadata Storage Server X9M-2 Extreme Flash Server Images

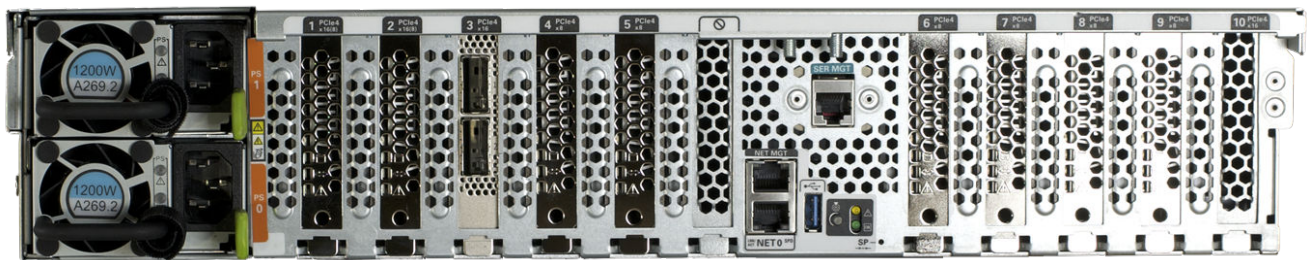
The following figure shows the front view of the Oracle Exadata Storage Server X9M-2 Extreme Flash (EF) servers.

**Figure 3-1 Front View of Oracle Exadata Storage Server X9M-2 Extreme Flash Server**



The following figure shows the rear view of the Oracle Exadata Storage Server X9M-2 Extreme Flash servers.

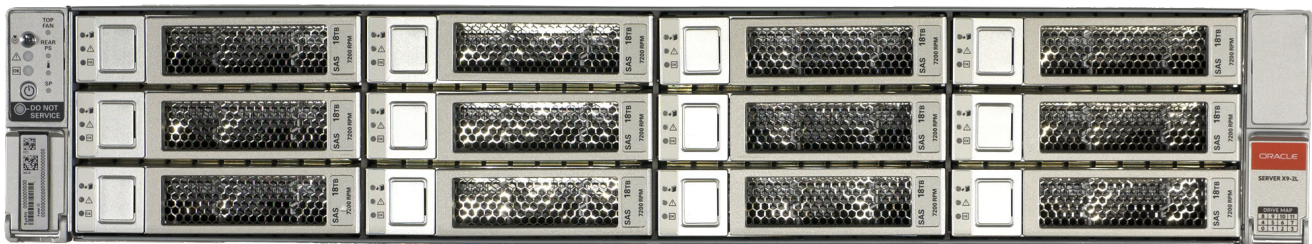
**Figure 3-2 Rear View of Oracle Exadata Storage Server X9M-2 Extreme Flash Server**



### 3.14.2 Oracle Exadata Storage Server X9M-2 High Capacity Server Images

The following figure shows the front view of the Oracle Exadata Storage Server X9M-2 High Capacity (HC) servers.

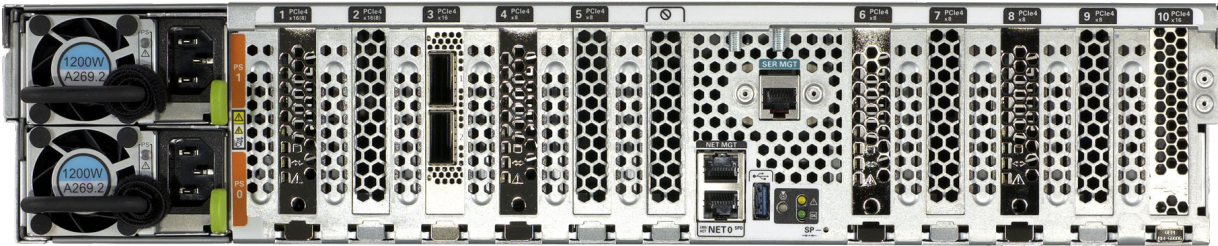
**Figure 3-3 Front View of Oracle Exadata Storage Server X9M-2 High Capacity servers**



The following figure shows the rear view of the Oracle Exadata Storage Server X9M-2 High Capacity servers.



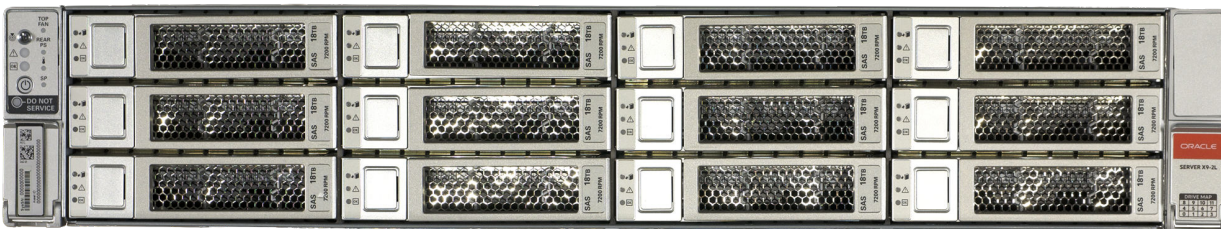
**Figure 3-4 Rear View of Oracle Exadata Storage Server X9M-2 High Capacity servers**



### 3.14.3 Oracle Exadata Storage Server X9M-2 Extended Server Images

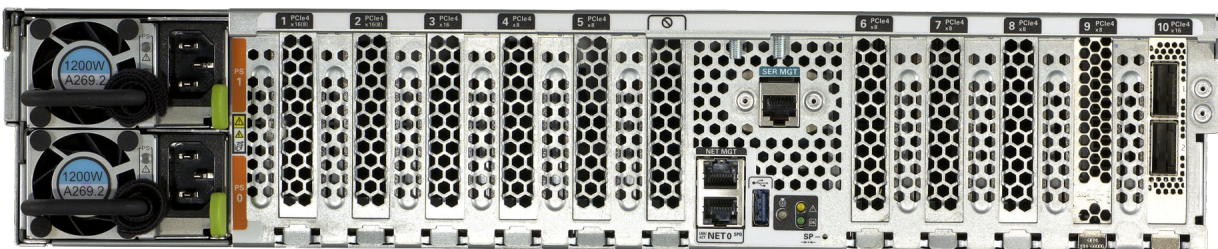
The following figure shows the front view of the Oracle Exadata Storage Server X9M-2 Extended (XT) servers.

**Figure 3-5 Front View of Oracle Exadata Storage Server X9M-2 Extended servers**



The following figure shows the rear view of the Oracle Exadata Storage Server X9M-2 Extended servers.

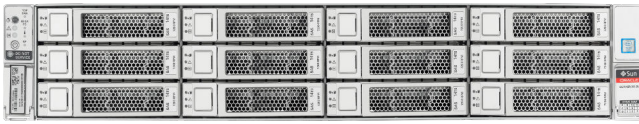
**Figure 3-6 Rear View of Oracle Exadata Storage Server X9M-2 Extended servers**



### 3.14.4 Oracle Exadata Storage Server X8M-2 and X8-2 High Capacity and Extended (XT) Server Images

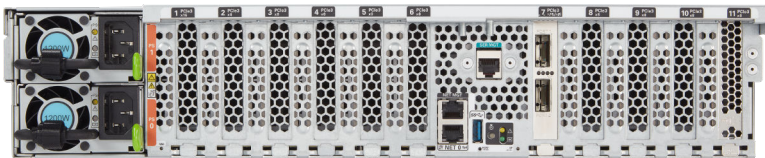
The following figure shows the front view of the Oracle Exadata Storage Server X8M-2 and X8-2 High Capacity and XT servers.

**Figure 3-7 Front View of Oracle Exadata Storage Server X8M-2 and X8-2 High Capacity and XT servers**



The following figure shows the rear view of the Oracle Exadata Storage Server X8M-2 and X8-2 High Capacity and XT servers.

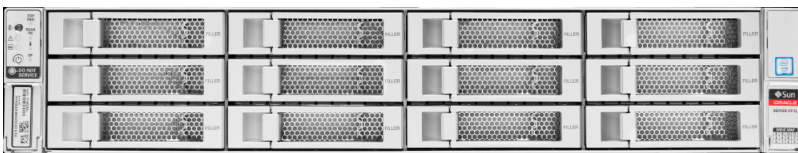
**Figure 3-8 Rear View of Oracle Exadata Storage Server X8M-2 and X8-2 High Capacity and XT servers**



### 3.14.5 Oracle Exadata Storage Server X8M-2 and X8-2 Extreme Flash Server Images

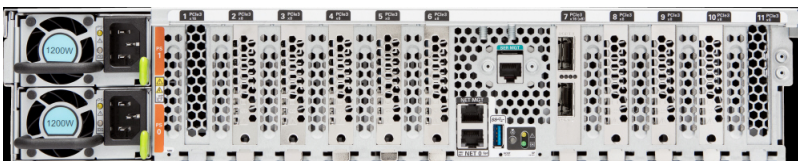
The front view of the Oracle Exadata Storage Server X8M-2 and X8-2 Extreme Flash server is almost identical to the X7-2 server. The main difference is the product logo.

**Figure 3-9 Front View of Oracle Exadata Storage Server X8M-2 and X8-2 Extreme Flash Server**



The rear view of the Oracle Exadata Storage Server X8M-2 and X8-2 Extreme Flash server is almost identical to the X7-2 server. The main difference is the product logo.

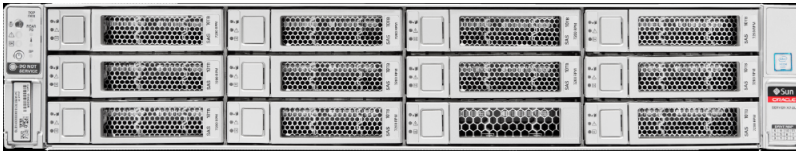
**Figure 3-10 Rear View of Oracle Exadata Storage Server X8M-2 and X8-2 Extreme Flash Server**



### 3.14.6 Oracle Exadata Storage Server X7-2 High Capacity Server Images

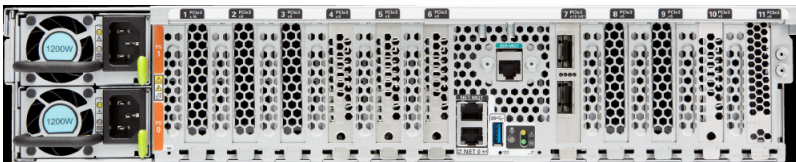
The following figure shows the front view of the Oracle Exadata Storage Server X7-2 High Capacity Server.

**Figure 3-11 Front View of Oracle Exadata Storage Server X7-2 High Capacity Server**



The following figure shows the rear view of the Oracle Exadata Storage Server X7-2 High Capacity Server.

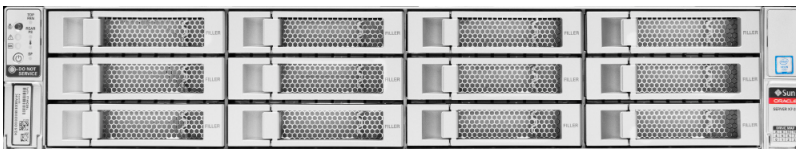
**Figure 3-12 Rear View of Oracle Exadata Storage Server X7-2 High Capacity Server**



### 3.14.7 Oracle Exadata Storage Server X7-2 Extreme Flash Server Images

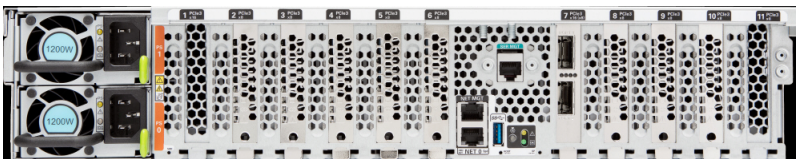
The following figure shows the front view of the Oracle Exadata Storage Server X7-2 Extreme Flash server.

**Figure 3-13 Front View of Oracle Exadata Storage Server X7-2 Extreme Flash Server**



The following figure shows the rear view of the Oracle Exadata Storage Server X7-2 Extreme Flash server.

**Figure 3-14 Rear View of Oracle Exadata Storage Server X7-2 Extreme Flash Server**





### 3.14.8 High Capacity Exadata Storage Server X6-2 Images

The following figure shows the front view of the Oracle Exadata Storage Server X6-2 High Capacity server.

**Figure 3-15 Front View of Oracle Exadata Storage Server X6-2 High Capacity Server**



The following figure shows the rear view of the Oracle Exadata Storage Server X6-2 High Capacity server.

**Figure 3-16 Rear View of Oracle Exadata Storage Server X6-2 High Capacity Server**



### 3.14.9 Extreme Flash Exadata Storage Server X6-2 Images

The following figure shows the front view of the Oracle Exadata Storage Server X6-2 Extreme Flash server.

**Figure 3-17 Front View of Oracle Exadata Storage Server X6-2 Extreme Flash Server**



The following figure shows the rear view of the Oracle Exadata Storage Server X6-2 Extreme Flash server.

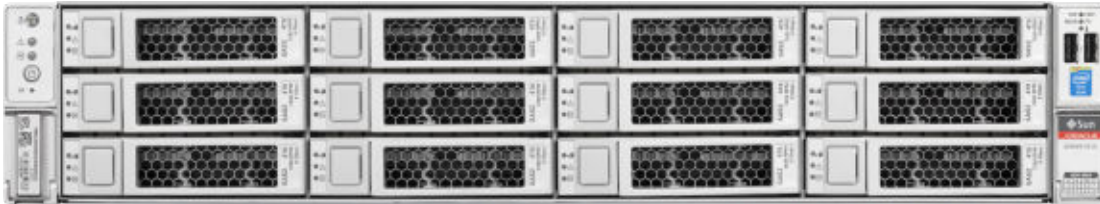
**Figure 3-18 Rear View of Oracle Exadata Storage Server X6-2 Extreme Flash Server**



### 3.14.10 High Capacity Exadata Storage Server X5-2 Images

The following image shows the front view of High Capacity Exadata Storage Server X5-2 Servers.

**Figure 3-19 Front View of High Capacity Exadata Storage Server X5-2 Servers**



The following image shows the rear view of High Capacity Exadata Storage Server X5-2 Servers.

**Figure 3-20 Rear View of High Capacity Exadata Storage Server X5-2 Servers**



### 3.14.11 Extreme Flash Exadata Storage Server X5-2 Images

The following image shows the front view of Extreme Flash Exadata Storage Server X5-2 Servers.

**Figure 3-21 Front View of Extreme Flash Exadata Storage Server X5-2 Servers**



The following image shows the rear view of Extreme Flash Exadata Storage Server X5-2 Servers.

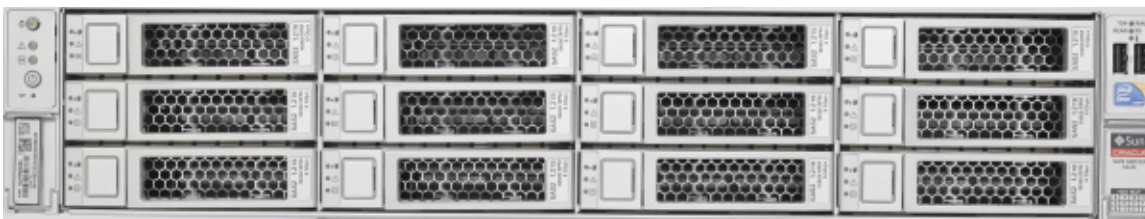
**Figure 3-22 Rear View of Extreme Flash Exadata Storage Server X5-2 Servers**



### 3.14.12 Exadata Storage Server X4-2L Images

The following image shows the front view of Exadata Storage Server X4-2L Servers. The hard drives are numbered from left to right, starting in the lower left. The drives in the bottom row are numbers 0, 1, 2, and 3. The drives in the middle row are numbers 4, 5, 6, and 7. The drives in the top row are numbers 8, 9, 10, and 11.

**Figure 3-23 Front View of Exadata Storage Server X4-2L Servers**



The following image shows the rear view of Exadata Storage Server X4-2L Servers.

**Figure 3-24 Rear View of Exadata Storage Server X4-2L Servers**

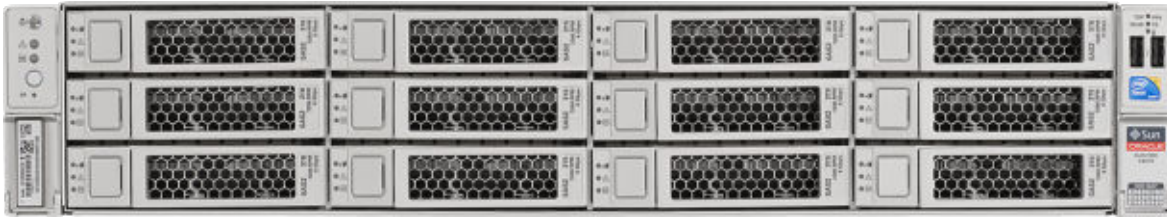


### 3.14.13 Exadata Storage Server X3-2L Images

The following image shows the front view of Exadata Storage Server X3-2L Servers. The hard drives are numbered from left to right, starting in the lower left. The drives in the bottom row are numbers 0, 1, 2, and 3. The drives in the middle row are numbers 4, 5, 6, and 7. The drives in the top row are numbers 8, 9, 10, and 11.



**Figure 3-25 Front View of Exadata Storage Server X3-2L Servers**



The following image shows the rear view of Exadata Storage Server X3-2L Servers.

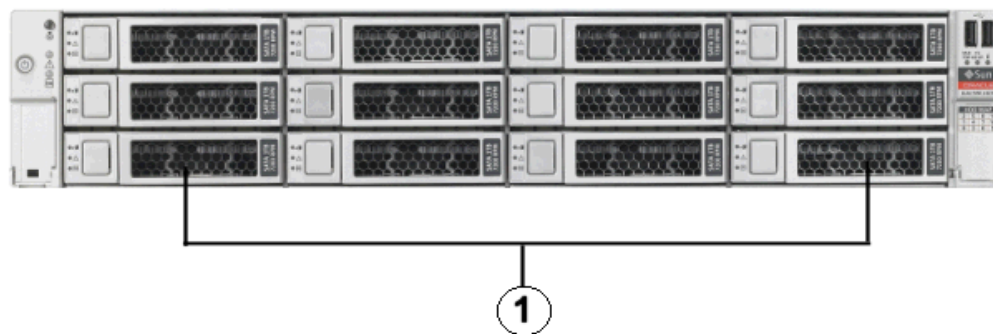
**Figure 3-26 Rear View of Exadata Storage Server X3-2L Servers**



### 3.14.14 Exadata Storage Server with Sun Fire X4270 M2 Images

The following image shows the front view of Exadata Storage Server with Sun Fire X4270 M2 Servers.

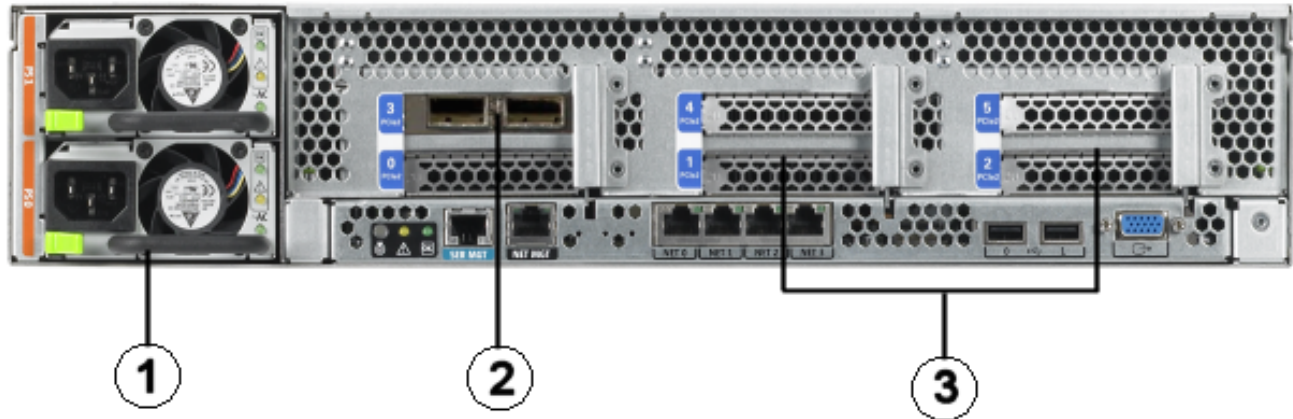
**Figure 3-27 Front View of Exadata Storage Server with Sun Fire X4270 M2 Servers**



1. Hard disk drives. The top drives are, from left to right, HDD2, HDD5, HDD8, and HDD11. The middle drives are, from left to right, HDD1, HDD4, HDD7, and HDD10. The bottom drives are, from left to right, HDD0, HDD3, HDD6, and HDD9.

The following image shows the rear view of Exadata Storage Server with Sun Fire X4270 M2 Servers.

Figure 3-28 Rear View of Exadata Storage Server with Sun Fire X4270 M2 Servers

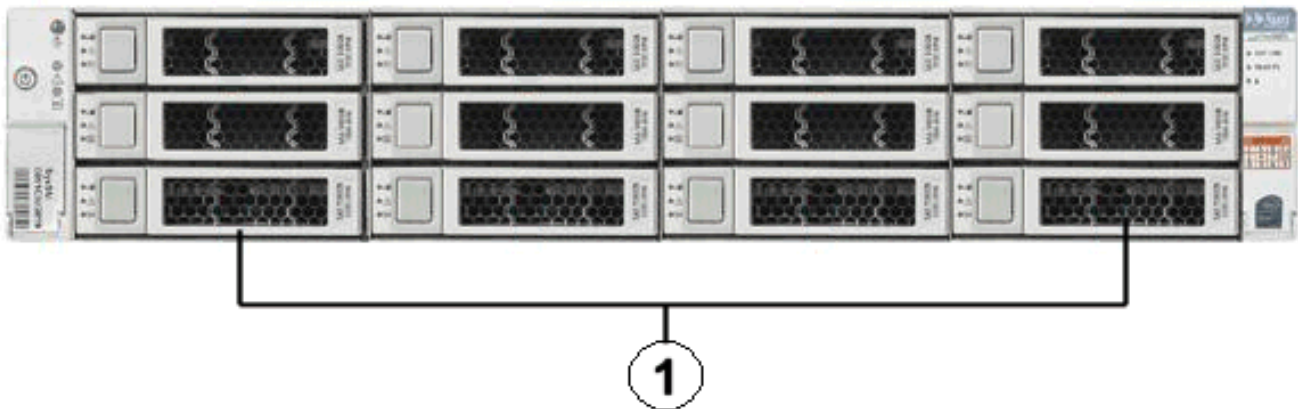


1. Power supplies.
2. InfiniBand host channel adapter PCI Express module.
3. Sun Flash Accelerator F20 PCIe Cards.

### 3.14.15 Exadata Storage Server with Sun Fire X4275 Images

The following figure shows the front view of Sun Fire X4275 Servers.

Figure 3-29 Front View of Sun Fire X4275 Server



1. Hard disk drives. The top drives are, from left to right, HDD2, HDD5, HDD8, and HDD11. The middle drives are, from left to right, HDD1, HDD4, HDD7, and HDD10. The bottom drives are, from left to right, HDD0, HDD3, HDD6, and HDD9.

The following figure shows the rear view of Sun Fire X4275 servers.

Figure 3-30 Rear View of Sun Fire X4275 Server



# 4

## Maintaining Other Exadata Database Machine Components

Besides the database and storage servers, Exadata Database Machine contains other components, such as power distribution units, ILOM, and network switches.

### Note:

- All procedures in this chapter are applicable to Exadata Database Machine and Oracle Exadata Storage Expansion Rack.
- For ease of reading, the name "Oracle Exadata Rack" is used when information refers to both Exadata Database Machine and Oracle Exadata Storage Expansion Rack.

- [Replacing a Power Distribution Unit](#)
- [Resetting a Non-Responsive ILOM](#)
- [Configuring Service Processor and ILOM Network Settings](#)
- [Verifying and Modifying the Link Speed on the Client Network Ports for X7 and Later Systems](#)  
You can configure 10 GbE connections or 25 GbE connections on the client network on Exadata Database Machine X7 and later database servers.
- [Verify the Link Speed on Network Ports](#)  
Ensure you are using the correct link speed for Oracle Exadata Database Machine X7-2 compute nodes.
- [Changing from 1 GbE Connections to 10 GbE Connections](#)  
1 GbE network connections can be changed to 10 GbE connections.
- [Maintaining the RoCE Network Fabric](#)  
The RoCE Network Fabric connects the database servers and Exadata Storage Servers through the bonded interface to the RoCE Network Fabric switches in the rack.
- [Maintaining the InfiniBand Network Fabric Network](#)  
The InfiniBand Network Fabric network connects the database servers and Exadata Storage Servers through the bonded interface to the InfiniBand Network Fabric switches in the rack.
- [Modifying the InfiniBand Network Fabric Configuration](#)  
You can change how the InfiniBand Network Fabric network is configured by changing the IP addresses or host names, or by implementing partitioning.
- [Configuring Network Routing on Database Servers](#)  
The tasks for network routing are for boot-time routing or real-time routing.

- [Changing the DNS Servers](#)  
The configuration settings for the Domain Name System (DNS) servers can be changed after initial setup.
- [Changing the NTP Servers](#)  
The configuration settings for the Network Time Protocol (NTP) servers can be changed after initial setup.
- [Changing the Time Zone Settings](#)  
You can change the time zones on Exadata Database Machine after initial configuration and deployment.
- [Managing the KVM Switch](#)  
The KVM switch is only available in Oracle Exadata Database Machine X2-2 racks and Oracle Exadata Storage Expansion Racks with Exadata Storage Server with Sun Fire X4270 M2 Servers.
- [LED Status Descriptions](#)  
The LEDs on the Oracle Exadata Rack components help you identify the component that needs servicing.

## 4.1 Replacing a Power Distribution Unit

Power distribution units (PDUs) can be replaced while Oracle Exadata Rack is online. PDU-A is on the left, and PDU-B is on the right when viewing the rack from the rear.

- [Reviewing the PDU Replacement Guidelines](#)  
Before replacing a PDU, review the following guidelines to ensure the procedure is safe and does not disrupt availability.
- [Replacing a PDU](#)  
This procedure describes how to replace a power distribution unit (PDU).

### 4.1.1 Reviewing the PDU Replacement Guidelines

Before replacing a PDU, review the following guidelines to ensure the procedure is safe and does not disrupt availability.

- Unlatching the RDMA Network Fabric cables while removing or inserting PDU-A may cause a loss of service due to nodes being removed from the cluster. This could cause the rack to be unavailable. Care should be taken when handling the RDMA Network Fabric cables, which are normally latched securely. Do not place excessive tension on the RDMA Network Fabric cables by pulling them.
- Unhooking the wrong power feeds causes the rack to shut down. Trace the power cables running from the PDU that will be replaced to the power source, and only unplug those feeds.
- Allow time to unpack and repack the PDU replacement parts. Note how the power cords are coiled in the packaging so the failed unit can be repacked the same way.
- Removal of the side panel lessens the amount of time needed to replace the PDU. However, it is not necessary to remove the side panel to replace the PDU.
- Use of a cordless drill or power screwdriver lessens the amount of time needed to replace the PDU. Allow more time for the replacement if using the hand wrench tool provided with the replacement rack. If using a screwdriver, then ensure that there are Torx T30 and T25 bits.

- It may be necessary to remove the server cable arms to move the power cables. If that is the case, then twist the plug connection and flex the cable arm connector to avoid having to unclip the cable arm. If it is necessary to unclip the cable arm, then support the cables with one hand, remove the power cord, and then clip the cable arm. Do not leave the cable arm hanging.
- When removing the T30 screws from the L-bracket, do not remove the T25 screws or nuts that attach the PDU to the bracket until the PDU is out of the rack.

## 4.1.2 Replacing a PDU

This procedure describes how to replace a power distribution unit (PDU).

1. Use the PDU monitor as follows to identify its network settings, if it is not the reason for the PDU replacement:
  - a. Press the reset button for 20 seconds until it starts to count from 5 to 0. While it is counting down, release the button, and then press it once.
  - b. Record the network settings, firmware version, and so on, displayed on the LCD screen as the monitor restarts.

 **Note:**

If the PDU monitor is not working, then retrieve the network settings by connecting to the PDU over the network, or from the network administrator.

2. Turn off all the PDU breakers.
3. Unplug the PDU power plugs from the AC outlets.

 **Note:**

- If the power cords use overhead routing, then put the power plugs in a location where they will not fall or hit anyone.
- If the rack is on a raised floor, then move the power cords out through the floor cutout. It may be necessary to maneuver the rack over the cutout in order to move the power cords out.

4. Do the following procedure for a PDU-B replacement when there is not side panel access, and the rack does not have an InfiniBand cable harness:

 **Note:**

Do not unstrap any cables attached to the cable arms.

- a. Unscrew the T25 screws holding the square cable arms to the rack.
  - b. Move the RDMA Network Fabric cables to the middle, out of the way.
5. Unplug all power cables going from the servers and switches to the PDU. Keep the power cables together in group bundles.



6. Remove the T30 screws from the top and bottom of the L-bracket, and note where the screws go.
7. Note where the PDU sits in the rack frame. It is usually 1 inch back from the rack frame to allow access to the breaker switches.
8. Angle and maneuver the PDU out of the rack.
9. Hold the PDU or lay it down, if there is enough room, while maneuvering the AC power cords through the rack. It may be necessary to cut the cable ties that hold the AC cord flush with the bottom side of the PDU.
10. Pull the cords as near to the bottom or top of the rack as possible where there is more room between the servers to get the outlet plug through the routing hole.
11. Remove the smaller Torx T25 screws, and loosen the nut on the top and bottom to remove the PDU from the L-bracket. The nut does not have to be removed.
12. Attach the L-bracket to the new PDU.
13. Lay the new PDU next to the rack.
14. Route the AC cords through the rack, and to where the outlets are.

 **Note:**

Do not cable tie the AC cord to the new PDU at this time.

15. Place the new PDU in the rack by angling and maneuvering it until the L-brackets sit on the top and bottom rails.
16. Line up the holes and slots so that the PDU sits about 1 inch back from the rack frame.
17. Attach the power cords using the labels on the cords as a guide.  
For example, G5-0 indicates PDU group 5 outlet 0 on the PDU.
18. Attach the InfiniBand cable holders if they were removed in step 4.  
Oracle recommends screwing the holders in by hand at first to avoid stripping the screws.
19. Attach the AC power cords to the outlets.
20. Turn on the breakers.
21. Cable and program the PDU monitor for the network, as needed.

**Related Topics**

- [Sun Rack II Power Distribution Units User's Guide](#)
- [Oracle Rack Cabinet 1242 User's Guide](#)
- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

## 4.2 Resetting a Non-Responsive ILOM

When Oracle Exadata System Software detects that the ILOM is unresponsive, it automatically resets the ILOM Service Processor.

Also, as a proactive measure, the ILOM is reset automatically every 90 days. To help predict the next automatic reset, you can retrieve the ILOM up-time by querying the ILOM directly or using `ipmitool`. For example:

- Using ILOM:

```
-> show /SP/clock uptime

/SP/clock
Properties:
uptime = 54 days, 15:41:51
```

- Using `ipmitool`:

```
ipmitool sunoem getval /SP/clock/uptime
Target Value: 54 days, 15:41:51
```

You can also manually reset the ILOM Service Processor using various methods:

- [Resetting the ILOM Using SSH](#)
- [Resetting the ILOM Using the ILOM Remote Console](#)
- [Resetting the ILOM Using IPMItool](#)
- [Resetting the ILOM Using the SP Reset Pin on Oracle Exadata Database Machine X2-2 Servers and Exadata Storage Servers](#)
- [Removing the SP from Sun Fire X4800 Oracle Database Servers and Sun Server X2-8 Oracle Database Servers](#)  
If you could not reset the ILOM on the Sun Fire X4800 Oracle Database Server or Sun Server X2-8 Oracle Database Server using SSH, the remote console or IPMItool, then remove the service processor (SP) from the server, and put it back.
- [Unplugging the ILOM Power Supply](#)



#### See Also:

Oracle Integrated Lights Out Manager (ILOM) Documentation at <http://www.oracle.com/goto/ilom/docs>

## 4.2.1 Resetting the ILOM Using SSH

The following procedure describes how to reset the ILOM by connecting to it using SSH:

1. Connect to the ILOM using SSH from another machine.
2. Enter the following command at the ILOM prompt:

```
reset /SP
```

## 4.2.2 Resetting the ILOM Using the ILOM Remote Console

If it is not possible to connect to the ILOM using SSH, then log in to the ILOM remote console. The following procedure describes how to reset the ILOM using the remote console.



1. Log in to the ILOM remote console.
2. Select **Reset SP** from the Maintenance tab.
3. Click **Reset SP**.

### 4.2.3 Resetting the ILOM Using IPMItool

If you could not connect to the ILOM using SSH or the remote console, then log in to the local host or another host on the ILOM network, and use IPMItool. The following procedure describes how to reset the ILOM using IPMItool:

1. Log in to local host or another host on the ILOM network.
2. Run the following IPMItool command:

- Using local host:

```
$ ipmitool mc reset cold
Sent cold reset command to MC
```

- Using another host:

```
$ ipmitool -H ILOM_host_name -U ILOM_user mc reset cold
Sent cold reset command to MC
```

In the preceding command, *ILOM\_host\_name* is the host name being used, and *ILOM\_user* is the user name for the ILOM.

### 4.2.4 Resetting the ILOM Using the SP Reset Pin on Oracle Exadata Database Machine X2-2 Servers and Exadata Storage Servers

If you could not connect to the ILOM using SSH, the remote console, or IPMItool on the Oracle Exadata Database Machine X2-2 server or Exadata Storage Server, then press the SP reset pin. The following procedure describes how to reset the ILOM using the SP reset pin.

1. Obtain a small, non-conductive stick.
2. Go to the rear of the rack.
3. Locate the SP reset pin opening. The SP reset pin opening is the first opening to the right of the NET MGT port.
4. Insert the stick into the opening and press the pin.

### 4.2.5 Removing the SP from Sun Fire X4800 Oracle Database Servers and Sun Server X2-8 Oracle Database Servers

If you could not reset the ILOM on the Sun Fire X4800 Oracle Database Server or Sun Server X2-8 Oracle Database Server using SSH, the remote console or IPMItool, then remove the service processor (SP) from the server, and put it back.

Messages are displayed at the operating system level. These messages can be ignored. The fans will speed up because there is no fan control.

- See "Removing and Installing the Service Processor Module (CRU)" in *Sun Fire X4800 Server Service Manual* at <http://docs.oracle.com/cd/E19140-01/html/821-0282/gjfv.html#scrolltoc>

## 4.2.6 Unplugging the ILOM Power Supply

If you could not reset to the ILOM using the preceding options, then unplug the power supply, and then plug it back in. This action power cycles the server as well as the ILOM.

## 4.3 Configuring Service Processor and ILOM Network Settings

The following procedure describes how to configure the service processor (SP) and ILOM network settings:

1. Log in to the SP as the `root` user using SSH.
2. Use the `version` command to check the SP/ILOM firmware release. The following is an example of the output from the command:

```
-> version
SP firmware 3.2.4.10
SP firmware build number: 93199
SP firmware date: Sat Oct 4 18:42:56 EDT 2014
SP filesystem version: 0.2.10
```

### Note:

The `ipmitool` can be used to log into the server SP. This is useful when the SP/ILOM is not accessible from the management network. The following command is used to connect to the SP:

```
ipmitool sunoem cli
Connected. Use ^D to exit.
-> version
SP firmware 3.2.4.10
SP firmware build number: 93199
SP firmware date: Sat Oct 4 18:42:56 EDT 2014
SP filesystem version: 0.2.10
```

3. Configure the DNS server settings using the `set` command as follows:

```
cd /SP/clients/dns/
/SP/clients/dns
show
/SP/clients/dns
Targets:
Properties:
 auto_dns = enabled
 nameserver = 0.0.0.0
 retries = 1
 searchpath =
 timeout = 5
Commands:
 cd
 set
 show
set nameserver=192.68.0.2
set searchpath=yourdomain.com
```

4. Configure the NTP server settings using the `set` command as follows.

```

cd /SP/clients/ntp/server/1/
/SP/clients/ntp/server/1
show
 /SP/clients/ntp/server/1
 Targets:
 Properties:
 address = 0.0.0.0
 Commands:
 cd
 set
 show
set address=192.68.0.1

```

 **Note:**

Two NTP servers can be configured. Set the first NTP server using the `set` command, and then use the path `SP/clients/ntp/server/2` to configure the second server.

5. Use the `set` command to configure the network settings as follows:

```

cd /SP/network
/SP/network
show
 /SP/network
 Targets:
 interconnect
 ipv6
 test
 Properties:
 commitpending = (Cannot show property)
 dhcp_clientid = none
 dhcp_server_ip = none
 ipaddress = 0.0.0.0
 ipdiscovery = dhcp
 ipgateway = 0.0.0.0
 ipnetmask = 0.0.0.0
 managementport = MGMT
 pendingipaddress = 0.0.0.0
 pendingipdiscovery = dhcp
 pendingipgateway = 0.0.0.0
 pendingipnetmask = 0.0.0.0
 pendingmanagementport = MGMT
 pendingvlan_id = (none)
 state = enabled
 vlan_id = (none)
 Commands:
 cd
 set
 show

```

6. Configure the corresponding `pendingip*` settings for the `ipaddress`, `ipdiscovery`, `ipgateway`, `ipnetmask`, and `vlan_id`, and then commit the pending settings using the following command:

```
set commitpending=true
```

7. Disconnect from the command line interface after the network configuration is complete.

 **Note:**

Use `^D` to exit the session when using the `ipmitool`.

## 4.4 Verifying and Modifying the Link Speed on the Client Network Ports for X7 and Later Systems

You can configure 10 GbE connections or 25 GbE connections on the client network on Exadata Database Machine X7 and later database servers.

 **Note:**

You should configure the client network ports using Oracle Exadata Deployment Assistant (OEDA) during system deployment. See *Using Oracle Exadata Deployment Assistant*.

The following steps may be necessary to configure a client access port if the OEDA deployment was not performed or was performed incorrectly. You can also use these steps to change the client network from 10 GbE to 25 GbE, or from 25 GbE to 10 GbE.

1. For each network interface (designated by `x`) that does not have the link detected, run the following commands:

- For 10GbE network interfaces:

```
ifdown ethx
ethtool -s ethx 10000 duplex full autoneg off
ifup ethx
ethtool ethx
```

For 10 Gb/s, you must use SFP+ transceivers; SFP28 transceivers do not support 10 Gb/s traffic.

- For 25GbE network interfaces:

```
ifdown ethx
ethtool -s ethx 25000 duplex full autoneg off
ifup ethx
ethtool ethx
```

2. Confirm that the output from the `ethtool` command shows `yes` for Link detected.

**Link detected: yes**

3. Edit the appropriate files in `/etc/sysconfig/network-scripts`, where `x` is the number associated with the network interface.
  - a. Locate the `/etc/sysconfig/network-scripts/ifcfg-ethx` file. Add the following lines, if they are not already present in the file:

- For 10 GbE network interfaces:

```
ONBOOT=YES
ETHTOOL_OPTS="speed 10000 duplex full autoneg off"
```

- For 25 GbE network interfaces:

```
ONBOOT=YES
ETHTOOL_OPTS="speed 25000 duplex full autoneg off"
```

- Repeat the previous step for all network interfaces that do not have the `ETHTOOL_OPTS` setting in the associated `ifcfg-ethx` file and are connected to 10 GbE or 25 GbE switches.

The network interface should now show the link as detected. These changes are persistent, and do not need to be repeated after a server reboot.

4. Check the ILOM on each compute node to validate the LAN on Motherboard is properly configured to detect the 25 GbE transceiver.

```
show /HOST/network
/HOST/network
 Targets:

 Properties:
 active_media = none
 auto_media_detection = enabled
 current_active_media = (none)

 Commands:
 cd
 set
 show
```

If the NIC is not working, change the `active_media` and `current_active_media` to the proper values:

- For 25 GbE transceivers (Fiber or Copper) these parameters should be set to `SPF28`
- For 10 GbE network using RJ-45 ended CAT6 cables, these parameters should be set to `RJ45`

## 4.5 Verify the Link Speed on Network Ports

Ensure you are using the correct link speed for Oracle Exadata Database Machine X7-2 compute nodes.

On Oracle Exadata Database Machine X7-2 compute nodes you might experience issues when connecting to 10GbE switches. These issues include links not being detected or being unable to connect to the gateway.

### Resolving 10GbE Network Speed Configuration on Client Network Ports

1. Log in as the `root` user.

2. Use the `cat` command to review the `/proc/net/bonding/bondeth0` file.
3. For each 10GbE network interface (designated by `x`) that does not have the link detected, run the following commands:

```
ifdown ethx
ethtool -s ethx 10000 duplex full autoneg off
ifup ethx
ethtool ethx
```

4. Confirm that the output from the `ethtool` command shows `yes` for Link detected.

**Link detected: yes**

5. Edit the appropriate files in `/etc/sysconfig/network-scripts`, where `x` is the number associated with the network interface.

- a. Locate the `/etc/sysconfig/network-scripts/ifcfg-ethx` file. Add the following line, if it is not already present in the file:

```
ETHTOOL_OPTS="speed 10000 duplex full autoneg off"
```

- b. Repeat the previous step for all network interfaces that do not have the `ETHTOOL_OPTS` setting in the associated `ifcfg-ethx` file and are connected to 10GbE switches.

The network interface should now show the link as detected. These changes are persistent, and do not need to be repeated after a server reboot.

## 4.6 Changing from 1 GbE Connections to 10 GbE Connections

1 GbE network connections can be changed to 10 GbE connections.

This procedure applies to Exadata Database Machine models X6 and earlier.

When changing the connections, note the following:

- To prevent a single point of failure for a bonded 10 GbE interface on Exadata Database Machine X2-8, use different ports on the Network Express Modules (NEMs) on the two cards, such as NEM0 NET1 and NEM1 NET0.
- The 10 GbE interfaces are identified as `eth4` and `eth5` on Sun Fire X4170 M2 Oracle Database Servers, and as `eth8` through `eth15` on Sun Fire X4800 Oracle Database Servers. Oracle recommends using following on Exadata Database Machine X2-8:
  - BONDETH0 using interfaces `eth9` and `eth15`
  - 10 GbE NEM0(left)/NET1
  - 10 GbE NEM1(right)/NET3
- Oracle Clusterware is shut down, and the database server is restarted during the procedure.
- [Task 1: Verify ping Functionality](#)
- [Task 2: Back up the Current Interface Files](#)
- [Task 3: Edit the 10 GbE Interface Settings](#)

- [Task 4: Verify the 10 GbE Interfaces](#)

## 4.6.1 Task 1: Verify ping Functionality

Verify the functionality of the `ping` command before any changes using the following commands. By verifying the `ping` command before any changes, you know what the results should be after changing the interfaces. Similar commands can be used to check other servers that connect to Oracle Exadata Database Machine.

```
grep "^nameserver" /etc/resolv.conf
nameserver ip_address_1
nameserver ip_address_2

ping -c 2 ip_address_1
PING ip_address_1 (ip_address_1) 56(84) bytes of data.
64 bytes from ip_address_1: icmp_seq=1 ttl=57 time=1.12 ms
64 bytes from ip_address_1: icmp_seq=2 ttl=57 time=1.05 ms

--- ip_address_1 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1002ms
rtt min/avg/max/mdev = 1.054/1.087/1.120/0.033 ms
```

If the test is not successful, showing 100% packet loss, then you should expect similar results when this same verification is run in "[Task 4: Verify the 10 GbE Interfaces](#)". If the test is successful, showing 0% packet loss, then you must see similar results after changing the 10 GbE connections.

## 4.6.2 Task 2: Back up the Current Interface Files

The following procedure describes how to back up the current interface files:

1. Log in as the `root` user.
2. Create hidden directories for the current and new 10 GbE files in the `/etc/sysconfig/network-scripts` directory similar to the following:

```
cd /etc/sysconfig/network-scripts
mkdir .Pre_10GigE_Settings
mkdir .Post_10GigE_Settings
```

### Note:

Linux startup scripts search for files that begin with `ifcfg`, and assume files beginning with `ifcfg` are used for network setup. Placing the backup files in hidden directories avoids them from being used to set up the network interface.

3. Identify the connected 10 GbE interfaces using the following command. Run the command for each 10 GbE interface.

```
ethtool interface
```

In the preceding command, `interface` is the 10 GbE interface. The interface is `eth4` and `eth5` for Sun Fire X4170 M2 Oracle Database Servers, and `eth8` through `eth15` for Sun Fire X4800 Oracle Database Servers.

The following is an example of the output from the command. The speed should be 10000Mb/s, Link detected should be yes, and Duplex should be full.

```
ethtool eth9
Settings for eth9:
 Supported ports: [FIBRE]
 Supported link modes: 1000baseT/Full
 10000baseT/Full
 Supports auto-negotiation: No
 Advertised link modes: 1000baseT/Full
 10000baseT/Full
 Advertised auto-negotiation: No
Speed: 10000Mb/s
Duplex: Full
 Port: FIBRE
 PHYAD: 0
 Transceiver: external
 Auto-negotiation: on
 Supports Wake-on: umbg
 Wake-on: umbg
 Current message level: 0x00000007 (7)
Link detected: yes
```

4. Verify the current bonded interface using the following command. An example of the output from the command is also shown.

```
grep -i bondeth0 ifcfg-eth*

ifcfg-eth1:MASTER=bondeth0
ifcfg-eth2:MASTER=bondeth0
```

5. Copy the 1 GbE interface files to the `.Pre_10GigE_Settings` directory using a command similar to the following:

```
cp -p ifcfg-eth1 ifcfg-eth2 ../Pre_10GigE_Settings/.
```

6. Copy the 10 GbE interface files to the `.Pre_10GigE_Settings` directory using a command similar to the following:

```
cp -p ifcfg-eth9 ifcfg-eth15 ../Pre_10GigE_Settings/.
```

7. Copy the files from the `.Pre_10GigE_Settings` directory to the `.Post_10GigE_Settings` directory using a command similar to the following:

```
cp -p ../Pre_10GigE_Settings/* ../Post_10GigE_Settings/.
```

### 4.6.3 Task 3: Edit the 10 GbE Interface Settings

The following procedure describes how to edit the `ifcfg` configuration files:

1. Edit the `ifcfg` configuration files as shown in the following table. The files must be edited in the `./Post_10GigE_Settings/.` directory.



File Name	Before Modification	After Modification
ifcfg-eth1	<pre>#### DO NOT REMOVE THESE LINES #### #### %GENERATED BY CELL% #### DEVICE=eth1 USERCTL=no ONBOOT=yes BOOTPROTO=none HOTPLUG=no IPV6INIT=no HWADDR=00:21:28:44:d2:5e MASTER=bondeth0 SLAVE=yes</pre>	<pre>#### DO NOT REMOVE THESE LINES #### #### %GENERATED BY CELL% #### DEVICE=eth1 USERCTL=no ONBOOT=no BOOTPROTO=none HOTPLUG=no IPV6INIT=no HWADDR=00:21:28:44:d2:5e</pre>
ifcfg-eth2	<pre>#### DO NOT REMOVE THESE LINES #### #### %GENERATED BY CELL% #### DEVICE=eth2 USERCTL=no ONBOOT=yes BOOTPROTO=none HOTPLUG=no IPV6INIT=no HWADDR=00:21:28:44:d2:f2 MASTER=bondeth0 SLAVE=yes</pre>	<pre>#### DO NOT REMOVE THESE LINES #### #### %GENERATED BY CELL% #### DEVICE=eth2 USERCTL=no ONBOOT=no BOOTPROTO=none HOTPLUG=no IPV6INIT=no HWADDR=00:21:28:44:d2:f2</pre>
ifcfg-eth4 on Oracle Exadata Database Machine X2-2	<pre>#### DO NOT REMOVE THESE LINES #### #### %GENERATED BY CELL% #### DEVICE=eth_interface ONBOOT=no BOOTPROTO=none HOTPLUG=no IPV6INIT=no HWADDR=00:1b:21:66:4b:c0</pre>	<pre>#### DO NOT REMOVE THESE LINES #### #### %GENERATED BY CELL% #### DEVICE=eth_interface USERCTL=no ONBOOT=yes BOOTPROTO=none HOTPLUG=no IPV6INIT=no HWADDR=00:1b:21:66:4b:c0 MASTER=bondeth0 SLAVE=yes</pre>
ifcfg-eth9 on Oracle Exadata Database Machine X2-8 Full Rack	<pre>In the preceding syntax, eth_interface is eth4 for Oracle Exadata Database Machine X2-2, or eth9 for Oracle Exadata Database Machine X2-8 Full Rack</pre>	<pre>In the preceding syntax, eth_interface is eth4 for Oracle Exadata Database Machine X2-2, or eth9 for Oracle Exadata Database Machine X2-8 Full Rack</pre>

File Name	Before Modification	After Modification
ifcfg-eth5 on Oracle Exadata Database Machine X2-2	#### DO NOT REMOVE THESE LINES #### #### %GENERATED BY CELL% #### DEVICE= <i>eth_interface2</i> ONBOOT=no BOOTPROTO=none HOTPLUG=no IPV6INIT=no HWADDR=00:1b:21:66:4b:c1	#### DO NOT REMOVE THESE LINES #### #### %GENERATED BY CELL% #### DEVICE= <i>eth_interface2</i> USERCTL=no ONBOOT=yes BOOTPROTO=none HOTPLUG=no IPV6INIT=no MASTER=bondeth0 SLAVE=yes HWADDR=00:1b:21:66:4b:c1
ifcfg- eth15 on Oracle Exadata Database Machine X2-8 Full Rack	In the preceding syntax, <i>eth_interface</i> is eth5 for Oracle Exadata Database Machine X2-2, or eth15 for Oracle Exadata Database Machine X2-8 Full Rack	In the preceding syntax, <i>eth_interface</i> is eth5 for Oracle Exadata Database Machine X2-2, or eth15 for Oracle Exadata Database Machine X2-8 Full Rack

- Copy the edited files to the `/etc/sysconfig/network-scripts` directory using the following command:

```
cp -fp /etc/sysconfig/network-scripts/.Post_10GigE_Settings/ifcfg-eth* \
/etc/sysconfig/network-scripts/.
```
- Restart the database server using the console.
- Monitor the boot sequence to ensure no errors occurred during `bondeth0` initialization.

#### 4.6.4 Task 4: Verify the 10 GbE Interfaces

The following procedure describes how to verify the 10 GbE interfaces:

- Log in as the `root` user.
- Use the `cat` command to review the `/proc/net/bonding/bondeth0` file. The following is an example of the command and output from the command:

```
cat /proc/net/bonding/bondeth0

Ethernet Channel Bonding Driver: v3.4.0 (October 7, 2008)

Bonding Mode: fault-tolerance (active-backup)
Primary Slave: None
Currently Active Slave: eth9
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 5000
Down Delay (ms): 5000

Slave Interface: eth9
MII Status: up
Link Failure Count: 0
Permanent HW addr: 00:1b:21:66:4b:c0

Slave Interface: eth15
MII Status: up
Link Failure Count: 0
Permanent HW addr: 00:1b:21:66:4b:c1
```

In the output, verify the slave interfaces are correct, and the MII statuses for the slave interface are up.

3. Use the `netstat -nr` command to check the routing table. The routing table should not have changed. The following is an example of the command and output:

```
netstat -nr

Kernel IP routing table
Destination Gateway Genmask Flags MSS Window irtt Iface
scan_subnet 0.0.0.0 255.255.255.0 U 0 0 0 bondeth0
192.168.80.0 0.0.0.0 255.255.254.0 U 0 0 0 bondib0
192.168.80.0 0.0.0.0 255.255.254.0 U 0 0 0 bondib1
192.168.80.0 0.0.0.0 255.255.254.0 U 0 0 0 bondib2
192.168.80.0 0.0.0.0 255.255.254.0 U 0 0 0 bondib3
mgmt_subnet 0.0.0.0 255.255.254.0 U 0 0 0 eth0
0.0.0.0 scan_gw 0.0.0.0 UG 0 0 0 bondeth0
```

4. Use the following commands to check the default gateway. The gateway is the SCAN network gateway, and should use bondeth0 on the 10 GbE interfaces.

```
grep GATEWAY /etc/sysconfig/network
GATEWAY=gw_address
GATEWAYDEV=bondeth0

ping -c 2 gw_address
PING gw_address (gw_address) 56(84) bytes of data.
64 bytes from gw_address: icmp_seq=1 ttl=57 time=1.12 ms
64 bytes from gw_address: icmp_seq=2 ttl=57 time=1.05 ms

--- gw_address ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1002ms
rtt min/avg/max/mdev = 1.054/1.087/1.120/0.033 ms
```

In the preceding commands and output, `gw_address` is the IP address of the default gateway.

5. If the name servers were responding to the `ping` command in "[Task 1: Verify ping Functionality](#)", then use the following commands to check the name servers. Similar commands can be used to check other servers that connect to Oracle Exadata Database Machine.

```
grep "^nameserver" /etc/resolv.conf
nameserver ip_address_1
nameserver ip_address_2

ping -c 2 ip_address_1
PING ip_address_1 (ip_address_1) 56(84) bytes of data.
64 bytes from ip_address_1: icmp_seq=1 ttl=57 time=1.12 ms
64 bytes from ip_address_1: icmp_seq=2 ttl=57 time=1.05 ms

--- ip_address_1 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1002ms
rtt min/avg/max/mdev = 1.054/1.087/1.120/0.033 ms
```

## 4.7 Maintaining the RoCE Network Fabric

The RoCE Network Fabric connects the database servers and Exadata Storage Servers through the bonded interface to the RoCE Network Fabric switches in the rack.

- [Backing Up Settings on the Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch](#)
- [Applying Golden Configuration Settings on Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches](#)  
The golden configuration settings are normally applied on the RoCE Network Fabric switches during initial deployment.
- [Restoring Settings on a Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch](#)  
You can restore the Cisco Nexus 9336C-FX2 RoCE Network Fabric switch settings from a backup.
- [Using Access VLANs with Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches](#)  
You can change the `switchport access vlan ID` setting to implement server-level isolation across the RoCE Network Fabric.
- [Replacing a Failed RoCE Network Fabric Switch](#)  
This procedure describes how to replace a failed RoCE Network Fabric switch.
- [Verifying the RoCE Network Fabric Configuration](#)  
This procedure describes how to verify the RoCE Network Fabric configuration.
- [Verifying RoCE Network Fabric Operation](#)  
Verify the RoCE Network Fabric is operating properly after making modifications to the underlying hardware.
- [Upgrading the Switch Firmware for RoCE Network Fabric](#)  
The `patchmgr` utility is used to upgrade the RoCE Network Fabric switches.
- [Downgrading the RoCE Network Fabric Switch Software](#)  
The `patchmgr` utility is used to downgrade the RoCE Network Fabric switches.

### 4.7.1 Backing Up Settings on the Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch

The following procedure describes how to back up the Cisco Nexus 9336C-FX2 RoCE Network Fabric switch settings. A backup is recommended after the switch is initially configured, and again after every configuration change.

1. Access the switch using SSH, and log in as the `admin` user and password.
2. Review the current configuration.

```
switch# show running-config
```

3. Copy the current configuration to a file.

You copy the current configuration to a file on the database server or storage server, using the format:

```
switch# copy running-config tftp://hostname/directory_name/switch_name-
start-config.back
```

You can use any of the supported transport schemes on the RoCE Network Fabric switch: `tftp`, `ftp`, `scp` or `sftp`. The *hostname* is the address or name of the remote server, and the *directory\_name* is the path to the directory that contains the file on the remote server. The *hostname*, *directory\_name*, and file name are case sensitive.

4. Exit from the session.

```
switch# exit
```

## 4.7.2 Applying Golden Configuration Settings on Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches

The golden configuration settings are normally applied on the RoCE Network Fabric switches during initial deployment.

### Note:

Do not apply the golden configuration settings to an active switch in the RoCE Network Fabric. Applying the golden configuration settings to an active switch may disrupt the RoCE Network Fabric and cause a system outage.

Starting with Oracle Exadata System Software release 20.1.0, you can use the following procedure to apply the golden configuration settings on the RoCE Network Fabric switches. For earlier releases, see [Restoring Settings on a Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch](#).

The following procedure applies the latest golden configuration settings to one or more switches in an Exadata single rack or multi-rack configuration. The switches must be powered on along with one server that has SSH access to the switches for accessing the switch configuration files.

1. Ensure you have a backup of the current switch configuration for each switch.  
**For each switch**, complete the steps in [Backing Up Settings on the Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch](#).
2. Log in to a server that has SSH access to the switch, and contains the latest RDMA Network Fabric patch ZIP file.  
To find the available RDMA Network Fabric patches, search for '**RDMA network switch**' in My Oracle Support document 888828.1. Download and use the latest patch for your Oracle Exadata System Software release.
3. Unzip the RDMA Network Fabric patch ZIP file and change directories to the location of the `patchmgr` utility.
4. Create a switch list file to drive the configuration of the RoCE Network Fabric switches.
  - a. Create a file that contains the host name or IP address of the switches that you want to configure. Place each switch on a separate line.

For example, create a file named `switches.lst`, which contains the host name of each switch on separate lines. On a single rack system, the file might contain:

```
switch123-rocea0
switch123-roceb0
```

**b.** Tag each line to specify the configuration type for each switch.

To specify the configuration type for each switch, append a colon (:) and tag to each switch host name or IP address in the switch list file. The following tags are supported:

- `leaf` - Identifies a leaf switch in a single rack system. This configuration type is assumed if no tag is specified.
- `mleaf` - Identifies a leaf switch in a multi-rack system.
- `sflleaf` - Identifies a leaf switch in a single rack system that is enabled to support Exadata Secure RDMA Fabric Isolation.
- `msflleaf` - Identifies a leaf switch in a multi-rack system that is enabled to support Exadata Secure RDMA Fabric Isolation.
- `mspine` - Identifies a spine switch. Note that one spine switch configuration supports all spine switches on single and multi-rack systems, with and without Exadata Secure RDMA Fabric Isolation.
- `leaf23` - Identifies a leaf switch in a single rack system that is configured with 23 host ports. This configuration is required only for 8-socket systems (X8M-8 and later) with 3 database servers and 11 storage servers.
- `mleaf23` - Identifies a leaf switch in a multi-rack system that is configured with 23 host ports. This configuration is required only for 8-socket X8M-8 systems with 3 database servers and 11 storage servers.
- `mleaf_u14` - Identifies a leaf switch in a multi-rack system that is configured with 14 inter-switch links. This is the typical multi-rack leaf switch configuration for X9M systems.
- `msflleaf_u14` - Identifies a leaf switch in a multi-rack system that is enabled to support Exadata Secure RDMA Fabric Isolation and is configured with 14 inter-switch links. This configuration is required for X9M systems with Secure Fabric enabled.
- `mleaf23_u13` - Identifies a leaf switch in a multi-rack system that is configured with 23 host ports and 13 inter-switch links. This configuration is required only for 8-socket X9M-8 systems with three database servers and 11 storage servers.

For example:

```
switch123-rocea0:leaf
switch123-roceb0:leaf
```

**c.** For multi-rack configurations only, specify a unique loopback octet for each switch.

The loopback octet is the last octet of the switch loopback address, which uniquely identifies a switch.

To specify the loopback octet for each switch, append a period (.) and numeric loopback octet value to each tagged switch entry in the switch list file. The range of valid loopback octet values is:

- 101-118 for leaf switches
- 201-208 for spine switches

For example, the switch list file for a 2-rack system might contain:

```
rack1sw-rocea0:mleaf.101
rack1sw-roceb0:mleaf.102
rack1sw-roces0:mspine.201
rack2sw-rocea0:mleaf.103
rack2sw-roceb0:mleaf.104
rack2sw-roces0:mspine.202
```

Or, if you were adding a 5th rack to an existing 4-rack system the switch list file might contain:

```
rack5sw-rocea0:mleaf.109
rack5sw-roceb0:mleaf.110
rack5sw-roces0:mspine.205
```

5. Use `patchmgr` to apply the latest golden configuration settings to the RoCE Network Fabric switches in the switch list file.

For example:

```
./patchmgr --roceswitches switches.lst --apply-config -log_dir
log-directory
```

#### See Also:

[Patchmgr Syntax for RoCE Network Fabric Switches.](#)

6. Use `patchmgr` to verify the configuration of the RoCE Network Fabric switches in the switch list file.

For example:

```
./patchmgr --roceswitches switches.lst --verify-config -log_dir
log-directory
```

#### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

## 4.7.3 Restoring Settings on a Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch

You can restore the Cisco Nexus 9336C-FX2 RoCE Network Fabric switch settings from a backup.

The following procedure outlines how to restore Cisco Nexus 9336C-FX2 RoCE Network Fabric switch settings from a backup, and can also be used to restore golden configuration settings on a switch.



### Note:

Oracle Exadata System Software release 20.1.0 contains a simplified and enhanced procedure for applying golden configuration settings on the RoCE Network Fabric switches. See [Applying Golden Configuration Settings on Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches](#).

1. Access the switch using SSH, and log in as the `admin` user and password.
2. Delete any existing backup configuration (or golden configuration) file on the switch for the configuration you are restoring.

The golden configuration files are:

- Single rack leaf: `roce_leaf_switch.cfg`
- Multi-rack leaf: `roce_leaf_switch_multi.cfg`
- Multi-rack spine: `roce_spine_switch_multi.cfg`
- Single rack leaf with Secure Fabric support: `roce_sf_leaf_switch.cfg`
- Multi-rack leaf with Secure Fabric support: `roce_sf_leaf_switch_multi.cfg`
- Single rack leaf configured with 23 host ports: `roce_leaf_switch_23hosts.cfg`
- Multi-rack leaf configured with 23 host ports: `roce_leaf_switch_23hosts_multi.cfg`
- Multi-rack leaf configured with 14 inter-switch links:  
`roce_leaf_switch_14uplinks_multi.cfg`
- Multi-rack leaf configured with 14 inter-switch links and with Secure Fabric support:  
`roce_sf_leaf_switch_14uplinks_multi.cfg`
- Multi-rack leaf configured with 23 host ports and 14 inter-switch links:  
`roce_leaf_switch_23hosts_13uplinks_multi.cfg`



 **Note:**

If you do not remove the file you are replacing, then when you attempt to overwrite the file you will get a 'permission denied' error.

```
rack3sw-rocea0# delete bootflash:roce_leaf_switch.cfg
Do you want to delete "/roce_leaf_switch.cfg" ? (yes/no/abort) [y] y
rack3sw-rocea0#
```

3. Copy the backup configuration file (or golden configuration file) to the switch.

You can restore the appropriate golden configuration file based on your system configuration and type of switch. The files are located within the patchmgr switch bundle in the `roce_switch_templates/` directory.

You can also restore a backup configuration file that was created manually and stored in a separate location.

```
[root@server_hostname ~]# scp roce_leaf_switch.cfg
admin@100.104.10.21:/
User Access Verification
Password:
roce_leaf_switch.cfg 100% 23KB 23.5KB/s 00:00
```

 **Note:**

You can use any of the supported transport schemes on the RoCE Network Fabric switch: `tftp`, `ftp`, `scp` or `sftp`.

4. Apply the backup configuration (or golden configuration) file using the `run-script` command.

```
rack3sw-rocea0# run-script bootflash:roce_leaf_switch.cfg | grep
'none'
```

 **Note:**

This command may take up to 1-2 minutes on single-rack switch and up to 3-4 minutes on a multi-rack switch.

5. Save the configuration onto flash.

```
rack3sw-rocea0# copy running-config startup-config
```

6. Exit from the session.

```
rack3sw-rocea0# exit
```

## 4.7.4 Using Access VLANs with Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches

You can change the `switchport access vlan` ID setting to implement server-level isolation across the RoCE Network Fabric.

By default, Exadata Database Machine uses Access VLAN ID 3888 for all RoCE Network Fabric private network traffic, on the server `re0` and `re1` interfaces.

If you change the default Access VLAN ID setting on the Cisco Nexus 9336C-FX2 RoCE Network Fabric switches, the corresponding server can no longer communicate with other database servers or storage servers using the default setting. The range of valid Access VLAN IDs is 2744-3967.

Use the following procedure to implement different Access VLANs on the RoCE Network Fabric.

### Note:

It is recommended that you shut down the servers while modifying the switch configuration to avoid any outages. If you are changing the Access VLAN ID for a subset of the servers in the rack, then only the affected servers need to be shut down prior to modifying the switch.

1. Identify the RoCE Network Fabric switch host name and port that is connected to the `re0` interface of the server that you are configuring.

```
[root@dbm01adm02 ~]# /opt/oracle.SupportTools/ibdiagtools/utils/
lldp_cap.py re0
| egrep 'SWITCH_PORT:|SWITCH_NAME:'
SWITCH_PORT: Ethernet1/21
SWITCH_PORT_DESCRIPTION: adm02
```

2. Log in to the lower leaf switch using the `SWITCH_NAME` identified in the previous step.

```
ssh admin@dbm01sw-rocea0
User Access Verification
Password: *****
```

3. Check the configuration for the switch port.

To specify the interface, use the `SWITCH_PORT` identified in Step 1.

```
dbm01sw-rocea0# show running-config interface ethernet 1/21
!Running configuration last done at: Wed Nov 13 10:34:58 2019
!Time: Wed Nov 13 14:55:48 2019

version 7.0(3)I7(6) Bios:version 05.33

interface Ethernet1/21
description adm02
```

```
switchport access vlan 3888
priority-flow-control mode on
spanning-tree port type edge
spanning-tree bpduguard enable
mtu 2300
speed 100000
duplex full
no negotiate auto
service-policy type qos input QOS_MARKING no-stats
```

4. Create the new VLAN ID on the switch.

This configuration is only required once on each switch. However, it is harmless to repeat the VLAN configuration on a switch.

For example, to create a new VLAN ID with the value 3889:

```
dbm01sw-rocea0# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
dbm01sw-rocea0(config)# vlan 3889
dbm01sw-rocea0(config-vlan)# exit
dbm01sw-rocea0(config)# exit
dbm01sw-rocea0#
```

5. Modify the switch port configuration to change the `switchport access vlan` setting.

Specify the same interface as in the previous steps. Then, remove the old VLAN ID (for example, 3888), add the new VLAN ID (for example, 3889), and exit configuration mode.

```
dbm01sw-rocea0# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
dbm01sw-rocea0(config)# interface ethernet 1/21
dbm01sw-rocea0(config-if)# no switchport access vlan 3888
dbm01sw-rocea0(config-if)# switchport access vlan 3889
dbm01sw-rocea0(config-if)# exit
dbm01sw-rocea0(config)# exit
dbm01sw-rocea0#
```

6. Verify that the switch interface is using the new VLAN ID.

Specify the same interface as in the previous steps.

```
dbm01sw-rocea0# show running-config interface ethernet 1/21

!Command: show running-config interface Ethernet1/21
!Running configuration last done at: Wed Nov 20 23:53:38 2019
!Time: Wed Nov 20 23:53:45 2019

version 7.0(3)I7(6) Bios:version 05.33

interface Ethernet1/21
 description adm02
 switchport access vlan 3889
 priority-flow-control mode on
```

```
spanning-tree port type edge
spanning-tree bpduguard enable
mtu 2300
speed 100000
duplex full
no negotiate auto
service-policy type qos input QOS_MARKING no-stats
```

7. Save the configuration.

```
dbm01sw-rocea0# copy running-config startup-config
[#####] 100%
Copy complete, now saving to disk (please wait)...
Copy complete.
```

8. Repeat Steps 1 to 7 for all of the database servers (bare metal servers or KVM hosts) and all of the storage servers that you want to change to the new VLAN ID.
9. Repeat Steps 1 to 8 for the `re1` interfaces, which are connected to the upper leaf switch.

## 4.7.5 Replacing a Failed RoCE Network Fabric Switch

This procedure describes how to replace a failed RoCE Network Fabric switch.

This procedure depends on having a backup of the configuration for the failed switch.

1. Power off both power supplies on the switch by removing the power plugs.
2. Verify each cable is labeled, and then disconnect each cable from the switch.

All RoCE Network Fabric cables should have labels at both ends indicating their locations. If there are any cables that do not have labels, then add a label before disconnecting the cable.

You will use these labels to cable the replacement switch.

3. Remove the switch from the rack.
  - a. Extend the rack cabinet's anti-tilt bar.
  - b. Attach an antistatic wrist strap.
  - c. Remove the switch from the rack.
4. Install the new switch in the rack.

After you have placed the new switch in the correct position in the rack, if there is no further equipment being serviced in the rack, then you can retract the anti-tilt bar on the rack cabinet.

5. Power on the switch by plugging in the power plugs.
6. Restore the switch settings using the backup, as described in [Restoring Settings on a Cisco Nexus 9336C-FX2 RoCE Network Fabric Switch](#).
7. Connect the cables to the new switch.

Use the labels on each cable to ensure that you connect each cable to the correct port on the new switch.

8. Complete the steps in [Verifying the RoCE Network Fabric Configuration](#).

**Related Topics**

- Part II Cabling Information for Oracle Exadata Database Machine

## 4.7.6 Verifying the RoCE Network Fabric Configuration

This procedure describes how to verify the RoCE Network Fabric configuration.

1. Verify the proper `oracle-rdma-release` software versions are being used on the database servers.

```
[root@dbm01adm08 ~]# rpm -qa |grep oracle-rdma-release
oracle-rdma-release-0.11.0-1.el7ora.x86_64
```

The `oracle-rdma-release` software and adapter firmware versions are automatically maintained on the Exadata Database Machine storage servers.

2. Check the adapter firmware versions on the database servers.

Use the `CheckHWnFWProfile` script to check firmware versions for the RDMA Network Fabric adapters.

```
/opt/oracle.SupportTools/CheckHWnFWProfile -action list
```

3. Visually check all the RDMA Network Fabric cable connections within the rack. The port lights should be on, and the LEDs should be on. Do not press each connector to verify connectivity.
4. Complete the steps described in My Oracle Support Doc ID 2587717.1

**Related Topics**

- [Verify RoCE Cabling on Oracle Exadata Database Machine X8M-2 and X8M-8 Servers \(My Oracle Support Doc ID 2587717.1\)](#)

## 4.7.7 Verifying RoCE Network Fabric Operation

Verify the RoCE Network Fabric is operating properly after making modifications to the underlying hardware.

If hardware maintenance has taken place with any component in the RoCE Network Fabric, including replacing an RDMA Network Fabric Adapter on a server, a switch, or a cable, or if the operation of the RoCE Network Fabric is suspected to be substandard, then verify the RoCE Network Fabric is operating properly. The following procedure describes how to verify network operation:

1. Complete the steps in [Verifying the RoCE Network Fabric Configuration](#).
2. Prepare for `infinicheck`.

You may need to run the following commands before you can use the `infinicheck` command to perform RoCE Network Fabric configuration, connectivity, and performance checks.

- If required, use the `-s` option set up user equivalence for password-less SSH across the RoCE Network Fabric. For example:

```
/opt/oracle.SupportTools/ibdiagtools/infinicheck -g hosttips -c
cellips -s
```

- You can use the `-z` option to clear the files that were created during the last run of the `infinicheck` command. For example:

```
/opt/oracle.SupportTools/ibdiagtools/infinicheck -g hosttips -c
cellips -z
```

In the previous commands, `hosttips` is the name of an input file that contains a comma-delimited list of database server host names or RoCE Network Fabric IP addresses, and `cellips` is the name of an input file that contains a list of RoCE Network Fabric IP addresses for the storage servers.

3. Run the `infinicheck` command to perform RoCE Network Fabric configuration, connectivity, and performance checks.

On a properly configured system, you can run the `infinicheck` command on any database server with minimal arguments. For example:

```
/opt/oracle.SupportTools/ibdiagtools/infinicheck
```

By default, the `infinicheck` command performs a group of configuration and connectivity checks on the RoCE Network Fabric. You can use the `-p` option to run the optional performance tests. Or, use the `-a` option to perform all checks, including the performance tests. For example:

```
/opt/oracle.SupportTools/ibdiagtools/infinicheck -a
```

 **Note:**

System performance may be impacted when the `infinicheck` command performs performance stress tests. Consequently, only run the `infinicheck` performance tests when required and preferably when there is no workload on the system.

You can also specify the servers in your system explicitly by using the `-g` option to specify the database servers and the `-c` option to specify the storage servers. For example:

```
/opt/oracle.SupportTools/ibdiagtools/infinicheck -g hosttips -c cellips
```

In the previous example, `hosttips` is the name of an input file that contains a comma-delimited list of database server host names or RoCE Network Fabric IP addresses, and `cellips` is the name of an input file that contains a list of RoCE Network Fabric IP addresses for the storage servers.

Instead of listing the database servers and storage servers in input files, you can supply a comma-separated list of host names or IP addresses on the command line.

The following example displays typical terminal output from the `infinicheck` command.

```
/opt/oracle.SupportTools/ibdiagtools/infinicheck -g hosts -c
cellips
 INFINICHECK
 [Network Connectivity, Configuration and
Performance]

 ##### FABRIC TYPE TESTS #####

System type identified: RoCE
Verifying User Equivalence of user=root from all DBs to all CELLS.

 ##### RoCE CONFIGURATION TESTS #####
 Checking for presence of RoCE devices on all DBs and CELLS
[SUCCESS].... RoCE devices on all DBs and CELLS look good
 Checking for RoCE Policy Routing settings on all DBs and
CELLS
[SUCCESS].... RoCE Policy Routing settings look good
 Checking for RoCE DSCP ToS mapping on all DBs and CELLS
[SUCCESS].... RoCE DSCP ToS settings look good
 Checking for RoCE PFC settings and DSCP mapping on all DBs
and CELLS
[SUCCESS].... RoCE PFC and DSCP settings look good
 Checking for RoCE interface MTU settings. Expected value :
2300
[SUCCESS].... RoCE interface MTU settings look good
 Verifying switch advertised DSCP on all DBs and CELLS ports
(~ 2 min)
[SUCCESS].... Advertised DSCP settings from RoCE switch looks good

 ##### CONNECTIVITY TESTS #####
 [COMPUTE NODES -> STORAGE CELLS]
 (60 seconds approx.)
 (Will walk through QoS values: 0-6)
[SUCCESS].....Results OK
[SUCCESS]..... All can talk to all storage cells
 [COMPUTE NODES -> COMPUTE NODES]
 (60 seconds approx.)
 (Will walk through QoS values: 0-6)
[SUCCESS].....Results OK
[SUCCESS]..... All hosts can talk to all other nodes
 Verifying Subnet Masks on all nodes
[SUCCESS] Subnet Masks is same across the network
```

## 4.7.8 Upgrading the Switch Firmware for RoCE Network Fabric

The `patchmgr` utility is used to upgrade the RoCE Network Fabric switches.

The switch firmware is upgraded in a rolling manner. `patchmgr` upgrades the switches in the order they are listed in the supplied file, for example `roceswitch.lst`.

Refer to [Upgrading RoCE Network Fabric Switch Firmware Software](#) for the instructions.

## 4.7.9 Downgrading the RoCE Network Fabric Switch Software

The `patchmgr` utility is used to downgrade the RoCE Network Fabric switches.

The switch firmware is downgraded in a rolling manner. `patchmgr` downgrades the switches in the order they are listed in the supplied file, for example `roceswitch.lst`.

Refer to [Downgrading RoCE Network Fabric Switch Firmware](#) for the instructions.

## 4.8 Maintaining the InfiniBand Network Fabric Network

The InfiniBand Network Fabric network connects the database servers and Exadata Storage Servers through the bonded interface to the InfiniBand Network Fabric switches in the rack.

- [Backing Up and Restoring InfiniBand Switch Settings](#)  
The procedure for backing up and restoring InfiniBand switch settings depends on the firmware on the switch.
- [Verifying the InfiniBand Network Fabric Configuration](#)  
This procedure describes how to verify the InfiniBand Network Fabric configuration.
- [Using the verify-topology Utility](#)  
The `verify-topology` utility can be used to identify various network connection problems.
- [Verifying InfiniBand Network Fabric Operation](#)  
Verify the InfiniBand Network Fabric network is operating properly after making modifications to the underlying hardware.
- [Understanding the Network Subnet Manager Master](#)  
The Subnet Manager manages all operational characteristics of the InfiniBand Network Fabric network.
- [Upgrading the Switch Firmware for InfiniBand Network Fabric](#)
- [Downgrading the InfiniBand Network Fabric Switch Software](#)  
Use `patchmgr` to downgrade the switch firmware.

### 4.8.1 Backing Up and Restoring InfiniBand Switch Settings

The procedure for backing up and restoring InfiniBand switch settings depends on the firmware on the switch.

The InfiniBand firmware release 1.1.3-2 or later has Integrated Lights Out Manager (ILOM) which provides backup and restore capability. The InfiniBand firmware release 1.0.1 does not have ILOM. You can either upgrade to the latest available firmware release and then use the procedure in [Backing Up Settings on a Switch with 2.1.3-4 Firmware](#), or you can manually perform the backup and restore of individual files.

- [Backing Up Settings on a Switch with 2.1.3-4 Firmware](#)
- [Backing Up Settings on a Switch with 1.1.3-2 Firmware](#)
- [Backing Up Settings on a Switch with 1.0.1 Firmware](#)
- [Restoring Settings on a Switch with 2.1.3-4 Firmware](#)



- [Restoring Settings on a Switch with 1.1.3-2 Firmware](#)
- [Restoring Settings on a Switch with 1.0.1 Firmware](#)



**See Also:**

Oracle Integrated Lights Out Manager (ILOM) Documentation at <http://www.oracle.com/goto/ilom/docs>

### 4.8.1.1 Backing Up Settings on a Switch with 2.1.3-4 Firmware

The following procedure describes how to back up a switch with 2.1.3-4 firmware. The backup only needs to be done once after the switch has been initially configured with the right settings.

1. Navigate to the switch ILOM URL in a browser. For example: `http://dbm002-il.us.example.com`.
2. Log in as the `ilom-admin` user.
3. Select the **Maintenance** tab.
4. Select the **Backup/Restore** tab.
5. Select the **Backup** operation and the **Browser** method.
6. Enter a passphrase. This is used to encrypt sensitive information, such as user passwords, in the backup.
7. Click **Run**, and save the resulting XML file in a secure location.
8. Log in to the Sun Datacenter InfiniBand Switch 36 switch as the `root` user.
9. Use the `scp` command to copy the following files:
  - root SSH keys: `/root/.ssh/authorized_keys`
  - nm2user SSH keys (if it exists): `/home/nm2user/.ssh/authorized_keys`
  - host file: `/etc/hosts`
  - host file: `/etc/opensm/opensm.conf` to preserve the openSM settings
10. Save the output from the `version` command.

### 4.8.1.2 Backing Up Settings on a Switch with 1.1.3-2 Firmware

The following procedure describes how to back up a switch with 1.1.3-2 firmware. The backup only needs to be done once after the switch has been initially configured with the right settings.

1. Navigate to the switch ILOM URL in a browser. For example: `http://dbm002-il.us.example.com`.
2. Log in as the `ilom-admin` user.
3. Select the **Maintenance** tab.
4. Select the **Backup/Restore** tab.
5. Select the **Backup** operation and the **Browser** method.

6. Enter a passphrase. This is used to encrypt sensitive information, such as user passwords, in the backup.
7. Click **Run**, and save the resulting XML file in a secure location.
8. Log in to the Sun Datacenter InfiniBand Switch 36 switch as the `root` user.
9. Use the `scp` command to copy the following files:
  - Network configuration: `/etc/sysconfig/network-scripts/ifcfg-eth0`
  - DNS information: `/etc/resolv.conf`
  - NTP information: `/etc/ntp.conf`
  - Time zone information: `/etc/localtime`
  - openSM settings: `/etc/opensm/opensm.conf`
  - Host name: `/etc/sysconfig/network`
  - root SSH keys: `/root/.ssh/authorized_keys`
  - nm2user SSH keys (if it exists): `/home/nm2user/.ssh/authorized_keys`
10. Run the `hostname` command, and then save the output. This is done in case the host name is not set in the `/etc/sysconfig/network` file.
11. Save the passwords for the `root` and `nm2user` accounts.
12. Run the `nm2version` command, and then save the output.

### 4.8.1.3 Backing Up Settings on a Switch with 1.0.1 Firmware

The following procedure describes how to back up the settings on a switch with 1.0.1 firmware:

1. Log in to the switch as the `root` user. If you do not have the password for the `root` user, then contact Oracle Support Services.
2. Make copies of the following files:
  - Network configuration: `/etc/sysconfig/network-scripts/ifcfg-eth0`
  - DNS information: `/etc/resolv.conf`
  - NTP information: `/etc/ntp.conf`
  - Time zone information: `/etc/localtime`
  - openSM settings: `/etc/opensm/opensm.conf`
  - Host name: `/etc/sysconfig/network`
  - root SSH keys: `/root/.ssh/authorized_keys`
  - nm2user SSH keys (if it exists): `/home/nm2user/.ssh/authorized_keys`
3. Run the `hostname` command and save the output, in case the host name is not set in the `/etc/sysconfig/network` file.
4. Save the passwords for the `root` and `nm2user` accounts.
5. Run the `nm2version` command and save the output.

### 4.8.1.4 Restoring Settings on a Switch with 2.1.3-4 Firmware

The following procedure describes how to restore the settings on a switch with 2.1.3-4 firmware:

1. Run the `version` command, and ensure that the switch is at the right firmware level. If not, then upgrade the switch to the correct firmware level.
2. Navigate to the switch ILOM URL in a browser. For example: `http://dbm002-il.us.example.com`.
3. Log in as the `ilom-admin` user.
4. Select the **Maintenance** tab.
5. Select the **Backup/Restore** tab.
6. Select the **Restore** operation and the **Browser** method.
7. Click **Browse**, and select the XML file that contains the switch configuration backup.
8. Enter the passphrase that was used during the backup.
9. Click **Run** to restore the configuration.
10. Log in to the Sun Datacenter InfiniBand Switch 36 switch as the `root` user.
11. Restore the following files from the backup:
  - root SSH keys: `/root/.ssh/authorized_keys`
  - nm2user SSH keys (if it exists): `/home/nm2user/.ssh/authorized_keys`
  - host file: `/etc/hosts`
  - host file: `/etc/opensm/opensm.conf`
12. Restart openSM from the switch CLI using the following commands:

```
disablesm
enablesm
```
13. Log in as the `root` user.
14. Restart the switch.

### 4.8.1.5 Restoring Settings on a Switch with 1.1.3-2 Firmware

The following procedure describes how to restore the settings on a switch with 1.1.3-2 firmware:

1. Run the `version` command, and ensure that the switch is at the right firmware level. If not, then upgrade the switch to the correct firmware level.
2. Navigate to the switch ILOM URL in a browser. For example: `http://dbm002-il.us.example.com`.
3. Log in as the `ilom-admin` user.
4. Select the **Maintenance** tab.
5. Select the **Backup/Restore** tab.

6. Select the **Restore** operation and the **Browser** method.
7. Click **Browse**, and select the XML file that contains the switch configuration backup.
8. Type in the passphrase that was used during the backup.
9. Click **Run** to restore the configuration.
10. Log in to the Sun Datacenter InfiniBand Switch 36 switch as the `root` user.
11. Restore the following files from the backup:
  - Network configuration: `/etc/sysconfig/network-scripts/ifcfg-eth0`
  - DNS information: `/etc/resolv.conf`
  - NTP information: `/etc/ntp.conf`
  - Time zone information: `/etc/localtime`
  - openSM settings: `/etc/opensm/opensm.conf`
  - Host name: `/etc/sysconfig/network`
  - root SSH keys: `/root/.ssh/authorized_keys`
  - nm2user SSH keys (if it exists): `/home/nm2user/.ssh/authorized_keys`
12. Restore the host name by adding the following line to the `/etc/sysconfig/network` file, if it not already in the file.

```
HOSTNAME=switch_host_name
```
13. Restore the passwords of the `root` and `nm2user` users using the `passwd` command.
14. Run the following commands in the order shown to restart the services and openSM:

```
service network restart
service ntpd restart
disablesm
enablesm
```
15. Log in as the `root` user.
16. Restart the switch.

#### 4.8.1.6 Restoring Settings on a Switch with 1.0.1 Firmware

The following procedure describes how to restore the settings to a switch with 1.0.1 firmware:

1. Log in to the switch as the `root` user. If you do not have the password for the `root` user, then contact Oracle Support Services.
2. Ensure that the switch is at the right firmware level. If not, then upgrade the switch to the correct firmware level.
3. Restore the following files from the backup:
  - Network configuration: `/etc/sysconfig/network-scripts/ifcfg-eth0`
  - DNS information: `/etc/resolv.conf`
  - NTP information: `/etc/ntp.conf`
  - Time zone information: `/etc/localtime`
  - openSM settings: `/etc/opensm/opensm.conf`

- Host name: `/etc/sysconfig/network`
  - root SSH keys: `/root/.ssh/authorized_keys`
  - nm2user SSH keys (if it exists): `/home/nm2user/.ssh/authorized_keys`
4. Restore the host name by adding a `HOSTNAME=switch_host_name` line to the `/etc/sysconfig/network` file, if not already present.
  5. Restore the passwords of the `root` and `nm2user` users using the `passwd` command.
  6. Run the following commands in the order shown to restart the services and openSM:

```
service network restart
service ntpd restart
disablesm
enablesm
```
  7. Log in as the `root` user.
  8. Restart the switch.

## 4.8.2 Verifying the InfiniBand Network Fabric Configuration

This procedure describes how to verify the InfiniBand Network Fabric configuration.

1. Verify the proper OpenFabrics Enterprise Distribution (OFED) software and HCA firmware versions are being used on the database servers.

The OFED software and HCA firmware versions are automatically maintained on the Exadata storage servers.
2. Verify the InfiniBand Network Fabric topology using the following command from a database server or Exadata Storage Server:

```
/opt/oracle.SupportTools/ibdiagtools/verify-topology
```

If any errors occur, then contact Oracle Support Services.

### Related Topics

- [Using the verify-topology Utility](#)

The `verify-topology` utility can be used to identify various network connection problems.
- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

## 4.8.3 Using the verify-topology Utility

The `verify-topology` utility can be used to identify various network connection problems.

The problems you can diagnose using `verify-topology` include:

- Missing InfiniBand Network Fabric cable
- Missing InfiniBand Network Fabric connection
- Incorrectly-seated cable

- Cable connected to the wrong endpoint

The utility is available in the `ibdiagtools` directory on all servers. To view the options for the `verify-topology` utility, use the following command:

```
./verify-topology -h

[DB Machine Infiniband Cabling Topology Verification Tool]
Usage: ./verify-topology
 [-v|--verbose]
 [-r|--reuse (cached maps)]
 [-m|--mapfile]
 [-ibn|--ibnetdiscover (specify location of ibnetdiscover output)]
 [-ibh|--ibhosts (specify location of ibhosts output)]
 [-ibs|--ibswitches (specify location of ibswitches output)]
 [-t|--topology [torus | fattree | halfrack] default is fattree]
```

#### Example 4-1 Using `verify-topology` to Identify Cables Seated Incorrectly

The following is an example shows the output when using the `verify-topology` utility. In the example, the error shows the cables are connected incorrectly. Both cables from the server are going to same InfiniBand Network Fabric switch. If the switch fails, then the server loses connectivity to InfiniBand Network Fabric network.

```
[DB Machine Infiniband Cabling Topology Verification Tool]

Bad link:Switch 0x21283a8371a0a0 Port 11A - Sun Port 11B
Reason : 2.5 Gbps Speed found. Could be 10 Gbps
Possible cause : Cable isn't fully seated in

Bad link:Switch 0x21283a89eba0a0 Port 11B - Sun Port 11A
Reason : 2.5 Gbps Speed found. Could be 10 Gbps
Possible cause : Cable isn't fully seated in

Is every external switch connected to every internal switch.....
[SUCCESS]
Are any external switches connected to each other.....
[SUCCESS]
Are any hosts connected to spine switch.....
[SUCCESS]
Check if all hosts have 2 CAs to different switches.....[ERROR]
Node trnA-db01 has 1 endpoints. (Should be 2)
Port 2 of this node is not connected to any switch

-----fattree End Point Cabling verification failed-----

Leaf switch check: cardinality and even distribution.....[ERROR]

Internal QDR Switch 0x21283a8371a0a0 has fewer than 4 compute nodes
It has only 3 links belonging to compute nodes
Check if each rack has a valid internal ring.....
[SUCCESS]
```

## 4.8.4 Verifying InfiniBand Network Fabric Operation

Verify the InfiniBand Network Fabric network is operating properly after making modifications to the underlying hardware.

If hardware maintenance has taken place with any component in the InfiniBand Network Fabric network, including replacing an InfiniBand HCA on a server, an InfiniBand Network Fabric switch, or an InfiniBand Network Fabric cable, or if operation of the InfiniBand Network Fabric is suspected to be substandard, then verify the InfiniBand Network Fabric is operating properly. The following procedure describes how to verify network operation:

 **Note:**

The following procedure can be used any time the InfiniBand Network Fabric is performing below expectations.

1. Complete the steps in [Verifying the InfiniBand Network Fabric Configuration](#).
2. Run the `ibdiagnet` command to verify the InfiniBand Network Fabric operation.

```
ibdiagnet -c 1000
```

All errors reported by this command should be investigated. This command generates a small amount of network traffic, and may be run while normal workload is running.

3. Run the `ibqueryerrors.pl` command to report on switch port error counters and port configuration information.

```
ibqueryerrors.pl -rR -s
RcvSwRelayErrors,XmtDiscards,XmtWait,VL15Dropped
```

Errors such as `LinkDowned`, `RcvSwRelayErrors`, `XmtDiscards`, and `XmtWait` are ignored when using the preceding command.

 **Note:**

- The InfiniBand Network Fabric counters are cumulative and the errors may have occurred at any time in the past. If there are errors reported, then Oracle recommends clearing the InfiniBand Network Fabric counters using the `ibclearcounters` command. After running the command, let the system run for a few minutes under load, and then run the `ibqueryerrors` command.
  - Some counters, such as `SymbolErrors` or `RcvErrors` can increment when servers are rebooted. Small values for these counters which are less than the `LinkDowned` counter are generally not a problem. The `LinkDowned` counter indicates the number of times the port has gone down usually for valid reasons, such as a reboot, and is not usually an error indicator by itself.
  - Any links reporting high, persistent errors especially `SymbolErrors`, `LinkRecovers`, `RcvErrors`, or `LinkIntegrityErrors` may indicate a bad or loose cable or port.
  - If there are persistent, high InfiniBand Network Fabric error counters, then investigate and correct the problem.
4. If there is no load running on any portion of the InfiniBand Network Fabric, such as no databases running, then run the `infinicheck` command to perform full InfiniBand Network Fabric configuration, connectivity and performance evaluation.

 **Note:**

This command evaluates full network maximum throughput and should not be run when there is workload running on any system on the InfiniBand Network Fabric.

This command relies on a fully-configured system. The first command clears the files that were created during the last run of the `infinicheck` command.

```
/opt/oracle.SupportTools/ibdiagtools/infinicheck -z

/opt/oracle.SupportTools/ibdiagtools/infinicheck
```

The following is an example of the output from the command:

```
Verifying User Equivalence of user=root to all hosts.
(If it isn't setup correctly, an authentication prompt will appear to
push keys
to all the nodes)

Verifying User Equivalence of user=root to all cells.
(If it isn't setup correctly, an authentication prompt will appear to
push keys
to all the nodes)
```



```
CONNECTIVITY TESTS
[COMPUTE NODES -> STORAGE CELLS]
(30 seconds approx.)
[SUCCESS].....Connectivity verified

[SUCCESS]..... All hosts can talk to all storage cells

Verifying Subnet Masks on Hosts and Cells
[SUCCESS] Subnet Masks is same across the network

Checking for bad links in the fabric
[SUCCESS]..... No bad fabric links found

[COMPUTE NODES -> COMPUTE NODES]
(30 seconds approx.)
[SUCCESS].....Connectivity verified

[SUCCESS]..... All hosts can talk to all other nodes

PERFORMANCE TESTS

[(1) Every COMPUTE NODE to its STORAGE CELL]
(15 seconds approx.)
[SUCCESS]..... Network Bandwidth looks OK.
..... To view only performance results run ./infinicheck -d -p

[(2) Every COMPUTE NODE to another COMPUTE NODE]
(10 seconds approx.)
[SUCCESS]..... Network Bandwidth looks OK.
..... To view only performance results run ./infinicheck -d -p

[(3) Every COMPUTE NODE to ALL STORAGE CELLS]
(45 seconds approx.) (looking for SymbolErrors)

[SUCCESS]..... No port errors found
```

## 4.8.5 Understanding the Network Subnet Manager Master

The Subnet Manager manages all operational characteristics of the InfiniBand Network Fabric network.

The operational characteristics of the Subnet Manager include:

- Discover the network topology
- Assign a local identifier to all ports connected to the network
- Calculate and program switch forwarding tables
- Monitor changes in the fabric

The InfiniBand Network Fabric network can have more than one Subnet Manager, but only one Subnet Manager is active at a time. The active Subnet Manager is the Master Subnet Manager. The other Subnet Managers are the Standby Subnet Managers. If a

Master Subnet Manager is shut down or fails, then a Standby Subnet Manager automatically becomes the Master Subnet Manager.

Each Subnet Manager has a priority that can be configured. When there is more than one Subnet Manager on the InfiniBand Network Fabric network, the Subnet Manager with the highest priority becomes the Master Subnet Manager. On Exadata Database Machine, the Subnet Managers on leaf switches should be configured as priority 5, and the Subnet Managers on spine switches should be configured as priority 8.

The following guidelines determine where Subnet Managers run on Exadata Database Machine:

- Only run Subnet Managers on the RDMA Network Fabric switches specified for use in your Oracle Engineered System. Running Subnet Manager on any other device is not supported.
- In Exadata-only configurations, when the InfiniBand Network Fabric network consists of one, two, or three racks cabled together, all switches should run Subnet Manager. The Master Subnet Manager should be run on a spine switch. If the network has only leaf switches, as in Exadata Database Machine Quarter Racks, then Subnet Manager Master runs on a leaf switch. When the InfiniBand Network Fabric network consists of four or more racks cabled together, then only spine switches should run Subnet Manager. The leaf switches should disable Subnet Manager.
- In multi-rack configurations, using different types of racks such as Oracle Exadata Database Machine and Oracle Exalogic Elastic Cloud, see My Oracle Support Doc ID 1682501.1.

#### Related Topics

- Setting the Subnet Manager Master on Oracle Exadata Database Machine Full Rack and Oracle Exadata Database Machine Half Rack
- [Setting up the Subnet Manager in a multi-rack cabling configuration containing Exalogic/Big Data Appliance and Exadata/SuperCluster \(My Oracle Support Doc ID 1682501.1\)](#)



#### See Also:

Sun Datacenter InfiniBand Switch 36 Firmware Version 2.1 Documentation at [http://docs.oracle.com/cd/E36265\\_01/index.html](http://docs.oracle.com/cd/E36265_01/index.html)

## 4.8.6 Upgrading the Switch Firmware for InfiniBand Network Fabric

The `patchmgr` utility is used to upgrade and downgrade the InfiniBand Network Fabric switches. The minimum switch firmware release that can use the `patchmgr` utility is release 1.3.3-2. If the switch firmware is at an earlier release, then it is necessary to upgrade the firmware to release 1.3.3-2 using the instructions in My Oracle Support note 888828.1.

Refer to [Updating InfiniBand Network Fabric Switch Firmware Software](#) for the instructions.

## 4.8.7 Downgrading the InfiniBand Network Fabric Switch Software

Use `patchmgr` to downgrade the switch firmware.

Refer to [Downgrading InfiniBand Network Fabric Switch Firmware](#) for the instructions.

## 4.9 Modifying the InfiniBand Network Fabric Configuration

You can change how the InfiniBand Network Fabric network is configured by changing the IP addresses or host names, or by implementing partitioning.

- [Configuring InfiniBand Partitioning](#)
- [Changing InfiniBand IP Addresses and Host Names](#)
- [Changing InfiniBand Network Information](#)

This procedure describes how to change the InfiniBand network information.

### 4.9.1 Configuring InfiniBand Partitioning

Configuring InfiniBand partitioning is described in [Implementing InfiniBand Partitioning across Oracle VM Oracle RAC Clusters on Exadata Database Machine](#). You can use InfiniBand partitioning with or without Oracle VM.

### 4.9.2 Changing InfiniBand IP Addresses and Host Names

It may be necessary to change the InfiniBand network information on an existing Oracle Exadata Rack. The change may be needed to support a media server with multiple InfiniBand cards, or keep InfiniBand traffic on a distinct InfiniBand network such as having production, test and QA environments in the same rack.

All InfiniBand addresses must be in the same subnet, with a minimum subnet mask of 255.255.240.0 (or /20). The subnet mask chosen should be wide enough to accommodate possible future expansion of the Oracle Exadata Rack and InfiniBand network.



**Note:**

It is not recommended to use SDP over InfiniBand on Exadata Database Machine.

### 4.9.3 Changing InfiniBand Network Information

This procedure describes how to change the InfiniBand network information.

The procedure described in this section is based on the following assumptions:

- All changes should be done as the `ilom-admin` user using the Integrated Lights Out Manager (ILOM) interface.

- Channel bonding is used for the client access network, such that the NET1 and NET2 interfaces are bonded to create BONDETH0. If channel bonding is not used, then replace BONDETH0 with NET1 in the procedure.
- On Exadata Database Machine X4-2 and later hardware, as of Oracle Exadata System Software release 11.2.3.3.0, the name used for InfiniBand bonding changed from BONDIB0 to IB0 and IB1. These interfaces are changed the same way as the `ifcfg-bondib0` interface.
- As of Oracle Exadata System Software release 11.2.2.1.0, the names used for bonding changed. The names are BONDIB0 for the InfiniBand bonding and BONDETH0 for Ethernet bonding. In earlier releases, the names were BOND0 and BOND1, respectively.
- The procedure uses the `dcli` utility and the `root` user. This significantly reduces the overall time to complete the procedure by running the commands in parallel on the database servers.
- The `dcli` utility requires SSH user-equivalence. If SSH user-equivalence is not configured, then some commands must be run explicitly on each database server.
- The database group file, `dbs_group`, must exist and be located in the `/root` directory.
- Ensure recent backups of the Oracle Cluster Registry (OCR) exist before changing the InfiniBand network information. OCR backups are located in the `$Grid_home/cdata/cluster-name` directory, where `Grid_home` represents the location of your Oracle Grid Infrastructure software installation.
- Starting with Oracle Grid Infrastructure 11g release 2 (11.2), the private network configuration is stored in the Grid Plug and Play (GPNP) profile as well as the OCR. If the GPNP definition is not correct, then Oracle Clusterware CRS does not start. Take a backup of the GPNP profile on all nodes before changing the InfiniBand network information using the following commands:

```
$ cd $Grid_home/gpnp/hostname/profiles/peer/
$ cp -p profile.xml profile.xml.bk
```

1. Determine if the `CLUSTER_INTERCONNECT` parameter is used in the Oracle Database and Oracle ASM instances.

```
SQL> SELECT inst_id, name,value FROM gv$parameter WHERE name = \
'cluster_interconnects';
```

If the `CLUSTER_INTERCONNECT` parameter is set in OCR, then no value is returned. If the `CLUSTER_INTERCONNECT` parameter is defined in the server parameter file (SPFILE), then the query returns an IP addresses for each instance, and they need to be changed to new IP addresses.

The following is an example of the commands to change the IP addresses for the Oracle ASM instances. In the example, the IP address 192.168.10.1 is the new IP address assigned to BONDIB0 on the server where the +ASM1 instance runs, 192.168.10.2 is the IP address for BONDIB0 on the server where the +ASM2 instance runs, and so on.

```
ALTER SYSTEM SET CLUSTER_INTERCONNECTS='192.168.10.1' SCOPE=SPFILE
SID='+ASM1';
ALTER SYSTEM SET CLUSTER_INTERCONNECTS='192.168.10.2' SCOPE=SPFILE
SID='+ASM2';
ALTER SYSTEM SET CLUSTER_INTERCONNECTS='192.168.10.3' SCOPE=SPFILE
```

```
SID='+ASM3';
...
```

Use a similar command to change the IP addresses for each Oracle Database instance that was returned.

2. Verify the assignment of the new InfiniBand network information for all servers.

Verification should include the InfiniBand IP addresses, netmask, broadcast, and network IP information.

3. Shut down all cluster-managed services on each database server as the `oracle` user.

```
$ srvctl stop home -o db_home -s state_filename -n node_name
```

In the preceding command, `db_home` is the full directory name for the Oracle Database home directory, `state_filename` is the path name where you want the state file to be written, and `node_name` is the name of the database server. The following is an example of the command:

```
$ srvctl stop home -o /u01/app/oracle/product/11.2.0.3/dbhome_1 -s \
/tmp/dm02db01_dbhome -n dm02db01
```

In the preceding example, `/u01/app/oracle/product/11.2.0.3/dbhome_1` is the Oracle Database home directory, `/tmp/dm02db01_dbhome` is the state file name, and `dm02db01` is the name of the database server.

4. Modify the cluster interconnect interface to use the BONDIB0 interface on the first database server.

 **Note:**

At this point, only Oracle Clusterware, Oracle Clusterware CRS, and Oracle ASM instances are started.

- a. Log in as the `oracle` user.
- b. Set `$ORACLE_HOME` to the Oracle Grid Infrastructure home.
- c. Set the base for the `ORACLE_SID` environment variable.

The `ORACLE_HOME` environment variable must be set to the Oracle Grid Infrastructure home.

```
$ ORACLE_SID=+ASM1
```

- d. List the available cluster interfaces.

```
$ oifcfg iflist
```

The following is an example of the output:

```
bondeth0 10.128.174.160
bondeth1 10.128.176.0
eth0 10.128.174.128
ib0 192.168.160.0
ib0 169.254.0.0
ib1 192.168.160.0
ib1 169.254.128.0
```

- e. List the currently-assigned cluster interfaces.

```
$ oifcfg getif
```

The following is an example of the output:

```
bondeth0 10.204.76.0 global public
ib0 192.168.16.0 global cluster_interconnect,asm
ib1 192.168.16.0 global cluster_interconnect,asm
```

- f. Assign the ib0 and ib1 interfaces new IP addresses as global cluster interconnect interfaces.

```
oifcfg setif -global ib0/192.168.8.0:cluster_interconnect
oifcfg setif -global ib1/192.168.8.0:cluster_interconnect
```

- g. List the current interfaces.

```
$ oifcfg getif
```

The following is an example of the output:

```
bondeth0 10.128.174.160 global public
ib0 192.168.8.0 global cluster_interconnect
ib1 192.168.8.0 global cluster_interconnect
```

The old private interface is removed at a later time.

5. Shut down Oracle Clusterware and Oracle Clusterware CRS on each database server.

- a. Log in as the `root` user.
- b. Shut down Oracle Clusterware CRS on each database server using the following command:

```
Grid_home/grid/bin/crsctl stop crs -f
```

- c. Disable automatic Oracle Clusterware CRS restart on each database server.

```
Grid_home/grid/bin/crsctl disable crs
```

6. Change the InfiniBand IP addresses on each Oracle Exadata Storage Server.

- a. Log in as the `root` user.

**b. Run the following commands:**

```
cellcli -e alter cell shutdown services all
 Stopping the RS, CELLSRV, and MS services... The SHUTDOWN of
services was successful.

service ocrvottargetd stop

ipconf
```

For the `service` command, respond to the prompts to change the BONDIB0 information. The following is an example of the prompts and responses for the `ipconf` command. Changes are applied after the prompt for basic Integrated Lights Out Manager (ILOM) settings.

```
Logging started to /var/log/cellos/ipconf.log
Interface ib0 is Linked. hca: mlx4_0
Interface ib1 is Linked. hca: mlx4_0
Interface eth0 is Linked. driver/mac: ixgbe/00:00:00:00:cd:01
Interface eth1 is ... Unlinked. driver/mac: ixgbe/
00:00:00:00:cd:02
Interface eth2 is ... Unlinked. driver/mac: ixgbe/
00:00:00:00:cd:03
Interface eth3 is ... Unlinked. driver/mac: ixgbe/
00:00:00:00:cd:04
```

```
Network interfaces
Name State IP address Netmask
Gateway Net type Hostname
ib0 Linked
ib1 Linked
eth0 Linked
eth1 Unlinked
eth2 Unlinked
eth3 Unlinked
```

```
Warning. Some network interface(s) are disconnected. Check
cables and swiches and retry
Do you want to retry (y/n) [y]: n
```

```
The current nameserver(s): 192.0.2.10 192.0.2.12 192.0.2.13
Do you want to change it (y/n) [n]:
The current timezone: America/Los_Angeles
Do you want to change it (y/n) [n]:
The current NTP server(s): 192.0.2.06 192.0.2.12 192.0.2.13
Do you want to change it (y/n) [n]:
```

```
Network interfaces
Name State IP address Netmask
Gateway Net type Hostname
eth0 Linked 192.0.2.151 255.255.252.0 192.0.2.15
Management myg.example.com
eth1 Unlinked
eth2 Unlinked
eth3 Unlinked
```

```

bondib0 ib0,ib1 192.168.13.101 255.255.252.0
Private myg-priv.example.com
Select interface name to configure or press Enter to continue: bondib0
Selected interface. bondib0
IP address or none [192.168.13.101]: 192.168.10.3
Netmask [255.255.252.0]:255.255.248.0
Fully qualified hostname or none [myg-priv.example.com]:
Continue configuring or re-configuring interfaces? (y/n) [y]: n

Select canonical hostname from the list below
1: myg.example.com
2: myg-priv.example.com
Canonical fully qualified domain name [1]:

Select default gateway interface from the list below
1: eth0
Default gateway interface [1]:

Canonical hostname: myg.example.com
Nameservers: 192.0.2.10 192.0.2.12 192.0.2.13
Timezone: America/Los_Angeles
NTP servers: 192.0.2.06 192.0.2.12 192.0.2.13
Default gateway device: eth0
Network interfaces
Name State IP address Netmask Gateway
Net type Hostname
eth0 Linked 192.0.2.151 255.255.252.0 192.0.2.15
Management myg.example.com
eth1 Unlinked
eth2 Unlinked
eth3 Unlinked
bondib0 ib0,ib1 192.168.10.3 255.255.248.0
Private myg-priv.example.com
Is this correct (y/n) [y]:

Do you want to configure basic ILOM settings (y/n) [y]: n

Starting the RS services...
Getting the state of RS services... running

Starting MS services...
The STARTUP of MS services was successful.
A restart of all services is required to put new network
configuration into
effect. MS-CELLSRV communication may be hampered until restart.
Cell myg successfully altered

Stopping the RS, CELLSRV, and MS services...
The SHUTDOWN of services was successful.
ipaddress1=192.168.10.3/21

```

**c. Restart the Oracle Exadata Storage Server.**

```
shutdown -r now
```



7. Restart the cell services.

```
cellcli -e alter cell restart services all
```

8. Verify the newly-assigned InfiniBand address on Oracle Exadata Storage Server.

```
cellcli -e list cell detail | grep ipaddress1
```

The following is an example of the output:

```
ipaddress1: 192.168.10.3/21
```

9. Change the InfiniBand IP addresses on each database server.

- a. Log in as the `root` user.
- b. Change to the `/etc/sysconfig/network-scripts` directory.
- c. Copy the `ifcfg-bondib0` file.

The copied file name must not start with `ifcfg`.

```
cp ifcfg-bondib0 orig_ifcfg-bondib0
```

- d. Edit the `ifcfg-bondib0` file to update the `IPADDR`, `NETMASK`, `NETWORK` and `BROADCAST` fields.

Example of original `ifcfg-bondib0` file:

```
DO NOT REMOVE THESE LINES
%GENERATED BY CELL%
DEVICE=bondib0
USERCTL=no
BOOTPROTO=none
ONBOOT=yes
IPADDR=192.168.20.8
NETMASK=255.255.248.0
NETWORK=192.168.16.0
BROADCAST=192.168.23.255
BONDING_OPTS="mode=active-backup miimon=100 downdelay=5000
updelay=5000"
IPV6INIT=no
MTU=65520
```

Example of updated `ifcfg-bondib0` file:

```
DO NOT REMOVE THESE LINES
%GENERATED BY CELL%
DEVICE=bondib0
USERCTL=no
BOOTPROTO=none
ONBOOT=yes
IPADDR=192.168.10.8
NETMASK=255.255.248.0
NETWORK=192.168.8.0
```

```
BROADCAST=192.168.15.255
BONDING_OPTS="mode=active-backup miimon=100 downdelay=5000
updelay=5000"
IPV6INIT=no
MTU=65520
```

 **Note:**

The MTU size for the InfiniBand interfaces on the database servers should be set as follows:

- For Oracle Exadata System Software release 11.2.3.3 and later, set the MTU size to 7000.
- For Oracle Exadata System Software releases earlier than release 11.2.3.3, set the MTU size to 65520 to ensure a high transfer rate to external devices using TCP/IP over InfiniBand such as media servers or NFS servers.

- e. Restart the database server.

```
shutdown -r now
```

- f. Verify the InfiniBand IP address information.

```
ifconfig -a
```

The following is an example of the BONDIB0 information. It shows the updated InfiniBand network information:

```
inet addr:192.168.10.8 Bcast:192.168.15.255 Mask:255.255.248.0
```

10. Update the `cellinit.ora` and `cellip.ora` files on each database server.

 **Note:**

Do not edit the `cellinit.ora` or `cellip.ora` files when the database or Oracle ASM instance are running. To make changes to the files, perform a procedure similar to the following:

- a. i. Create a copy of the file.

```
cp cellinit.ora cellinit.new
```

- ii. Edit the `cellinit.new` file with a text editor.

- iii. Replace the old `cellinit.ora` file with the updated `cellinit.new` file.

```
mv cellinit.new cellinit.ora
```

- a. Log in as the `root` user.
- b. Change to the `/etc/oracle/cell/network-config` directory.
- c. Make a backup copy of the `cellip.ora` file.

```
cp cellip.ora orig_cellip.ora
```

 **Note:**

If you are using SSH user-equivalence, then the `dcli` utility can be used. The following is an example of the `dcli` command:

```
dcli -l root -g /root/dbs_group "cp cellip.ora
orig_cellip.ora"
```

- d. Make a backup copy of the `cellinit.ora` file.

The following is an example of the command:

```
cp cellinit.ora orig_cellinit.ora
```

 **Note:**

If you are using SSH user-equivalence, then the `dcli` utility can be used. The following is an example of the `dcli` command:

```
dcli -l root -g /root/dbs_group "cp cellinit.ora \
orig_cellinit.ora"
```

- e. Change the InfiniBand IP addresses in the `cellip.ora` file.

Example of original file:

```
cell="192.168.20.1"
cell="192.168.20.2"
cell="192.168.20.3"
cell="192.168.20.4"
cell="192.168.20.5"
cell="192.168.20.6"
cell="192.168.20.7"
```

Example of updated file:

```
cell="192.168.10.1"
cell="192.168.10.2"
cell="192.168.10.3"
cell="192.168.10.4"
```

```
cell="192.168.10.5"
cell="192.168.10.6"
cell="192.168.10.7"
```

 **Note:**

If you are using SSH user-equivalence, then the `dcli` utility can be used to copy the updated file from the first database server to the other database servers. The following is an example of using the `dcli` command:

```
dcli -l root -g /root/dbs_group -f \
/etc/oracle/cell/network-config/cellip.ora

dcli -l root -g /root/dbs_group "mv /root/cellip.ora \
/etc/oracle/cell/network-config/"
```

- f. Change the InfiniBand IP addresses in the `cellinit.ora` file.

The file is updated with the subnet ID and its subnet mask.

Example of original file:

```
ipaddress="192.168.20.8/21"
```

Example of updated file:

```
ipaddress="192.168.10.8/21"
```

Update the `cellinit.ora` file on each database server. The contents of the file is specific to the database server. The `dcli` utility cannot be used for this step.

- g. Run the `ALTER DBSERVER` command on each database server to update the `/etc/oracle/cell/network-config/cellinit.ora` file.

```
dbmcli -e alter dbserver interconnect1 = "ib0"
dbmcli -e alter dbserver interconnect2 = "ib1"
dbmcli -e alter dbserver interconnect3 = "ib2"
dbmcli -e alter dbserver interconnect4 = "ib3"
```

11. Update the `/etc/hosts` file on each database server and Oracle Exadata Storage Servers to use the new InfiniBand IP addresses.

- a. Log in as the `root` user.
- b. Make a backup copy of the `/etc/hosts` file.

```
cp /etc/hosts /etc/orig_hosts
```

- c. Change the InfiniBand IP addresses for the database servers and Oracle Exadata Storage Server files.

12. Start Oracle Clusterware as the `root` user on each server.

```
Grid_home/grid/bin/crsctl start crs
```

13. Verify the cluster interconnect is using the RDS protocol on each database server by examining the Oracle ASM `alert.log`.

The log is in the directory `/u01/app/oracle/diag/asm/+asm/+ASM1/trace`. An entry similar to the following should be listed for the most-recent Oracle ASM restart:

```
CELL interconnect IPC version: Oracle RDS/IP (generic)
```

For Oracle Exadata System Software releases 11.2.0.2 and later, the following command can be used to verify cluster interconnect. The command is run as the `oracle` user on each database server.

```
$ORACLE_HOME/bin/skgxpinfo
```

The output from the command should be `rds`.

If the instance is not using the RDS protocol over InfiniBand, then relink the Oracle software using the following steps:

 **Note:**

Do not use the `relink all` command to relink the Oracle software.

- a. As the `oracle` user, shut down any processes using Oracle software.
- b. If you are relinking the Oracle Grid Infrastructure home, then as the `root` user, run one of the following commands. Do not perform this step if you are not relinking the Oracle Grid Infrastructure home.

- For Oracle Grid Infrastructure release 12.2.0.1 or higher:

```
Grid_home/crs/install/rootcrs.sh -unlock
```

- For Oracle Grid Infrastructure release 12.1.0.1 or 12.1.0.2:

```
Grid_home/crs/install/rootcrs.pl -unlock
```

- c. As the `oracle` user, change to the `$ORACLE_HOME/rdbms/lib` directory.
- d. As the `oracle` user, run the following command:

```
$ make -f ins_rdbms.mk ipc_rds ioracle
```

- e. If you are relinking the Oracle Grid Infrastructure home, then as the `root` user, run one of the following commands. Do not perform this step if you are not relinking the Oracle Grid Infrastructure home.

- For Oracle Grid Infrastructure release 12.2.0.1 or higher:

```
Grid_home/crs/install/rootcrs.sh -unlock
Grid_home/bin/crsctl start crs
```

- For Oracle Grid Infrastructure release 12.1.0.1 or 12.1.0.2:

```
Grid_home/crs/install/rootcrs.pl -patch
```

**14.** Start all cluster-managed services using the SRVCTL utility.

- a. Log in as the `oracle` user.
- b. Start the database using the following command, where `Oracle_home` is your Oracle home directory:

```
$ srvctl start home -o Oracle_home \
-s /tmp/dm02db01_dbhome -n dm02db01
```

- c. Verify the database instances are running.

```
$ srvctl status database -d dbm
```

**15.** Verify the Oracle ASM and database instances are using the new network settings.

- a. Log in to an Oracle ASM and database instance using SQL\*Plus.
- b. Query the cluster interconnect information.

```
SQL> SELECT inst_id, name,value FROM gv$parameter WHERE name = \
'cluster_interconnects';
```

**16.** Delete the old private network.

```
$ oifcfg delif -global bondib0/192.168.16.0
```

**17.** Verify that the old interface is not present.

```
$ oifcfg getif
bondeth0 10.204.76.0 global public
bondib0 192.168.8.0 global cluster_interconnect
```

**18.** Enable Oracle Clusterware CRS automatic restart on each database server.

- a. Log in as the `root` user.
- b. Enable Oracle Clusterware CRS.

```
Grid_home/grid/bin/crsctl enable crs
```

 **Note:**

To use the `dcli` utility to enable Oracle Clusterware CRS.

```
dcli -l root -g dbs_group "Grid_home/grid/bin/crsctl \
enable crs"
```

19. Perform a full restart of Oracle Clusterware on all nodes.
20. Perform a health check of Oracle Exadata Rack using the steps described in My Oracle Support Doc ID 1070954.1.

 **Note:**

Oracle EXAchk utility collects data for key software, hardware, and firmware releases, and configuration best practices for Oracle Exadata Rack.

Oracle recommends you periodically review the current data for key components of Oracle Exadata Rack, and compare them to the supported release levels, and recommended best practices.

Oracle EXAchk is not a database, network, or SQL performance analysis tool. It is not a continuous monitoring utility, and does not duplicate other monitoring or alerting tools, such as ILOM, or Oracle Enterprise Manager Cloud Control.

21. Verify the private network configuration using the clusterware verification utility, `cluvfy`.

**Related Topics**

- [Using the ipconf Utility](#)
- [Server Control Utility Reference](#)
- [Oracle Exadata Database Machine exachk or HealthCheck \(My Oracle Support Doc ID 1070954.1\)](#)
- [Cluster Verification Utility \(CLUVFY\) FAQ \(My Oracle Support Doc ID 316817.1\)](#)

## 4.10 Configuring Network Routing on Database Servers

The tasks for network routing are for boot-time routing or real-time routing.

- [About Network Routing on Database Servers](#)  
There are three logical network interfaces configured on the database servers.
- [Task 1: Configure for Boot-Time Routing](#)
- [Task 2: Configure for Real-Time Routing](#)
- [Task 3: Verify Network Routing Rules and Routes](#)
- [Removing Network Routing Configuration for Troubleshooting](#)

- [Returning to Default Routing](#)  
To return to the default network routing, delete the supplemental files from the `/etc/sysconfig/network-scripts` directory, and then restart the server.

## 4.10.1 About Network Routing on Database Servers

There are three logical network interfaces configured on the database servers.

The interface names are:

- Management network: `eth0`
- Client access network: `bond1` or `bondeth0`
- RDMA Network Fabric network: `bond0`, `bondib0`, or `ib0` and `ib1`, or `re0` and `re1`



### Note:

The tasks in this section are for Oracle Exadata Database Servers that were configured prior to Oracle Exadata System Software release 11.2.3.2.1.

Starting with Oracle Exadata System Software release 11.2.2.3.0, connections that come in on the management network have their responses sent out on the management network interface, and connections on the client access network have their responses sent out on the client access network interface.

For Oracle Exadata System Software releases earlier than release 11.2.2.3.0, the default route for outbound traffic not destined for an IP address on the management or private InfiniBand network is sent out using the client access network. The tasks in this section modify the routing such that traffic that comes in on the management network has the responses sent out on the management network. Similarly, traffic coming in on the client network has the responses sent out on the client network.

The tasks for network routing are for boot-time routing or real-time routing. The following apply to both types of routing:

- These tasks are for database servers running a release earlier than Oracle Exadata System Software release 11.2.2.3.0.
- The following sample IP addresses, netmasks, and gateways are used in the tasks:
  - Management network has IP address 10.149.49.12, netmask 255.255.252.0 (network 10.149.48.0/22), and gateway 10.149.48.1.
  - Client access network has IP address 10.204.78.15, netmask 255.255.255.0 (network 10.204.78.0/24), and gateway 10.1.78.1.



### Note:

If the database server has additional networks configured, then files should be set up for the additional networks.



## 4.10.2 Task 1: Configure for Boot-Time Routing

To configure network routing for boot-time routing, rule and routing files must be created for each database server. The rule and routing files must be located in the `/etc/sysconfig/network-scripts` directory on each database server. For each Ethernet interface on the management network that has a configured IP address, the database server must have `route-ethn` and `rule-ethn` files. For each bonded Ethernet interface, the database server must have `route-bondethn` and `rule-bondethn` files. The following are examples of the content in the files:

File	Content
<code>/etc/sysconfig/network-scripts/rule-eth0</code>	from 10.149.49.12 table 220 to 10.149.49.12 table 220
<code>/etc/sysconfig/network-scripts/route-eth0</code>	10.149.48.0/22 dev eth0 table 220 default via 10.149.48.1 dev eth0 table 220
<code>/etc/sysconfig/network-scripts/rule-bondeth0</code>	from 10.204.78.0/24 table 210 to 10.204.78.0/24 table 210
<code>/etc/sysconfig/network-scripts/route-bondeth0</code>	10.204.78.0/24 dev bondeth0 table 210 default via 10.204.78.1 dev bondeth0 table 210

## 4.10.3 Task 2: Configure for Real-Time Routing

To configure the rules on a running system, use the `/sbin/ip` command to create the same configuration that is performed at startup. The following commands result in the same configuration as the boot-time files:

```
/sbin/ip rule add from 10.149.49.12 table 220
/sbin/ip rule add to 10.149.49.12 table 220
/sbin/ip route add 10.149.48.0/22 dev eth0 table 220
/sbin/ip route add default via 10.149.48.1 dev eth0 table 220

/sbin/ip rule add from 10.204.78.0/24 table 210
/sbin/ip rule add to 10.204.78.0/24 table 210
/sbin/ip route add 10.204.78.0/24 dev bondeth0 table 210
/sbin/ip route add default via 10.204.78.1 dev bondeth0 table 210
```

Oracle recommends restarting the database server after running the commands to validate that the boot-time configuration is correct.

## 4.10.4 Task 3: Verify Network Routing Rules and Routes

Use the following command to verify the network routing rules. The command output shows all the rules on the system.

```
/sbin/ip rule list
0: from all lookup 255
32762: from all to 10.204.78.0/24 lookup 210
32763: from 10.204.78.0/24 lookup 210
```

```
32764: from all to 10.149.49.12 lookup 220
32765: from 10.149.49.12 lookup 220
32766: from all lookup main
32767: from all lookup default
```

The default routing table is not changed because two new routing tables are created during the preceding tasks. The new routing tables are used when the rules dictate their use. The following commands show how to check the default and new routing tables:

- To check the default routing table. The following is an example of the command and output.

```
/sbin/ip route list
10.204.78.0/24 dev bondeth0 proto kernel scope link src 10.204.78.15
192.168.10.0/24 dev bondib0 proto kernel scope link src 192.168.10.8
10.149.48.0/22 dev eth0 proto kernel scope link src 10.149.49.12
default via 10.149.52.1 dev bondeth0
```

- To check that the supplemental tables include the table name with the command. The following is an example of the command and output.

```
/sbin/ip route list table 220
10.149.48.0/22 dev eth0 scope link
default via 10.149.48.1 dev eth0
root@dbhost# ip route list table 210
10.204.78.0/24 dev bondeth0 scope link
default via 10.204.78.1 dev bondeth0
```

## 4.10.5 Removing Network Routing Configuration for Troubleshooting

The network routing configuration can be removed to configure or troubleshoot Oracle Exadata Database Machine. Use the following commands to remove the rules and routes:

```
/sbin/ip route del default via 10.149.48.1 dev eth0 table 220
/sbin/ip route del 10.149.48.0/22 dev eth0 table 220
/sbin/ip rule del to 10.149.49.12 table 220
/sbin/ip rule del from 10.149.49.12 table 220

/sbin/ip route del default via 10.204.78.1 dev bondeth0 table 210
/sbin/ip route del 10.204.78.0/24 dev bondeth0 table 210
/sbin/ip rule del to 10.204.78.0/24 table 210
/sbin/ip rule del from 10.204.78.0/24 table 210
```

## 4.10.6 Returning to Default Routing

To return to the default network routing, delete the supplemental files from the `/etc/sysconfig/network-scripts` directory, and then restart the server.

The following is an example of the commands to remove the files, and restart the server:

```
/bin/rm -f /etc/sysconfig/network-scripts/rule-eth0
/bin/rm -f /etc/sysconfig/network-scripts/route-eth0
/bin/rm -f /etc/sysconfig/network-scripts/rule-bondeth0
/bin/rm -f /etc/sysconfig/network-scripts/route-bondeth0
shutdown -r now
```

## 4.11 Changing the DNS Servers

The configuration settings for the Domain Name System (DNS) servers can be changed after initial setup.

All servers and switches in Exadata Database Machine should reference the same DNS servers. All domains that Exadata Database Machine references should be resolvable through each individual DNS server.

The following topics contain the tasks and procedures for setting the Exadata Database Machine servers and switches to the same DNS servers. Oracle recommends changing the servers one at a time.

- [Change the DNS Server Address on the Database Server](#)  
This procedure describes how to change the DNS server address on the database servers.
- [Change the DNS Server on Oracle Exadata Storage Server](#)  
This procedure describes how to change the DNS server on Oracle Exadata Storage Servers.
- [Change the DNS Server Address on the Cisco RoCE Network Fabric Switches](#)  
This procedure describes how to change the DNS server address on the Cisco RoCE Network Fabric switches.
- [Change the DNS Server Address on the Cisco 9300 Series Management Network Switch](#)  
This procedure describes how to change the DNS server address on the Cisco 9300 Series Management Network Switch.
- [Change the DNS Server Address on the Cisco 4948 Ethernet Switch](#)  
This procedure describes how to change the DNS server address on the Cisco 4948 Ethernet switch.
- [Change the DNS Server Address on the InfiniBand Network Fabric Switch](#)
- [Change the DNS Server on the KVM Switch](#)  
This procedure describes how to change the DNS server configuration using the KVM switch.

### 4.11.1 Change the DNS Server Address on the Database Server

This procedure describes how to change the DNS server address on the database servers.

1. If you are using Oracle Exadata System Software 20.1.0 or later, use `ipconf` with the `-update` and `-dns` options to modify the DNS settings.
  - a. Log in to the database server as the `root` user.
  - b. Check to make sure there are no configuration issues with the new settings.

Use the following command, where `IP_addr_list` is a comma-separated list of IP addresses for the DNS servers. If you also want to check the DNS servers for ILOM, then include the `-ilom-dns` parameter and replace `ILOM_DNS_list` with

a comma-separated list of up to 3 IP addresses for the DNS servers. Including the `-dry` parameter means the settings are checked, but not applied.

```
ipconf -update -dns IP_addr_list [-ilom-dns ILOM_DNS_list] -dry
```

- c. Update the DNS settings using the following `ipconf` command:

In the following command, `IP_addr_list` is a comma-separated list of IP addresses for the DNS servers. If you also want to change the DNS servers for ILOM, then include the `-ilom-dns` parameter and replace `ILOM_DNS_list` with a comma-separated list of up to 3 IP addresses for the DNS servers. Include the `-force` parameter to force the update, bypassing all checks.

```
ipconf -update -dns IP_addr_list [-ilom-dns ILOM_DNS_list] [-force]
```

Here is an example of the command and its output.

```
[root@dbm03adm02]# ipconf -update -dns 10.31.138.25,10.231.225.65
[Info]: ipconf command line: ipconf -update -dns 10.31.138.25,10.231.225.65
Logging started to /var/log/cellos/ipconf.log
[Info]: Updating dns/ntp
[Info]: Backup existing cell configuration file /opt/oracle.cellos/cell.conf
to /var/log/exadatatmp/cell.conf_2020_01_13-17_59_44
[Info]: Custom changes have been detected in /etc/resolv.conf
[Info]: Original file /etc/resolv.conf will be saved in /etc/
resolv.conf.backupbyExadata
[Done]: Update cell configuration file /opt/oracle.cellos/cell.conf OK
```

- d. Repeat these steps for each database server.
2. If you are using Oracle Exadata System Software 19.3.x or earlier, then use the following steps to modify the DNS servers.
    - a. Log in to the database server as the `root` user.
    - b. Edit the `/etc/resolv.conf` file.

Set the DNS server and domain name using an editor such as `vi`. There should be a name server line for each DNS server.

```
search example.com
nameserver 10.7.7.3
```

- c. Set the DNS server in the server ILOM.

```
ipmitool sunoem cli 'set /SP/clients/dns nameserver=dns_ip'
```

In the preceding command, `dns_ip` is the IP address of the DNS server. If there is more than one DNS server, then enter a comma-separated list such as `set /SP/clients/dns nameserver=dns_ip1,dns_ip2,dns_ip3`.

- d. Repeat these steps for each database server.

## 4.11.2 Change the DNS Server on Oracle Exadata Storage Server

This procedure describes how to change the DNS server on Oracle Exadata Storage Servers.



### Note:

The NTP settings can also be set during this procedure.

1. Log in to the Oracle Exadata Storage Server as the `root` user.
2. Specify a time interval to repair the disk and bring it back online.

The default `DISK_REPAIR_TIME` attribute value of 3.6 hours should be long enough for most environments.

- a. Check the repair time for all mounted disk groups.

Log in to the Oracle ASM instance and run the following query:

```
SQL> SELECT dg.name,a.value
 FROM V$ASM_DISKGROUP dg, V$ASM_ATTRIBUTE a
 WHERE dg.group_number=a.group_number AND
 a.name='disk_repair_time';
```

- b. Adjust the `DISK_REPAIR_TIME` parameter, if needed.

In the following command, *h.n* is the amount of time in hours, such as 4.6.

```
SQL> ALTER DISKGROUP DATA SET ATTRIBUTE
'DISK_REPAIR_TIME'='h.nH';
```

3. Check that putting the grid disks offline will not cause a problem for Oracle ASM.

```
cellcli -e LIST GRIDDISK ATTRIBUTES
name,asmmodestatus,asmdeactivationoutcome
```

The value `Yes` should be returned for the grid disks. If one or more disks does not return a `Yes` value, then restore data redundancy for the disk group, and repeat the command until all disk groups return a `Yes` value.

4. Inactivate all grid disks on the cell.

```
cellcli -e ALTER GRIDDISK ALL INACTIVE
```

This command may take more than 10 minutes to complete. Inactivating the grid disks automatically sets the disk to `offline` in the Oracle ASM instance.

5. Confirm the grid disks are offline.

- a. Check the status of the grid disks.

```
cellcli -e LIST GRIDDISK ATTRIBUTES name, asmmodestatus, \
asmdeactivationoutcome
```

The output should show `asmmodestatus=OFFLINE` or `asmmodestatus=UNUSED`, and `asmdeactivationoutcome=Yes` for all grid disks.

- b. List the grid disk to confirm that they are offline.

```
cellcli -e LIST GRIDDISK
```

6. Shut down the cell services and `ocrvottargetd` service.

```
cellcli -e ALTER CELL SHUTDOWN SERVICES ALL
service ocrvottargetd stop
```

 **Note:**

The `ocrvottargetd` service is not included in some releases.

7. Use the `ipconf` utility to change the DNS settings.

- a. If you are using Oracle Exadata System Software 20.1.0 or later, use `ipconf` with the `-update` and `-dns` options to modify the DNS settings as follows:

- i. Check to make sure there are no configuration issues with the new settings.

Use the following command, where `IP_addr_list` is a comma-separated list of IP addresses for the DNS servers. If you also want to check the DNS servers for ILOM, then include the `-ilom-dns` parameter and replace `ILOM_DNS_list` with a comma-separated list of up to 3 IP addresses for the DNS servers. Including the `-dry` parameter means the settings are checked, but not applied.

 **Note:**

If you use host names for the DNS servers instead of IP addresses, then the cellwall service will fail when restarted. Use only IP addresses when defining NTP and DNS servers.

```
ipconf -update -dns IP_addr_list [-ilom-dns ILOM_DNS_list] -dry
```

- ii. Update the DNS settings using the following `ipconf` command:

In the following command, `IP_addr_list` is a comma-separated list of IP addresses for the DNS servers. If you also want to change the DNS servers for ILOM, then include the `-ilom-dns` parameter and replace `ILOM_DNS_list` with a comma-

separate list of up to 3 IP addresses for the DNS servers. Include the `-force` parameter to force the update, bypassing all checks.

```
ipconf -update -dns IP_addr_list [-ilom-dns ILOM_DNS_list]
[-force]
```

Here is an example of the command and its output.

```
[root@dbm03celadm06]# ipconf -update -dns
10.31.138.25,10.231.225.65
[Info]: ipconf command line: ipconf -update -dns
10.31.138.25,10.231.225.65
Logging started to /var/log/cellos/ipconf.log
[Info]: Updating dns/ntp
[Info]: Backup existing cell configuration file /opt/
oracle.cellos/cell.conf
to /var/log/exadatatmp/cell.conf_2020_01_13-17_59_44
[Info]: Custom changes have been detected in /etc/resolv.conf
[Info]: Original file /etc/resolv.conf will be saved in /etc/
resolv.conf.backupbyExadata
[Done]: Update cell configuration file /opt/oracle.cellos/
cell.conf OK
```

- b.** If you are using Oracle Exadata System Software 19.3.x or earlier, then run the `ipconf` utility and follow the prompts to change the DNS settings.

```
/usr/local/bin/ipconf
```

- 8.** Restart the cell services and `ocrvottargetd` service.

```
service ocrvottargetd start
cellcli -e ALTER CELL STARTUP SERVICES ALL
```

The server does not need to reboot.

- 9.** Activate the grid disks when the cell comes online.

```
cellcli -e ALTER GRIDDISK ALL ACTIVE
```

- 10.** Verify the disks are active.

```
cellcli -e LIST GRIDDISK
```

The output should show `active`.

- 11.** Verify the grid disk status.

- a.** Check that all grid disks are online.

```
cellcli -e LIST GRIDDISK ATTRIBUTES name, asmmodestatus
```

- b.** Wait for Oracle ASM synchronization to complete for all grid disks.

Each disk will go to a `SYNCING` state first then `ONLINE`.

The following is an example of the output:

```
DATA_CD_00_dm01cel01 ONLINE
DATA_CD_01_dm01cel01 SYNCING
DATA_CD_02_dm01cel01 OFFLINE
DATA_CD_03_dm01cel01 OFFLINE
DATA_CD_04_dm01cel01 OFFLINE
DATA_CD_05_dm01cel01 OFFLINE
DATA_CD_06_dm01cel01 OFFLINE
DATA_CD_07_dm01cel01 OFFLINE
DATA_CD_08_dm01cel01 OFFLINE
DATA_CD_09_dm01cel01 OFFLINE
DATA_CD_10_dm01cel01 OFFLINE
DATA_CD_11_dm01cel01 OFFLINE
```

Oracle ASM synchronization is complete when all grid disks show `asmmodestatus=ONLINE`.

12. Repeat this procedure for each Oracle Exadata Storage Server.

### 4.11.3 Change the DNS Server Address on the Cisco RoCE Network Fabric Switches

This procedure describes how to change the DNS server address on the Cisco RoCE Network Fabric switches.

1. Access the switch using SSH, and log in as the `admin` user with the administrator password.

#### Note:

If SSH has not been configured, then use Telnet to access the switch as the `admin` user.

2. Review the current configuration.

```
Switch# show running-config
...
ip domain-name example.com
ip name-server 192.0.2.2 198.51.100.4 203.0.113.2 use-vrf management
...
```

3. Erase the current DNS server information.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# no ip name-server 192.0.2.2 use-vrf management
Switch(config)# no ip name-server 198.51.100.4 use-vrf management
Switch(config)# no ip name-server 203.0.113.2 use-vrf management
Switch(config)# end
Switch# copy running-config startup-config
[#####] 100%
```



```
Copy complete, now saving to disk (please wait)...
Copy complete.
```

 **Note:**

Each current DNS IP address to be changed needs to be erased. Invalid IP addresses must also be erased.

4. Set the domain name and then configure up to three DNS servers, as shown in the following example:

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# ip domain-name example.com
Switch(config)# ip name-server 192.0.2.3 use-vrf management
Switch(config)# ip name-server 198.51.100.5 use-vrf management
Switch(config)# ip name-server 203.0.113.1 use-vrf management
Switch(config)# end
Switch# copy running-config startup-config
[#####] 100%
Copy complete, now saving to disk (please wait)...
Copy complete.
```

5. Verify the changes.

```
Switch# show running-config
```

The command output should include entries for the new DNS servers.

For example:

```
!Command: show running-config
...
ip domain-name example.com
ip name-server 192.0.2.3 198.51.100.5 203.0.113.1 use-vrf management
...
```

6. Exit the session.

```
Switch# exit
```

## 4.11.4 Change the DNS Server Address on the Cisco 9300 Series Management Network Switch

This procedure describes how to change the DNS server address on the Cisco 9300 Series Management Network Switch.

1. Access the switch using SSH, and log in as the `admin` user with the administrator password.

 **Note:**

If SSH has not been configured, then use Telnet to access the switch as the admin user.

2. Review the current configuration.

```
Switch# show running-config
...
ip domain-name example.com
ip name-server 192.0.2.2 198.51.100.4 203.0.113.2
...
```

3. Erase the current DNS server information.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# no ip name-server 192.0.2.2
Switch(config)# no ip name-server 198.51.100.4
Switch(config)# no ip name-server 203.0.113.2
Switch(config)# end
Switch# copy running-config startup-config
[#####] 100%
Copy complete, now saving to disk (please wait)...
Copy complete.
```

 **Note:**

Each current DNS IP address to be changed needs to be erased. Invalid IP addresses must also be erased.

4. Set the domain name and then configure up to three DNS servers, as shown in the following example:

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# ip domain-name example.com
Switch(config)# ip name-server 192.0.2.3
Switch(config)# ip name-server 198.51.100.5
Switch(config)# ip name-server 203.0.113.1
Switch(config)# end
Switch# copy running-config startup-config
[#####] 100%
Copy complete, now saving to disk (please wait)...
Copy complete.
```

5. Verify the changes.

```
Switch# show running-config
```

The command output should include entries for the new DNS servers.

For example:

```
!Command: show running-config
...
ip domain-name example.com
ip name-server 192.0.2.3 198.51.100.5 203.0.113.1
...
```

6. Exit the session.

```
Switch# exit
```

## 4.11.5 Change the DNS Server Address on the Cisco 4948 Ethernet Switch

This procedure describes how to change the DNS server address on the Cisco 4948 Ethernet switch.

1. Access the switch using one of the following methods, based on the firmware release:

- Firmware release 12.2 or later:

Access the switch using SSH, and log in as the `admin` user with the administrator password.

 **Note:**

If SSH has not been configured, then use Telnet to access the switch as the `admin` user.

- Firmware earlier than release 12.2:

Access the switch using Telnet, and log in as the administrator using the administrative password.

2. Change to `enable` mode.

```
Switch> enable
```

When prompted for a password, use the administrator password.

3. Review the current configuration.

```
Switch# show running-config
```

4. Erase the current DNS server information.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# no ip name-server 192.0.2.2
Switch(config)# no ip name-server 198.51.100.4
```

```
Switch(config)# no ip name-server 203.0.113.2
Switch(config)# end
Switch# write memory
Building configuration...
Compressed configuration from 2603 bytes to 1158 bytes [OK]
```

 **Note:**

Each current DNS IP address to be changed needs to be erased. Invalid IP addresses must also be erased.

5. Set the domain name and then configure up to three DNS servers, as shown in the following example:

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# ip domain-name example.com
Switch(config)# ip name-server 192.0.2.3
Switch(config)# ip name-server 198.51.100.5
Switch(config)# ip name-server 203.0.113.1
Switch(config)# end
Switch# write memory
Building configuration...
Compressed configuration from 2603 bytes to 1158 bytes [OK]
```

6. Verify the changes.

```
Switch# show running-config
```

The command output should include entries for the new DNS servers.

For example:

```
Building configuration...
...
ip domain-name example.com
ip name-server 192.0.2.3
ip name-server 198.51.100.5
ip name-server 203.0.113.1
...
```

7. Save the configuration.

```
Switch# copy running-config startup-config
Destination filename [startup-config]?
Building configuration...
Compressed configuration from 14343 bytes to 3986 bytes[OK]
```

8. Exit the session.

```
Switch# exit
```

## 4.11.6 Change the DNS Server Address on the InfiniBand Network Fabric Switch

All configuration procedures should be done as the `ilom-admin` user using the Integrated Lights Out Manager (ILOM) interface. Use one of the following procedures to change the DNS server, depending on firmware release:

1. If your switch is using firmware 2.0.4 or later:
  - a. Log in to the InfiniBand Network Fabric switch as the `ilom-admin` user.
  - b. Set the DNS address using one of the following options:
    - Using the ILOM web interface:  
Select the **Configuration** tab and set the DNS server addresses.
    - Using the command line interface, set the DNS server using the following command:

```
set /SP/clients/dns nameserver=dns_ip
```

In the preceding command, *dns\_ip* is the new IP address of the DNS server. If there is more than one DNS server, then enter a comma-separated list such as `set /SP/clients/dns nameserver=dns_ip1,dns_ip2,dns_ip3`.

2. If your switch is using firmware earlier than 2.0.4:
  - a. Log in to the InfiniBand Network Fabric switch as the `root` user.
  - b. Edit the `/etc/resolv.conf` file.  
Set the DNS server and domain name using an editor such as `vi`. There should be a line for each DNS server.
  - c. Save the file.

## 4.11.7 Change the DNS Server on the KVM Switch

This procedure describes how to change the DNS server configuration using the KVM switch.

### Note:

- The KVM switch is only available in Oracle Exadata Database Machine X2-2 racks and Oracle Exadata Storage Expansion Racks with Oracle Exadata Storage Server with Sun Fire X4270 M2 Servers.
- The KVM switch does not support NTP.

1. Log in to the KVM switch. You can log in directly on the KVM switch or access the switch using the host name or IP address over the Internet.
2. Select **Appliance** from Unit View.

3. Select **DNS** from Appliance Settings.
4. Select **DNS Configuration**.
5. Enter the DNS configuration. The following configuration options are available:
  - DNS Mode (Manual, DHCP, DHCPv6)
  - DNS Server Addresses (Primary, Secondary, Tertiary)
6. Click **Save**.

## 4.12 Changing the NTP Servers

The configuration settings for the Network Time Protocol (NTP) servers can be changed after initial setup.

All servers and switches in Exadata Database Machine should reference the same NTP servers so that the servers are synchronized to the same time.

The following topics contain the tasks and procedures for setting the Exadata Database Machine servers and switches to the same NTP server addresses. Oracle recommends changing the servers one at a time.



### Note:

- These procedures assume that there is not a large time discrepancy between the two NTP servers. Use the command `ntpq -p` to see if the system is healthy first before performing the NTP server update.
- Up to two NTP servers can be configured for use with Exadata Database Machine.

- [Set the NTP Server Address on the Database Servers](#)  
You can set or change the Network Time Protocol (NTP) server address on the database server of Exadata Database Machine.
- [Set the NTP Server on Oracle Exadata Storage Servers](#)  
You can set or change the Network Time Protocol (NTP) server on Oracle Exadata Storage Servers.
- [Set the NTP Server Address on the Cisco RoCE Network Fabric Switches](#)  
This procedure describes how to change the Network Time Protocol (NTP) server address on the Cisco RoCE Network Fabric switches.
- [Set the NTP Server Address on the Cisco 9300 Series Management Network Switch](#)  
This procedure describes how to change the Network Time Protocol (NTP) server address on the Cisco 9300 Series Management Network Switch.
- [Set the NTP Server Address on the Cisco 4948 Ethernet Switch](#)  
You can set or change the Network Time Protocol (NTP) server on the Cisco 4948 Ethernet switch.
- [Set the NTP Server Address on the InfiniBand Network Fabric Switch](#)  
You can set or change the Network Time Protocol (NTP) server address on the InfiniBand Network Fabric switch.

## 4.12.1 Set the NTP Server Address on the Database Servers

You can set or change the Network Time Protocol (NTP) server address on the database server of Exadata Database Machine.

1. If you are using Oracle Exadata System Software 20.1.0 or later, use `ipconf` to set or change the NTP server.
  - a. Log in to the database server as the `root` user.
  - b. Check to make sure there are no configuration issues with the new settings.

Use the following command, where `IP_addr_list` is a comma-separate list of IP addresses for the NTP servers. If you are also modifying the NTP servers for ILOM, then include the `-ilom-ntp` parameter and replace `ILOM_NTP_list` with a comma-separate list of up to 2 IP addresses for the NTP servers. Including the `-dry` parameter means the settings are checked, but not applied.

```
ipconf -update -ntp IP_addr_list [-ilom-ntp ILOM_NTP_list] -dry
```

- c. Update the NTP settings using the following `ipconf` command:

In the following command, `IP_addr_list` is a comma-separate list of IP addresses for the NTP servers. If you also want to change the NTP servers for ILOM, then include the `-ilom-ntp` parameter and replace `ILOM_NTP_list` with a comma-separate list of up to 3 IP addresses for the NTP servers. Include the `-force` parameter to force the update, bypassing all checks.

```
ipconf -update -ntp IP_addr_list [-ilom-ntp ILOM_NTP_list] [-force]
```

If the timestamp obtained from the new NTP server differs from the current time known to the system by more than 1 second (*time step*), then the command errors out and does not update the NTP settings. You can use the `-force` option with the command line to override this check.

Here is an example of the command and its output.

```
[root@dbm03adm02 oracle.cellos]# ipconf -update -ntp
10.31.138.20,10.31.16.1
-ilom-ntp 10.31.138.20,10.31.16.1
[Info]: ipconf command line: ipconf -update -ntp
10.31.138.20,10.31.16.1
-ilom-ntp 10.31.138.20,10.31.16.1
Logging started to /var/log/cellos/ipconf.log
[Info]: Updating dns/ntp
[Warning]: ntpd service is not running
[Info]: Backup existing cell configuration file /opt/
oracle.cellos/cell.conf to
/var/log/exadatatmp/cell.conf_2020_01_13-17_54_56
[Info]: Restart ntpd service
Shutting down ntpd:
[OK]
Starting ntpd:
[OK]
```

```
[Done]: Update cell configuration file /opt/oracle.cellos/cell.conf OK
```

- d. Repeat these steps for each database server.
2. If the database server operating system is Oracle Linux 7 but using Oracle Exadata System Software 19.3.x or earlier, follow these instructions:

- a. Stop the time synchronization service on the database server.

```
systemctl stop chronyd
```

- b. Update the `/etc/chrony.conf` file with the IP address of the new NTP server.
- c. Start the time synchronization service on the database server.

```
systemctl start chronyd
```

- d. Repeat Steps 2.a through 2.c for each database server.

3. If the database server operating system is Oracle Linux 5 or 6:

- a. Stop the NTP services on the database server.

```
service ntpd stop
```

- b. Update the `ntp.conf` file with the IP address of the new NTP server.
- c. Start the NTP services on the database server.

```
service ntpd start
```

- d. Repeat Steps 3.a through 3.c for each database server.

## 4.12.2 Set the NTP Server on Oracle Exadata Storage Servers

You can set or change the Network Time Protocol (NTP) server on Oracle Exadata Storage Servers.

### Note:

The DNS settings can also be set during this procedure.

### WARNING:

If you do not complete the steps to ensure taking the grid disks offline will not impact Oracle ASM operation, then you can cause a database outage.

1. Log in to the cell as the `root` user.
2. Check the time interval to repair the disk and bring it back online and specify a new value if needed.

The default `DISK_REPAIR_TIME` attribute value of 3.6 hours should be long enough for most environments.



- a. Check the repair time for all mounted disk groups by logging in to the Oracle ASM instance, and running the following query:

```
SQL> SELECT dg.name,a.value FROM v$asm_diskgroup dg,
v$asm_attribute a \
WHERE dg.group_number=a.group_number AND
a.name='disk_repair_time';
```

- b. If a non-default value has been specified, and the current setting will not give you enough time to implement the NTP server change, then adjust the parameter by issuing a command similar to the following:

```
SQL> ALTER DISKGROUP DATA SET ATTRIBUTE
'DISK_REPAIR_TIME'='3.6H';
```

Specify the time using the format *h.nH*, where *h.n* is the amount of time in hours, such as 4.6.

3. Check that taking the grid disks offline will not cause a problem for Oracle ASM.

```
cellcli -e list griddisk attributes
name,asmmodestatus,asmdeactivationoutcome
```

The `asmdeactivationoutcome` value should be `Yes` for the grid disks. If one or more disks does not return a `Yes` value, then restore data redundancy for the disk group, and repeat the command until all grid disks return a `Yes` value.

4. Inactivate all grid disks on the cell.

```
cellcli -e alter griddisk all inactive
```

This command may take more than 10 minutes to complete. Inactivating the grid disks automatically sets the disks to `offline` in the Oracle ASM instance.

5. Confirm the grid disks are offline.

- a. Check the status of the grid disks.

```
cellcli -e list griddisk attributes name, asmmodestatus, \
asmdeactivationoutcome
```

The output should show `asmmodestatus=OFFLINE` or `asmmodestatus=UNUSED`, and `asmdeactivationoutcome=Yes` for all grid disks.

- b. List the grid disk to confirm that they are offline.

```
cellcli -e list griddisk
```

6. Shut down the cell services and `ocrvottargetd` service.

```
cellcli -e alter cell shutdown services all
service ocrvottargetd stop
```

 **Note:**

The `ocrvottargetd` service is not included in some releases.

7. Use the `ipconf` utility to change the NTP settings.
  - a. If you are using Oracle Exadata System Software 20.1.0 or later, use `ipconf` with the `-update` and `-ntp` options to modify the NTP settings as follows:
    - i. Check to make sure there are no configuration issues with the new settings.

Use the following command, where `IP_addr_list` is a comma-separated list of IP addresses for the NTP servers. If you are also modifying the NTP servers for ILOM, then include the `-ilom-ntp` parameter and replace `ILOM_NTP_list` with a comma-separated list of up to 2 IP addresses for the NTP servers. Including the `-dry` parameter means the settings are checked, but not applied.

 **Note:**

If you use host names for the NTP servers instead of IP addresses, then the `cellwall` service will fail when restarted. Use only IP addresses when defining NTP and DNS servers.

```
ipconf -update -ntp IP_addr_list [-ilom-ntp ILOM_NTP_list] -dry
```

- ii. Update the NTP settings using the following `ipconf` command:

In the following command, `IP_addr_list` is a comma-separated list of IP addresses for the NTP servers. If you also want to change the NTP servers for ILOM, then include the `-ilom-ntp` parameter and replace `ILOM_NTP_list` with a comma-separated list of up to 3 IP addresses for the NTP servers. Include the `-force` parameter to force the update, bypassing all checks.

```
ipconf -update -ntp IP_addr_list [-ilom-ntp ILOM_NTP_list] [-force]
```

If the timestamp obtained from the new NTP server differs from the current time known to the system by more than 1 second (*time step*), then the command fails and does not update the NTP settings. You can use the `-force` option with the command line to override this check.

Here is an example of the command and its output.

```
[root@dbm03adm02 oracle.cellos]# ipconf -update -ntp
10.31.138.20,10.31.16.1
-ilom-ntp 10.31.138.20,10.31.16.1
[Info]: ipconf command line: ipconf -update -ntp
10.31.138.20,10.31.16.1
-ilom-ntp 10.31.138.20,10.31.16.1
Logging started to /var/log/cellos/ipconf.log
[Info]: Updating dns/ntp
[Warning]: ntpd service is not running
[Info]: Backup existing cell configuration file /opt/oracle.cellos/
```

```
cell.conf to
/var/log/exadatatmp/cell.conf_2020_01_13-17_54_56
[Info]: Restart ntpd service
Shutting down ntpd:
[OK]
Starting ntpd:
[OK]
[Done]: Update cell configuration file /opt/oracle.cellos/
cell.conf OK
```

- b.** If you are using Oracle Exadata System Software 19.3.x or earlier, then run the `ipconf` utility and follow the prompts to change the NTP settings.

```
/usr/local/bin/ipconf
```

- 8.** Restart the cell services. The server does not need to reboot.

```
service ocrvottargetd start
cellcli -e alter cell startup services all
```

- 9.** Activate the grid disks when the cell services come online.

```
cellcli -e alter griddisk all active
```

- 10.** Verify the disks are active.

```
cellcli -e list griddisk
```

The output should show `active`.

- 11.** Verify the grid disks are online.

```
cellcli -e list griddisk attributes name, asmmodestatus
```

The following is an example of the output:

```
DATA_CD_00_dm01cel01 ONLINE
DATA_CD_01_dm01cel01 SYNCING
DATA_CD_02_dm01cel01 OFFLINE
DATA_CD_03_dm01cel01 OFFLINE
DATA_CD_04_dm01cel01 OFFLINE
DATA_CD_05_dm01cel01 OFFLINE
DATA_CD_06_dm01cel01 OFFLINE
DATA_CD_07_dm01cel01 OFFLINE
DATA_CD_08_dm01cel01 OFFLINE
DATA_CD_09_dm01cel01 OFFLINE
DATA_CD_10_dm01cel01 OFFLINE
DATA_CD_11_dm01cel01 OFFLINE
```

Each disk will go to a `SYNCING` state first then `ONLINE`. Wait for Oracle ASM synchronization to complete for all grid disks.

Oracle ASM synchronization is complete when all grid disks show a value of `ONLINE` for `asmmodestatus`.

12. Repeat this procedure for each Oracle Exadata Storage Server.

#### Related Topics

- [Change the DNS Server on Oracle Exadata Storage Server](#)  
This procedure describes how to change the DNS server on Oracle Exadata Storage Servers.

## 4.12.3 Set the NTP Server Address on the Cisco RoCE Network Fabric Switches

This procedure describes how to change the Network Time Protocol (NTP) server address on the Cisco RoCE Network Fabric switches.

1. Access the switch using SSH, and log in as the `admin` user with the administrator password.

#### Note:

If SSH has not been configured, then use Telnet to access the switch as the `admin` user.

2. Review the current configuration.

```
Switch# show running-config
```

The command output includes entries for the current NTP servers.

For example:

```
...
ntp server 10.10.10.1 prefer use-vrf management
ntp server 10.8.8.1 use-vrf management
...
```

3. Erase the current NTP server configuration.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# no ntp server 10.10.10.1
Switch(config)# no ntp server 10.8.8.1
Switch(config)# end
Switch# copy running-config startup-config
[#####] 100%
Copy complete, now saving to disk (please wait)...
Copy complete.
```

 **Note:**

Each current NTP IP address being changed needs to be erased. Invalid IP addresses must also be erased.

**4. Configure up to two NTP servers.**

In this example, the new NTP server IP addresses are 10.7.7.1 and 10.9.9.1.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# feature ntp
Switch(config)# ntp server 10.7.7.1 prefer use-vrf management
Switch(config)# ntp server 10.9.9.1 use-vrf management
Switch(config)# clock protocol ntp
Switch(config)# end
```

**5. Verify the changes.**

```
Switch# show running-config
```

The command output should include entries for the new NTP servers.

For example:

```
...
ntp server 10.7.7.1 prefer use-vrf management
ntp server 10.9.9.1 use-vrf management
...
```

**6. Save the configuration.**

```
Switch# copy running-config startup-config
[#####] 100%
Copy complete, now saving to disk (please wait)...
Copy complete.
```

**7. Exit from the session.**

```
Switch# exit
```

## 4.12.4 Set the NTP Server Address on the Cisco 9300 Series Management Network Switch

This procedure describes how to change the Network Time Protocol (NTP) server address on the Cisco 9300 Series Management Network Switch.

1. Access the switch using SSH, and log in as the `admin` user with the administrator password.

 **Note:**

If SSH has not been configured, then use Telnet to access the switch as the admin user.

2. Review the current configuration.

```
Switch# show running-config
```

The command output includes entries for the current NTP servers.

For example:

```
...
ntp server 10.10.10.1 prefer use-vrf default
ntp server 10.8.8.1 use-vrf default
...
```

3. Erase the current NTP server configuration.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# no ntp server 10.10.10.1
Switch(config)# no ntp server 10.8.8.1
Switch(config)# end
Switch# copy running-config startup-config
[#####] 100%
Copy complete, now saving to disk (please wait)...
Copy complete.
```

 **Note:**

Each current NTP IP address being changed needs to be erased. Invalid IP addresses must also be erased.

4. Configure up to two NTP servers.

In this example, the new NTP server IP addresses are 10.7.7.1 and 10.9.9.1.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# feature ntp
Switch(config)# ntp server 10.7.7.1 prefer
Switch(config)# ntp server 10.9.9.1
Switch(config)# clock protocol ntp
Switch(config)# end
```

5. Verify the changes.

```
Switch# show running-config
```

The command output should include entries for the new NTP servers.

For example:

```
...
ntp server 10.7.7.1 prefer use-vrf default
ntp server 10.9.9.1 use-vrf default
...
```

6. Save the configuration.

```
Switch# copy running-config startup-config
[#####] 100%
Copy complete, now saving to disk (please wait)...
Copy complete.
```

7. Exit from the session.

```
Switch# exit
```

## 4.12.5 Set the NTP Server Address on the Cisco 4948 Ethernet Switch

You can set or change the Network Time Protocol (NTP) server on the Cisco 4948 Ethernet switch.

1. Access the switch using one of the following methods, based on the firmware version:
  - a. Firmware versions earlier than version 12.2: Access the switch using Telnet, and log in as the administrator using the administrative password.
  - b. Firmware version 12.2 or later: Access the switch using SSH, and log in as the `admin` user with the admin password.

 **Note:**

If SSH has not been configured, then use Telnet to access the switch as the `admin` user.

2. Change to enable mode. When prompted for a password, use the administrator password.

```
Switch> enable
```

3. Review the current configuration.

```
Switch# show running-config
```

4. Erase the current NTP server configuration.

In this example, the current IP addresses are 10.10.10.1 and 10.8.8.1.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
```

```
Switch(config)# no ntp server 10.10.10.1
Switch(config)# no ntp server 10.8.8.1
Switch(config)# end
Switch# write memory
Building configuration...
Compressed configuration from 2603 bytes to 1158 bytes [OK]
```

 **Note:**

Each current NTP IP address being changed needs to be erased. Invalid IP addresses must also be erased.

5. Configure up to two NTP servers.

```
Switch# configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
Switch(config)# ntp server 10.7.7.1 prefer
Switch(config)# ntp server 10.9.9.1
Switch(config)# end
Switch# write memory
Building configuration...
Compressed configuration from 2603 bytes to 1158 bytes [OK]
```

6. Verify the changes.

```
Switch# show running-config
```

The command output should include entries for the new NTP servers.

For example:

```
Building configuration...
...
ntp server 10.7.7.1 prefer
ntp server 10.9.9.1
...
```

7. Save the configuration.

```
Switch# copy running-config startup-config
Destination filename [startup-config]?
Building configuration...
Compressed configuration from 14343 bytes to 3986 bytes[OK]
```

8. Exit from the session.

```
Switch# exit
```



## 4.12.6 Set the NTP Server Address on the InfiniBand Network Fabric Switch

You can set or change the Network Time Protocol (NTP) server address on the InfiniBand Network Fabric switch.

 **Note:**

Do not manually edit the files on the InfiniBand Network Fabric switches.

1. Log in as the `ilom-admin` user.
2. Set the date, time zone, and NTP server using one of the following methods:
  - Using the Configuration page on the Integrated Lights Out Manager (ILOM) graphical interface.
  - Manually, using the following commands:

```
set /SP/clock timezone=preferred_tz
set /SP/clients/ntp/server/1 address=ntp_ip1
set /SP/clients/ntp/server/2 address=ntp_ip2
set /SP/clock usntpserver=enabled
```

In the preceding commands, *preferred\_tz* is the preferred time zone, and *ntp\_ip1* and *ntp\_ip2* are the NTP server IP addresses. It is not necessary to configure both NTP servers, but at least one should be configured.

## 4.13 Changing the Time Zone Settings

You can change the time zones on Exadata Database Machine after initial configuration and deployment.

The following components need to be modified when changing the time zone settings:

- Storage servers
- Database servers
- RDMA Network Fabric switches
- Ethernet switch

 **Note:**

Cell services and Oracle Clusterware services must be stopped before changing the time zone settings.

- [Change Time Zone Settings on Storage Servers](#)  
Use these steps to change the time zone setting on storage servers.

- [Change Time Zone Settings on the Database Servers](#)  
After modifying the time zone setting on the storage cells, you can change the time zone setting on the database servers.
- [Change Time Zone Settings on the InfiniBand Network Fabric Switches](#)  
You can change the time zone setting on the InfiniBand Network Fabric switches.
- [Change Time Zone Settings on the Cisco RoCE Network Fabric Switches](#)  
You can change the time zone setting on the Cisco RoCE Network Fabric Switches.
- [Change Time Zone Settings on the Cisco Management Network Switch](#)  
You can change the time zone setting on the Ethernet switch.

### 4.13.1 Change Time Zone Settings on Storage Servers

Use these steps to change the time zone setting on storage servers.

Complete the setting changes to all storage servers before changing the settings on the database servers.

1. Log in as the `root` user on the database server node.
2. Stop the Oracle Clusterware stack on all nodes.

Use a command similar to the following, where *Grid\_home* is the location of your Oracle Grid Infrastructure software installation.

```
Grid_home/bin/crsctl stop crs
```

3. Log in as the `root` user on the storage server.
4. Stop the processes on the storage server.

```
cellcli -e alter cell shutdown services all
```

5. Run the `ipconf` script.

```
/opt/oracle.cellos/ipconf
```

- a. Proceed through the script prompts until you get to the time zone prompts. Do not change any other settings.

Each Country is identified by a number and once it is selected, you will see another set of numbers for the different time zones within that country. The following is an example of the time zone prompts for changing the time zone from Antarctica to the United States. The number for the United States is 230.

```
The current timezone: Antarctica/McMurdo
Do you want to change it (y/n) [n]: y
```

```
Setting up local time...
```

```
1) Andorra
2) United Arab Emirates
3) Afghanistan
.
.
```

```
.
15) Aruba
16) Aaland Islands
Select country by number, [n]ext, [l]ast: 230

Selected country: United States (US). Now choose a zone

1) America/New_York
2) America/Detroit
3) America/Kentucky/Louisville
.
.
.
15) America/North_Dakota/New_Salem
16) America/Denver
Select zone by number, [n]ext: 1

Selected timezone: America/New_York
Is this correct (y/n) [y]:
```

- b.** Proceed through the rest of the script prompts, but do not change any other values. Do not change ILOM settings.

After responding to all change requests, the script generates new files.

```
Info. Run /opt/oracle.cellos/validations/init.d/saveconfig
/opt/oracle.cellos/validations/init.d/saveconfig started at
2017_05_12_10_28
Copy cell configs from /opt/oracle/cell/cellsrv/deploy/config
to /opt/oracle.cellos/iso/lastGoodConfig/cell/cellsrv/deploy/config
[INFO] Copying ssh host settings from //etc/ssh to /opt/
oracle.cellos/iso/lastGoodConfig/etc/ssh ...
uid=0(root) gid=0(root)
groups=0(root),1(bin),2(daemon),3(sys),4(adm),6(disk),10(wheel)
```

- 6.** Verify the time zone changes have been propagated to the following files. Examples of the changes are shown for the files.

- /opt/oracle.cellos/cell.conf:

```
$VAR1 = {
 'Hostname' => 'xdserver.example.com',
 'Ntp servers' => [
 '10.141.138.1'
],
 'Timezone' => 'America/New_York',
```

- /etc/sysconfig/clock:

```
ZONE="America/New_York"
UTC=false
ARC=false
#ZONE="Antarctica/McMurdo"
#ZONE="America/New_York"
#ZONE="America/Los_Angeles"
```

The uncommented `ZONE` value (line not preceded by `#`) is the current setting.

- `/etc/localtime:`

Run the command `strings /etc/localtime` command to verify the change. The last line shows the time zone.

```
~^Ip
EST5EDT,M3.2.0,M11.1.0
```

7. Restart the storage server.
8. Use the `date` command to see the current time zone. The following is an example of the output from the command:

```
date
Tue Jan 29 12:44:21 EDT 2019
```

9. Review the `$ADR_BASE/diag/asm/cell/host_name/alert.log` file. The time that processes were restarted should match the current and correct time.
10. Repeat steps 3 through 9 on each cell.

## 4.13.2 Change Time Zone Settings on the Database Servers

After modifying the time zone setting on the storage cells, you can change the time zone setting on the database servers.

Before starting this procedure, you should have already stopped the Oracle Clusterware stack and modified the time zone on the storage cells, as described in [Change Time Zone Settings on Storage Servers](#).

1. As the `root` user, copy the file `/etc/localtime` from any of the storage cells to the database server.

```
scp root@cell_name:/etc/localtime /etc/localtime
```

2. Copy the file `/etc/sysconfig/clock` from any of the storage cells to the database server.

```
scp root@cell_name:/etc/sysconfig/clock /etc/sysconfig/clock
```

3. Change the Oracle Clusterware settings to prevent the CRS stack from starting automatically after restarting the database server.

```
Grid_home/bin/crsctl disable crs
```

4. Reboot the database server.

```
shutdown -r now
```

5. Verify the date has been changed on the database server.

Use the `date` command verify the change for the time zone.

```
date
Tue Jan 29 13:08:46 EDT 2017
```

6. Change the Oracle Clusterware settings to automatically restart the CRS stack after restarting the database server.

```
Grid_home/bin/crsctl enable crs
```

7. Start the CRS stack on the database server.

```
Grid_home/bin/crsctl start crs
```

### 4.13.3 Change Time Zone Settings on the InfiniBand Network Fabric Switches

You can change the time zone setting on the InfiniBand Network Fabric switches.

1. Connect to the InfiniBand Network Fabric switch using SSH as the `ilom-admin` user.

```
ssh -l ilom-admin switch_hostname
```

2. Use the `version` command to check the version of the switch software.

The following is an example of the output from the command:

```
$ version
SUN DCS 36p version: 2.2.2-7
Build time: Aug 26 2016 10:00:25
SP board info:
Manufacturing Date: 2012.06.23
Serial Number: "ABCDE1234"
Hardware Revision: 0x0007
Firmware Revision: 0x0000
BIOS version: SUN0R100
BIOS date: 06/22/2010
```

3. Administer the switch as follows, depending on the software version:
  - If the software version is 1.1.3-2 or later, then administration of the switch is done using ILOM as follows:
    - a. Log in to ILOM using the web address `http://switch_alias`.
    - b. Select the **Configuration** tab. Or, in the left-side navigation, select **ILOM Administration**, then **Date and Time**.
    - c. Select the **Clock** tab.
    - d. Click the **Enabled** check box next to Synchronize Time Using NTP.
    - e. In the Server 1 text box, enter the correct IP address for the primary NTP server.

- f.** In the Server 2 text box, type the IP address of the secondary NTP server you want to use.
  - g.** Click **Save**.
- If the software version is earlier than 1.1.3-2, then log in to the switch using SSH as follows:

- a.** Log in to the switch using the following command:

```
ssh -l root {switch_ip | switch_name}
```

- b.** Stop the `ntpd` daemon using the following command:

```
service ntpd stop
```

- c.** Save a copy of the `/etc/localtime` file using the following command:

```
cp /etc/localtime /etc/localtime.backup
```

- d.** Identify the file in the `/usr/share/zoneinfo` directory for the time zone. The following is an example for the United States:

```
#cd /usr/share/zoneinfo/US
#ls
Alaska Aleutian Arizona Central Eastern East-Indiana Hawaii
Indiana-Starke Michigan Mountain Pacific Samoa
```

- e.** Copy the appropriate file to the `/etc/localtime` directory. The following is an example of the command:

```
cp /usr/share/zoneinfo/US/Eastern /etc/localtime
```

- f.** Manually set the current date and time to values near the current time.

- g.** Synchronize the time to the NTP server for the new time zone using the `date` command with the `MMddHHmmCCyy` format for Month, Day, Hour, Minute, Century, Year. The following is an example of the command:

```
date 013110452013
ntpd -q -g
```

- h.** Validate the date using the following command:

```
date
```

- i.** Restart the `ntpd` daemon using the following command:

```
service ntpd start
```

## 4.13.4 Change Time Zone Settings on the Cisco RoCE Network Fabric Switches

You can change the time zone setting on the Cisco RoCE Network Fabric Switches.

1. Access the switch using SSH, and log in as the `admin` user.
2. Use the `configure terminal` command to begin configuration.
3. Set the clock using the following commands:
  - a. To change the time zone:

```
clock timezone zone hours_offset minutes_offset
```

In the preceding syntax:

- `zone` is the name of the time zone to be displayed when standard time is in effect. The default time zone is UTC.
  - `hours_offset` is the hours offset from UTC
  - `minutes_offset` are the minutes offset from UTC
- b. To set summer time (daylight savings time) in areas where it starts and ends on a particular day of the week each year, use the following command:

```
clock summer-time zone recurring [week day month hh:mm week day month hh:mm[offset]]
```

In the preceding syntax, the values of `week day month hh:mm` are listed twice, once for the starting time and again for the ending time.

- `recurring` specifies that summer time starts and ends on the specified days every year. Summer time is disabled by default. If you specify `clock summer-time zone recurring` without any other parameters, the summer time rules default to the United States rules.
- `week` is the week of the month, between 1 to 5. The first occurrence of week is the start date and the second occurrence is the end date.
- `day` is the day of the week, such as Sunday or Monday. The first occurrence of week is the start date and the second occurrence is the end date.
- `month` is the month, such as January or June. The first occurrence of week is the start date and the second occurrence is the end date.
- `hh:mm` is the time in 24-hour format in hours and minutes, such as 15:42
- `offset` is the number of minutes to add during summer time. The default is 60.

**Example 4-2 Setting the Time Zone on the RoCE Network Fabric Switch**

The following is an example of setting the time zone to US Eastern time with summer time enabled:

```
dbm0sw-rocea0#configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
dbm0sw-rocea0(config)#clock timezone EST -5 0
dbm0sw-rocea0(config)#clock summer-time EDT recurring
dbm0sw-rocea0(config)#end
dbm0sw-rocea0#copy running-config startup-config
Building configuration...
Compressed configuration from 6421 bytes to 2041 bytes[OK]
dbm0sw-rocea0#show clock
12:03:43.516 EDT Wed May 12 2012
dbm0sw-rocea0#
```

## 4.13.5 Change Time Zone Settings on the Cisco Management Network Switch

You can change the time zone setting on the Ethernet switch.

1. Use Telnet to connect to the Ethernet switch.
2. Use the `enable` command to enter privileged mode.
3. Use the `configure terminal` command to begin configuration.
4. Set the clock using the following commands:
  - a. To change the time zone:

```
clock timezone zone hours_offset minutes_offset
```

In the preceding syntax:

- *zone* is the name of the time zone to be displayed when standard time is in effect. The default time zone is UTC.
  - *hours\_offset* is the hours offset from UTC
  - *minutes\_offset* are the minutes offset from UTC
- b. To set summer time (daylight savings time) in areas where it starts and ends on a particular day of the week each year, use the following command:

```
clock summer-time zone recurring [week day month hh:mm week day
month hh:mm[offset]]
```

In the preceding syntax, the values of *week day month hh:mm* are listed twice, once for the starting time and again for the ending time.

- `recurring` specifies that summer time starts and ends on the specified days every year. Summer time is disabled by default. If you specify `clock summer-time zone recurring` without any other parameters, the summer time rules default to the United States rules.



- *week* is the week of the month, between 1 to 5. The first occurrence of week is the start date and the second occurrence is the end date.
- *day* is the day of the week, such as Sunday or Monday. The first occurrence of week is the start date and the second occurrence is the end date.
- *month* is the month, such as January or June. The first occurrence of week is the start date and the second occurrence is the end date.
- *hh:mm* is the time in 24-hour format in hours and minutes, such as 15:42
- *offset* is the number of minutes to add during summer time. The default is 60.

### Example 4-3 Setting the Time Zone on the Ethernet Switch

The following is an example of setting the time zone to US Eastern time with summer time enabled:

```
$ telnet dbmcisco-ip
Connected to switch name
Escape character is '^]'.

User Access Verification

Password:
dmbcisco-ip>enable
Password:
dmbcisco-ip#configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
dmbcisco-ip(config)#clock timezone EST -5 0
dmbcisco-ip(config)#clock summer-time EDT recurring
dmbcisco-ip(config)#end
dmbcisco-ip#write memory
Building configuration...
Compressed configuration from 6421 bytes to 2041 bytes[OK]
dmbcisco-ip#show clock
12:03:43.516 EDT Wed May 12 2012
dmbcisco-ip#
```

## 4.14 Managing the KVM Switch

The KVM switch is only available in Oracle Exadata Database Machine X2-2 racks and Oracle Exadata Storage Expansion Racks with Exadata Storage Server with Sun Fire X4270 M2 Servers.

- [Configuring the KVM Switch](#)  
This procedure describes how to configure the KVM (Keyboard, Video, Mouse) switch.
- [Configuring the KVM Switch to Access a Server](#)
- [Accessing a Server Using the KVM Switch](#)

## 4.14.1 Configuring the KVM Switch

This procedure describes how to configure the KVM (Keyboard, Video, Mouse) switch. The switch is configured with all the connected components powered off.

 **Note:**

The KVM switch is only available in Oracle Exadata Database Machine X2-2 racks and Oracle Exadata Storage Expansion Racks with Exadata Storage Server with Sun Fire X4270 M2 Servers.

1. Pull the KVM tray out from the front of the rack, and open it using the handle.
2. Touch the touch pad.
3. Toggle between the host and KVM interface by pressing the `Ctrl` key on the left side twice, similar to a double-click on a mouse.
4. Select **Target Devices** from the Unit View of the user interface. The number of sessions shown should be 22 for Oracle Exadata Database Machine Full Rack, 11 for Oracle Exadata Database Machine Half Rack, and 5 for Oracle Exadata Database Machine Quarter Rack. The number of sessions should be 18 for Oracle Exadata Storage Expansion Full Rack, 9 for Oracle Exadata Storage Expansion Half Rack, and 4 for Oracle Exadata Storage Expansion Quarter Rack.

 **Note:**

If all sessions are not shown, then select IQ Adaptors from the Ports heading. Click the table heading, and then **Port**, to sort the sessions by port number. Note any missing items. The sessions are numbered from the bottom of the rack to the top.

5. Return to the Target Devices screen.
6. Select **Local** from User Accounts.
7. Click **Admin** under Users.
8. Set a password for the Admin account. Do not modify any other parameters.
9. Click **Save**.
10. Select **Network** from Appliance Settings. The Network Information screen appears.
11. Select IPv4 or IPv6.
12. Enter the values for Address, Subnet, Gateway, and the IP addresses of the DNS servers.
13. Click **Save**.
14. Connect the KVM LAN1 Ethernet port to the management network.

15. Verify the port has been configured correctly by checking the MAC address on the Network Information screen. The address should match the label next to the LAN1/LAN2 ports on the rear of the KVM switch.
16. Select **Overview** from Appliance.
17. Enter a name for the KVM switch.
18. Click **Save**.
19. Restart the KVM switch by selecting **Reboot** under Overview.
20. Examine the firmware version of the switch by selecting **Versions** from Appliance Settings. There are two version numbers shown, Application and Boot, as shown in the following:

```
Required version is:
Application 1.2.10.15038
Boot 1.6.15020
```

 **Note:**

The recommended firmware version is 1.2.8 or later.

If the firmware is 1.2.3 or earlier, then it can be upgraded from a network browser. If it is version 1.2.3 or later, then it can be upgraded from the local keyboard using a flash drive plugged in to the KVM USB port. To upgrade the firmware, do the following:

- a. Select **Overview** from Appliance.
- b. Select **Upgrade Firmware** from the Tools list.
- c. Select the method to upgrade.
- d. Click **Upgrade**.
- e. Confirm the firmware version.

 **See Also:**

Avocent Web site for information about KVM switch Management Information Base (MIB) at <https://www.vertivco.com/en-us/support/software-download/it-management/avocent-mergepoint-unity-switches-software-downloads/>

## 4.14.2 Configuring the KVM Switch to Access a Server

The following procedure describes how to configure the KVM switch to access the servers:

 **Note:**

The KVM switch is only available in Oracle Exadata Database Machine X2-2 racks and Oracle Exadata Storage Expansion Racks with Exadata Storage Server with Sun Fire X4270 M2 Servers.

1. Select **Target Devices** from Unit View.
2. Power on the server. The power button is on the front panel. If the button seems stuck, then use a small tool to loosen the button.
3. Click the system name in the Name column using the left mouse button.
4. Click **Overview**, and overwrite the name with the Oracle standard naming format of customer prefix, node type, and number. For example, `trnace103` has the prefix `trna`, and is storage cell 3 from the bottom of the rack, and `trnadb02` has the prefix `trna`, and is database server 2 from the bottom of the rack.
5. Press **Save**.
6. Repeat steps 2 through 5 for each server in the rack. Each server boots up through BIOS, and boots the operating system with the default factory IP configuration.

### 4.14.3 Accessing a Server Using the KVM Switch

The following procedure describes how to access a server using the KVM switch:

 **Note:**

The KVM switch is only available in Oracle Exadata Database Machine X2-2 racks and Oracle Exadata Storage Expansion Racks with Exadata Storage Server with Sun Fire X4270 M2 Servers.

1. Select **Target Devices** from Unit View.
2. Click the system name in the Name column using the left mouse button.
3. Click the KVM session.

## 4.15 LED Status Descriptions

The LEDs on the Oracle Exadata Rack components help you identify the component that needs servicing.

- [Sun Datacenter InfiniBand Switch 36 Switch LEDs](#)
- [Cisco Nexus 9336C-FX2 Switch LEDs](#)  
This topic describes the color codes of the LEDs on Cisco Nexus 9336C-FX2 switches.

## 4.15.1 Sun Datacenter InfiniBand Switch 36 Switch LEDs

Table 4-1 describes the color codes of the LEDs on Sun Datacenter InfiniBand Switch 36 switches.

**Table 4-1 Sun Datacenter InfiniBand Switch 36 Switch LED Status Descriptions**

Component	LED Status
Sun Datacenter InfiniBand Switch 36 chassis	<ul style="list-style-type: none"> <li>Locator LED is white: It flashes when identifying itself. It is on when there is no function, and off when disabled.</li> <li>Attention LED is amber: There is a fault condition. It flashes when there is no function.</li> <li>OK LED is green: Switch is functioning correctly. It flashes when there is no function.</li> </ul>
Sun Datacenter InfiniBand Switch 36 link status	Link LED is green: It is on when link is established. It is off when link is down, and it flashes when there are symbol errors.
Sun Datacenter InfiniBand Switch 36 network management ports	<ul style="list-style-type: none"> <li>Link speed LED: Green indicates 1000BASE-T. Amber indicates 100BASE-T. Off indicates no link. Flashing indicates no function.</li> <li>Activity LED: Flashing indicates packet activity. On indicates no function. Off indicates no activity.</li> </ul>
Sun Datacenter InfiniBand Switch 36 power supply	<ul style="list-style-type: none"> <li>OK LED is green: Indicates 12 VDC is supplied. Flashing indicates no function.</li> <li>Attention LED is amber: There is a fault, and 12 VDC shut down. Flashing indicates no function.</li> <li>AC LED is green: AC power is present and good. Flashing indicates no function.</li> </ul>

## 4.15.2 Cisco Nexus 9336C-FX2 Switch LEDs

This topic describes the color codes of the LEDs on Cisco Nexus 9336C-FX2 switches.

**Table 4-2 Cisco Nexus 9336C-FX2 Switch LED Status Descriptions**

Component	LED Status	Location
Switch chassis Beacon (BCN) LED	<ul style="list-style-type: none"> <li><b>Flashing blue:</b> Operator activated LED to identify the switch.</li> <li><b>Off:</b> Default status, switch is not being identified.</li> </ul>	The BCN LED is located on the left side of the front of the switch
Switch chassis Status (STS) LED	<ul style="list-style-type: none"> <li><b>Flashing amber:</b> Switch is booting up.</li> <li><b>Amber or Red:</b> Temperature of the switch has exceeded the minor alarm threshold</li> <li><b>Off:</b> The switch has no power.</li> </ul>	The STS LED is located on the left side of the front of the switch

Table 4-2 (Cont.) Cisco Nexus 9336C-FX2 Switch LED Status Descriptions

Component	LED Status	Location
Switch chassis Environment (ENV) LED	<ul style="list-style-type: none"> <li>• <b>Green:</b> The fans and power supplies are functioning correctly.</li> <li>• <b>Amber:</b> At least one power supply or fan has stopped functioning.</li> </ul>	The ENV LED is located on the left side of the front of the switch
Cisco Nexus 9336C-FX2 network ports	<ul style="list-style-type: none"> <li>• <b>Green:</b> The port admin state is enabled, Small form-factor pluggable (SFP) transceiver is present, and the interface is connected.</li> <li>• <b>Amber:</b> The port admin state is disabled, the SFP is absent, or both.</li> <li>• <b>Off:</b> The port admin state is enabled, SFP is present, but the interface is not connected.</li> </ul>	The port LEDs appear as triangles pointing up or down to the nearest port.
Switch chassis lane link	<ul style="list-style-type: none"> <li>• <b>One or more are lit:</b> The lane is being checked.</li> <li>• <b>None:</b> All four lanes are being checked.</li> </ul>	The Lane Link LEDs indicate which of the break out lanes are being checked.
Power supply OK and Fault LEDs	<ul style="list-style-type: none"> <li>• <b>OK-Green, Fault-Off:</b> Power supply is functioning correctly.</li> <li>• <b>OK-Flashing Green, Fault-Off:</b> Power supply is connected to a power source but not outputting power to the switch. The power supply might not be installed in the chassis.</li> <li>• <b>OK-Off, Fault-Off:</b> Power supply is not receiving power or is shut down.</li> <li>• <b>OK-Green, Fault-Flashing Amber:</b> Power supply warning, indicating possible high voltage, high power, low voltage, power supply warning condition, or power supply fan not operating correctly.</li> <li>• <b>OK-Green, Fault-Amber:</b> Power supply failure.</li> </ul>	The power supply LEDs are located on the left front portion of the power supply. Combinations of states indicated by the Okay and Fault LEDs indicate the status for the module.
Fan Module STS LED	<ul style="list-style-type: none"> <li>• <b>Green:</b> Fan module is operating correctly.</li> <li>• <b>Red:</b> Fan module is not operational.</li> <li>• <b>Off:</b> Fan module does not have power.</li> </ul>	The fan module LED is located below the air holes on the front of the module.

# 5

## Managing Oracle VM User Domains

Oracle VM is a Xen-based virtualization technology, which is used across Oracle Exadata Database Machine systems that use InfiniBand Network Fabric.

- [Oracle VM and Exadata Database Machine](#)  
When deploying Exadata Database Machine, you can decide to implement Oracle VM on the database servers.
- [Migrating a Bare Metal Oracle RAC Cluster to an Oracle RAC Cluster in Oracle VM](#)
- [Showing Running Domains](#)
- [Monitoring a User Domain Console](#)
- [Monitoring Oracle VMs with Oracle Enterprise Manager](#)  
The Exadata plug-in for Oracle Enterprise Manager discovers, manages, and monitors virtualized Oracle Exadata Database Machine in conjunction with the Virtualization Infrastructure plug-in of Oracle Enterprise Manager.
- [Starting a User Domain](#)
- [Managing Automatic Startup of Oracle VM User Domains](#)  
By default, when you create a user domain, it is configured to automatically start when the management domain is started. You can enable and disable this feature as needed.
- [Shutting Down a User Domain From Within the User Domain](#)
- [Shutting Down a User Domain From Within the Management Domain](#)
- [Backing Up and Restoring Oracle Databases on Oracle VM User Domains](#)  
Backing up and restoring Oracle databases on Oracle VM user domains is the same as backing up and restoring Oracle databases on physical nodes.
- [Modifying the Memory Allocated to a User Domain](#)
- [Modifying the Number of Virtual CPUs Allocated to a User Domain](#)
- [Increasing the Disk Space in a User Domain](#)  
You can increase the size of Logical Volume Manager (LVM) partitions, swap space, and file systems in a user domain.
- [Expanding /EXAVMIMAGES After Adding the Database Server Disk Expansion Kit](#)  
With the addition of a disk expansion kit to the database server, it is important to follow proper procedures to add this additional space to the /EXAVMIMAGES file system.
- [Adding an Oracle VM Cluster](#)  
You can use Oracle Exadata Deployment Assistant (OEDA) to create a new Oracle VM cluster on an existing Exadata Database Machine.
- [Expanding an Oracle RAC Cluster on Oracle VM Using OEDACLI](#)  
You can expand an existing Oracle RAC cluster on Oracle VM by adding user domains using the Oracle Exadata Deployment Assistant command-line interface (OEDACLI).
- [Creating a User Domain Without Oracle Grid Infrastructure and Oracle Database](#)
- [Moving a User Domain to a Different Database Server](#)  
User domains can move to different database servers.

- [Backing up the Management Domain and User Domains in an Oracle VM Deployment](#)
- [Recovering an Oracle VM Deployment](#)  
You can recover an Oracle VM from a snapshot-based backup when severe disaster conditions damage the Oracle VM, or when the server hardware is replaced to such an extent that it amounts to new hardware.
- [Removing an Oracle RAC Cluster Running in Oracle VM](#)  
You can remove all Oracle RAC nodes of an Oracle VM cluster, including the databases running within the cluster and all data stored on the Oracle Exadata Storage Server used by those databases.
- [Removing a User Domain](#)  
You can remove an user domain in Oracle VM using either OEDACLI or the `domu_maker` utility.
- [Implementing Tagged VLAN Interfaces](#)  
This topic describes the implementation of tagged VLAN interfaces in Oracle VM environments on Exadata.
- [Implementing InfiniBand Partitioning across Oracle VM Oracle RAC Clusters on Exadata Database Machine](#)  
For Oracle Real Application Clusters (Oracle RAC) clusters running in Oracle VM on Exadata Database Machine, you can isolate the network traffic on the InfiniBand network for each Oracle RAC clusters using custom InfiniBand partitioning, dedicated partition keys, and partitioned tables.
- [Running Oracle EXAchk in Oracle VM Environments](#)  
Oracle EXAchk version 12.1.0.2.2 and higher supports virtualization on Exadata Database Machine.

## 5.1 Oracle VM and Exadata Database Machine

When deploying Exadata Database Machine, you can decide to implement Oracle VM on the database servers.

Oracle VM Server and one or more Oracle VM guests are installed on every database server. You can configure Oracle VM environments on your initial deployment using scripts created by Oracle Exadata Deployment Assistant (OEDA) or you can migrate an existing environment to Oracle VM.



### Note:

Oracle VM is not supported on 8-socket servers, such as X7-8.

- [About Oracle VM](#)  
Oracle VM enables you to deploy the Oracle Linux operating system and application software within a supported virtualization environment.
- [Oracle VM Deployment Specifications and Limits](#)  
This topic describes the deployment specifications and limits for using Oracle VM on Oracle Exadata Database Machine.
- [Supported Operations in the Management Domain \(dom0\)](#)  
Manually modifying the dom0 can result in configuration issues for Oracle VM Server, which can degrade performance or cause a loss of service.



- [Oracle VM Resources](#)  
Two fundamental parts of the Oracle VM infrastructure – networking and storage – are configured outside of the Oracle VM.

## 5.1.1 About Oracle VM

Oracle VM enables you to deploy the Oracle Linux operating system and application software within a supported virtualization environment.

If you use Oracle VM on Oracle Exadata Database Machine, then they provide CPU, memory, operating system, and sysadmin isolation for your workloads. You can combine VMs with network and I/O prioritization to achieve full stack isolation. For consolidation, you can create multiple trusted databases or pluggable databases in an Oracle VM, allowing resources to be shared more dynamically.

An Oracle VM environment consists of an Oracle VM Server, virtual machines, and resources. An Oracle VM Server is a managed virtualization environment providing a lightweight, secure, server platform which runs virtual machines, also known as domains.

Oracle VM Server is installed on a bare metal computer. The hypervisor present on each Oracle VM Server is an extremely small-footprint virtual machine manager and scheduler. It is designed so that it is the only fully privileged entity in the system. It controls only the most basic resources of the system, including CPU and memory usage, privilege checks, and hardware interrupts.

The hypervisor securely executes multiple virtual machines on one host computer. Each virtual machine runs in its own domain and has its own guest operating system. A primary management domain, dom0, an abbreviation for *domain zero*, also runs as a guest on top of the hypervisor. Dom0 has privileged access to the hardware and device drivers.

A user domain (domU) is an unprivileged domain that can access the InfiniBand HCA. The domU is started and managed on an Oracle VM Server by dom0. Because a domU operates independently of other domains, a configuration change applied to the virtual resources of a domU does not affect any other domains. A failure of the domU does not impact any other domains.

The terms "domain", "guest", and "virtual machine" are often used interchangeably, but they have subtle differences:

- A domain is a configurable set of resources, including memory, virtual CPUs, network devices and disk devices, in which virtual machines run.
- A domain or virtual machine is granted virtual resources and can be started, stopped and restarted independently of other domains or the host server itself.
- A guest is a virtualized operating system running within a domain. Guest operating systems each have their own management domain called a user domain, abbreviated to domU.

Up to 8 guests can run on the same Oracle VM Server, each within its own domain. These domains are unprivileged domains that can access the InfiniBand HCA. Each domU is started alongside dom0 running on Oracle VM Server. Other domains never interact with dom0 directly. Their requirements are handled by the hypervisor itself. Dom0 only provides a means to administer the hypervisor.

You use Oracle Exadata Deployment Assistant (OEDA) to create and configure Oracle VMs on Oracle Exadata Database Machine.

## 5.1.2 Oracle VM Deployment Specifications and Limits

This topic describes the deployment specifications and limits for using Oracle VM on Oracle Exadata Database Machine.

**Table 5-1 Oracle VM Deployment Specifications and Limits**

Attribute	Value for X3-2	Value for X4-2	Value for X5-2	Value for X6-2	Value for X7-2	Value for X8-2
Maximum number of Oracle VM user domains on each database server	8	8	8	8	8	8
Total physical memory on each database server with Oracle VM	Default: 256 GB Maximum: 512 GB	Default: 256 GB Maximum: 512 GB	Default: 256 GB Maximum: 768 GB	Default: 256 GB Maximum: 768 GB	Default: 384 GB Maximum: 768 GB	Default: 384 GB Maximum: 768 GB
Total available memory on each database server for all Oracle VM user domains	Maximum: 464 GB	Maximum: 464 GB	Maximum: 720 GB	Maximum: 720 GB	Maximum: 720 GB	Maximum: 720 GB
Minimum memory limit for each Oracle VM user domain	16 GB	16 GB	16 GB	16 GB	16 GB	16 GB
Total CPU cores (vCPUs) on each database server	16 (32)	24 (48)	32 (64)	44 (88)	48 (96)	48 (96)

**Table 5-1 (Cont.) Oracle VM Deployment Specifications and Limits**

Attribute	Value for X3-2	Value for X4-2	Value for X5-2	Value for X6-2	Value for X7-2	Value for X8-2
CPU core (vCPU) limits for each Oracle VM user domain	Minimum: 2 (4) Maximum: 14 (28)	Minimum: 2 (4) Maximum: 22 (44)	Minimum: 2 (4) Maximum: 30 (60)	Minimum: 2 (4) Maximum: 42 (84)	Minimum: 2 (4) Maximum: 46 (92)	Minimum: 2 (4) Maximum: 46 (92)
<b>Note:</b> CPU over-provisioning is allowed but may cause performance conflicts.						
Total usable disk storage for Oracle VM user domains on each database server	700 GB	1.6 TB	1.6 TB (3.7 TB with Storage Expansion Kit)	1.6 TB (3.7 TB with Storage Expansion Kit)	1.6 TB (3.7 TB with Storage Expansion Kit)	3.2 TB



**Note:**

1 CPU core = 1 OCPU = 2 vCPUs = 2 hyper-threads

### 5.1.3 Supported Operations in the Management Domain (dom0)

Manually modifying the dom0 can result in configuration issues for Oracle VM Server, which can degrade performance or cause a loss of service.



**WARNING:**

Oracle does not support any changes that are made to the dom0 beyond what is documented. Oracle does not support running any third party software within the dom0.

If you are in doubt whether an operation on the dom0 is supported, contact Oracle Support Services.

## 5.1.4 Oracle VM Resources

Two fundamental parts of the Oracle VM infrastructure – networking and storage – are configured outside of the Oracle VM.

### Networking

When specifying the configuration details for your Oracle Exadata Rack using Oracle Exadata Deployment Assistant (OEDA), you provide input on how the required network IP addresses for Oracle VM environments should be created. The generated OEDA setup files are transferred to the Oracle Exadata Rack and used to create the network addresses.

### Storage

Oracle VM always requires a location to store environment resources that are essential to the creation and management of virtual machines. These resources include ISO files (virtual DVD images), VM configuration files and VM virtual disks. The location of such a group of resources is called a storage repository.

On Exadata Database Machine, storage for the Oracle VMs is configured as OCFS2 (Oracle Cluster File System) storage.

On X5-2 and later 2-socket Oracle Exadata Database Machine systems only, you can purchase a disk expansion kit to increase storage capacity. You can use the additional disk space to support more Oracle VM guests (up to a maximum of 8) by expanding `/EXAVMIMAGES` or to increase the size of the `/u01` partition in each `domU`.

### Maximum Supported VMs on Exadata

For any existing Exadata Database Server, the maximum number of supported VMs is eight. For software prerequisites, refer to My Oracle Support notes 888828.1 and 1270094.1.

- [Storage Configuration for Management Domain](#)  
The management domain (`dom0`) contains the image repository and Oracle VM configuration files.
- [Storage Configuration for User Domain](#)  
The user domain (`domU`) is a virtualized database node.

### Related Topics

- [Adding the Disk Expansion Kit to Database Servers: X8M-2 and Prior](#)
- [Expanding /EXAVMIMAGES After Adding the Database Server Disk Expansion Kit](#)  
With the addition of a disk expansion kit to the database server, it is important to follow proper procedures to add this additional space to the `/EXAVMIMAGES` file system.
- [Using Oracle Exadata Deployment Assistant](#)
- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Exadata Critical Issues \(My Oracle Support Doc ID 1270094.1\)](#)

### 5.1.4.1 Storage Configuration for Management Domain

The management domain (dom0) contains the image repository and Oracle VM configuration files.

The management domain contains the following directories:

- `/EXAVMIMAGES`, where the images used to create the guests are stored. The ZIP files in this directory contain the ISO files.
- `/conf`
- `/GuestImages`, where the files representing each user domain are stored.

The management domain exists on the physical disk `/dev/sda`. There are three disk partitions:

- `/dev/sda1` — Mounted as `/boot`.
- `/dev/sda2` — Used for `swap`.
- `/dev/sda3` — Used for the volume group `VGExaDb`.

The logical volumes created for the management domain are:

- `/dev/VGExaDb/LVDbSys2` — Used by `dbnodeupdate.sh` while performing a backup
- `/dev/VGExaDb/LVDbSys3` — Mounted as `/`
- `/dev/VGExaDb/LVDbSwap1` — Used for swap space
- `/dev/VGExaDb/LVDoNotRemoveOrUse` — Used by `dbnodeupdate.sh` while performing a backup
- `/dev/VGExaDb/LVDbExaVMImages` — Mounted as `/EXAVMIMAGES`

The `/EXAVMIMAGES` directory is where the configuration files for each virtual machine are located. The files are named using the format `/EXAVMIMAGES/GuestImages/nodename/vm.cfg`. Each virtual machine also has image files that point back to the ISO files in the image repository. The following files, except for `pv1_vgeexadb.img`, are created with reflinks, an OCFS2 feature:

- `/EXAVMIMAGES/GuestImages/user-domain-name/System.img`
- `/EXAVMIMAGES/GuestImages/user-domain-name/gridversion.img`
- `/EXAVMIMAGES/GuestImages/user-domain-name/dbversion.img`

### 5.1.4.2 Storage Configuration for User Domain

The user domain (domU) is a virtualized database node.

Each user domain has 4 virtual disks at the management domain (dom0) level. This can be seen from `/EXAVMIMAGES/GuestImages/user_domain_name/vm.cfg`. These 4 virtual disks are in turn soft linked to 4 files under `/EXAVMIMAGES/GuestImages/user_domain_name`, which are the real disk image files as described below:

- `/dev/xvda`, for the system image file `System.img`

- `/dev/xvdb`, for the Oracle Grid Infrastructure image file, for example, `grid12.1.0.2.2.img`. This virtual disk is 50 GB in size and is mounted as `/u01/app/version/grid`.
- `/dev/xvdc`, for the Oracle Database image file, for example, `db12.1.0.2.2-3.img`. This virtual disk is 50 GB in size and is mounted as `/u01/app/oracle/product/version/dbhome_1`.
- `/dev/xvdd`, for the `pv1_vgexadb.img` image file

The `System.img` (`/dev/xvda`) disk has 2 partitions created on pre-grub2 images and 3 partitions on grub2 images.

- Pre-Grub2 image
  - Partition 1 — The boot partition (`/boot`) for the user domain (512 MB), represented as `xvda1` in the user domain.
  - Partition 2 — Where the bios-grub is stored (24.5 GB), represented as `xvda2` in the user domain.
- Grub2 image
  - Partition 1 — The boot partition (`/boot`) for the user domain (512 MB), represented as `xvda1` in the user domain.
  - Partition 2 — The EFI boot partition on Oracle Exadata Database MachineX7 and later systems
  - Partition 3 — Where the bios-grub is stored (24.5 GB), represented as `xvda3` in the user domain.

The `pv1_vgexadb.img` (`/dev/xvdd`) disk has 1 partition. The disk partition `/dev/xvdd1` is 62 GB in size.

For pre-grub2 images, 2 physical volumes (PVs) are laid on top of the `xvda2` and `xvdd1` partitions. On grub2 images, 2 physical volumes (PVs) are laid on top of the `xvda3` and `xvdd1` partitions. A volume group (`VgExaDb`) of size 86.49G is laid on top of these physical volumes. This volume group contains the following logical volumes (LVs):

- `/dev/VGExaDb/LVDbSys1` (24 GB) — used for the root file system `/`. This LVM is confined to the `xvda2` partition (for pre-grub2 images) or the `xvda3` partition (for grub2 images).
- `/dev/VGExaDb/LVDbSys2` (24 GB) — used for `dbnodeupdate` backups.
- `/dev/VGExaDb/LVDbOra1` (24 GB) — used for the `/u01` file system which holds the `diagnostic_dest` area.
- `/dev/VGExaDb/LVDbSwap1` (16 GB) — used for `swap`
- `/dev/VGExaDb/LVDbDoNotRemoveOrUse` (1 GB) — a reserved LVM used by `dbnodeupdate`
- `/dev/VGExaDb/LVDbVnodenamedgname` (128 MB) — for the quorum disks.

All but the first and last LVs in the list above are contained in the `xvdd1` partition.

## 5.2 Migrating a Bare Metal Oracle RAC Cluster to an Oracle RAC Cluster in Oracle VM

 **Note:**

This topic applies only to two-socket x86 servers. It does not apply to eight-socket servers such as Oracle Exadata Database Machine X5-8.

The migration of a bare metal Oracle RAC cluster to an Oracle RAC cluster in Oracle VM can be achieved in the following ways:

- Migrate to Oracle RAC cluster in Oracle VM using the existing bare metal Oracle RAC cluster with zero downtime.
- Migrate to Oracle RAC cluster in Oracle VM by creating a new Oracle RAC cluster in Oracle VM with minimal downtime.
- Migrate to Oracle RAC cluster in Oracle VM using Oracle Data Guard with minimal downtime.
- Migrate to Oracle RAC cluster in Oracle VM using Oracle Recovery Manager (RMAN) backup and restore with complete downtime.

The conversion of a bare metal Oracle RAC cluster to an Oracle RAC cluster in Oracle VM has the following implications:

- Each of the database servers will be converted to an Oracle VM Server on which a management domain (dom0) is created along with one or more user domains, depending on the number of Oracle RAC clusters being deployed. Each user domain on a database server will belong to a particular Oracle RAC cluster.
- As part of the conversion procedure, the bare metal Oracle RAC cluster will be converted to one Oracle RAC cluster in Oracle VM to start with. There will be one user domain per database server.
- At the end of the conversion, the cell disk and grid disk configuration of the storage cells are the same as they were at the beginning of the conversion.
- The management domain will use a small portion of the system resources on each database server. Typically a management domain uses 8 GB of memory and 4 virtual CPUs. This has to be taken into consideration while sizing the SGA of the databases running on the Oracle RAC cluster in Oracle VM.
- Refer to My Oracle Support note 2099488.1 for the complete instructions.

### Related Topics

- [Migration of a Bare metal RAC cluster to an OVM RAC cluster on Exadata \(Doc ID 2099488.1\)](#)

## 5.3 Showing Running Domains

The following procedure describes how to show running domains:

1. Connect to the management domain (domain zero, or dom0).

2. Run the `xm list` command. The following is an example of the output:

```
Example
xm list
Name ID Mem VCPUs State Time(s)
Domain-0 0 8192 4 r----- 409812.7
dm01db01vm01 8 8192 2 -b----- 156610.6
dm01db01vm02 9 8192 2 -b----- 152169.8
dm01db01vm03 10 10240 4 -b----- 150225.9
dm01db01vm04 16 12288 8 -b----- 113519.3
dm01db01vm05 12 12288 8 -b----- 174101.6
dm01db01vm06 13 12288 8 -b----- 169115.9
dm01db01vm07 14 8192 4 -b----- 175573.0
```

## 5.4 Monitoring a User Domain Console

The following procedure describes how to monitor a user domain console:

1. Connect as the `root` user to the management domain.
2. Obtain the domain name using the `xm list` command.
3. Use the following command to attach to the user domain console:

```
xm console DomainName
```

In the preceding command, *DomainName* is the name of the domain.

4. Press `CTRL+]` to disconnect from the console.

## 5.5 Monitoring Oracle VMs with Oracle Enterprise Manager

The Exadata plug-in for Oracle Enterprise Manager discovers, manages, and monitors virtualized Oracle Exadata Database Machine in conjunction with the Virtualization Infrastructure plug-in of Oracle Enterprise Manager.

With virtualized Exadata, one Exadata Database Machine target is created for each physical Database Server instead of one DB Machine target for each DB cluster deployed through Oracle Exadata Deployment Assistant (OEDA). Compute nodes, Exadata Storage Servers, InfiniBand switches, compute node ILOM, PDU, KVM, and Cisco switch targets are discovered by the Exadata plug-in. The physical server (physical Oracle Server target), Dom0 (Virtual Platform target), and DomU (virtual Oracle Server target) are discovered and monitored by the Virtualization Infrastructure (VI) plug-in.

- Refer to Virtualized Exadata Database Machine in *Oracle Enterprise Manager Exadata Management Getting Started Guide* for instructions on how to discover Oracle VM domains on Oracle Exadata Database Machine.

## 5.6 Starting a User Domain

The following procedure describes how to start a user domain:

- Use the following command to start the user domain:

```
xm create /EXAVMIMAGES/GuestImages/DomainName/vm.cfg
Using config file "/EXAVMIMAGES/GuestImages/dm01db01vm04/vm.cfg".
Started domain dm01db01vm04 (id=23)
```



In the preceding command, *DomainName* is the name of the domain.

 **Note:**

To see Oracle Linux boot messages during user domain startup, connect to the console during startup using the `-c` option. To disconnect from the console after startup is complete, press `CTRL+]`.

## 5.7 Managing Automatic Startup of Oracle VM User Domains

By default, when you create a user domain, it is configured to automatically start when the management domain is started. You can enable and disable this feature as needed.

- [Enabling User Domain Automatic Start](#)
- [Disabling User Domain Automatic Start](#)

### 5.7.1 Enabling User Domain Automatic Start

The following procedure describes how to enable a user domain to start automatically when the management domain is started:

1. Connect to the management domain.
2. Locate the user domain configuration file under `/OVS/Repositories`.

Investigate the files under `/OVS/Repositories` to determine the file belonging to the domain that you are acting on. The user domain configuration file is located at `/OVS/Repositories/UUID/vm.cfg`, where *UUID* is the unique domain identifier.

3. Create a symbolic link to the user domain configuration file in the `/etc/xen/auto` directory using the following command:

```
ln -s /OVS/Repositories/UUID/vm.cfg /etc/xen/auto/DomainName.cfg
```

In the preceding command, *UUID* is the unique domain identifier and *DomainName* is the fully qualified name of the domain that you are acting on.

For example:

```
ln -s /OVS/Repositories/fa8f413f8f544dce8b065a814a0c491c/
vm.cfg /etc/xen/auto/dbm01adm05vm01.example.com.cfg
```

### 5.7.2 Disabling User Domain Automatic Start

The following procedure describes how to disable a user domain from automatically starting when the management domain is started:

1. Connect to the management domain.

2. Remove the symbolic link to the user domain configuration file in the `/etc/xen/auto` directory using the following command:

```
rm /etc/xen/auto/DomainName.cfg
```

In the preceding command, *DomainName* is the fully qualified name of the domain that you are acting on.

For example:

```
rm /etc/xen/auto/dbm01adm05vm01.example.com.cfg
```

## 5.8 Shutting Down a User Domain From Within the User Domain

The following procedure describes how to shut down a user domain from within a user domain:

1. Connect as the `root` user to the user domain.
2. Use the following command to shut down the domain:

```
shutdown -h now
```

## 5.9 Shutting Down a User Domain From Within the Management Domain

The following procedure describes how to shut down a user domain from within a management domain:

1. Connect as the `root` user to the management domain.
2. Use the following command to shut down the domain:

```
xm shutdown DomainName -w
```

In the preceding command, *DomainName* is the name of the domain.

### Note:

Use the `-w` option so that the `xm` command waits until the domain shutdown completes before returning. The `xm shutdown` command performs the same orderly shutdown as running `shutdown -h now` within the user domain.

To shut down all user domains within the management domain, use the following command:

```
xm shutdown -a -w
```

The following is an example of the output:

```
Domain dm01db01vm04 terminated
All domains terminated
```

## 5.10 Backing Up and Restoring Oracle Databases on Oracle VM User Domains

Backing up and restoring Oracle databases on Oracle VM user domains is the same as backing up and restoring Oracle databases on physical nodes.

- When backing up Oracle databases on Oracle VM user domains you must set the following four parameters in the `/etc/sysctl.conf` file on the database nodes (user domains). If you are using Oracle Exadata storage to hold the backups, the parameters need to be set in the `/etc/sysctl.conf` file on the Exadata storage cells as well.

```
net.core.rmem_default = 4194304
net.core.wmem_default = 4194304
net.core.rmem_max = 4194304
net.core.wmem_max = 4194304
```

- If you are using Exadata storage, each Oracle VM RAC cluster requires its own Oracle Automatic Storage Management (Oracle ASM) disk group to be designated as the fast recovery area (FRA) such as `+RECO`. Refer to the "[Exadata Database Machine Backup and Restore Configuration and Operational Best Practices](#)" technical reference paper for details.
- If you are using Oracle ZFS Storage Appliance, refer to the "[Protecting Exadata Database Machine with the Oracle ZFS Storage Appliance: Configuration Best Practices](#)" technical reference paper for details.

## 5.11 Modifying the Memory Allocated to a User Domain

The following procedure describes how to modify the memory allocated to a user domain:

### Note:

This operation requires user domain restart. It is not supported to modify memory allocation using the `xm mem-set` command.

1. If you are decreasing the amount of memory used by the user domain, then you must first review and adjust Oracle Database memory usage and the operating system huge pages configuration settings inside the user domain.
  - a. Review the SGA size of databases and reduce if necessary.
 

If you do not first reduce the memory requirements of the databases running in the user domain, then the user domain might fail to restart because too much memory is reserved for huge pages when the Oracle Linux operating system attempts to boot. See My Oracle Support Doc ID 361468.1 for details.
  - b. Review the operating system configuration and reduce the memory allocation for huge pages if necessary.

- c. If you modify the huge pages settings in the operating system kernel configuration file (`/etc/sysctl.conf`), regenerate the `initramfs` file to reflect the system configuration change.

You should backup up the existing `initramfs` file and then regenerate it by using the `dracut` command. For example:

```
ls -l /boot/initramfs-$(uname -r).img
-rw----- 1 root root 55845440 Jan 8 10:34 /boot/
initramfs-4.14.35-2047.508.3.3.el7uek.x86_64.img

cp /boot/initramfs-$(uname -r).img backup_directory

dracut --force
```

2. Connect to the management domain (`dom0`).

The remainder of this procedure is preformed inside the management domain.

3. Use the following command to determine the amount of free memory available, when increasing the allocation:

```
xm info | grep free_memory
```

#### Note:

When assigning free memory to a user domain, approximately 1 to 2 percent of free memory is used for metadata and control structures. Therefore, the amount of memory increase possible is 1 to 2 percent less than free memory value.

4. Shut down the user domain gracefully using the name obtained from the `xm list` command. Use the `-w` option so the `xm` command waits until the domain is shut down before returning.

```
xm shutdown DomainName -w
```

In the preceding command, *DomainName* is the name of the domain.

5. Create a backup copy of the `/EXAVMIMAGES/GuestImages/DomainName/vm.cfg` file.
6. Edit the `memory` and `maxmem` settings in the `/EXAVMIMAGES/GuestImages/DomainName/vm.cfg` file using a text editor. The `memory` and `maxmem` settings must be identical values.

#### Note:

If the `memory` and `maxmem` parameters are not identical values, then InfiniBand network interfaces are not configured during user domain start, which prevents proper Oracle CRS and database startup.

- Use the following command to start the user domain:

```
xm create /EXAVMIMAGES/GuestImages/DomainName/vm.cfg
```

 **Note:**

To see Oracle Linux boot messages during user domain startup, connect to the console during startup using the `-c` option. To disconnect from the console after startup is complete, press `CTRL+]`.

## 5.12 Modifying the Number of Virtual CPUs Allocated to a User Domain

Note the following about modifying the number of virtual CPUs (vCPUs):

- All actions to modify the number of vCPUs allocated to a user domain are performed in the management domain.
- The number of vCPUs allowed for a user domain may be changed dynamically to a lower value or to a higher value provided it does not exceed the setting of `maxvcpus` parameter for the user domain.
- It is possible to over-commit vCPUs such that the total number of vCPUs assigned to all domains exceeds the number of physical CPUs on the system. However, over-committing CPUs should be done only when competing workloads for oversubscribed resources are well understood and concurrent demand does not exceed physical capacity.

The following procedure describes how to modify the number of virtual CPUs allocated to a user domain:

- Determine the number of physical CPUs as follows:
  - Use the following command in the management domain:

```
xm info | grep -A3 nr_cpus
nr_cpus : 24
nr_nodes : 2
cores_per_socket : 6
threads_per_core : 2
```

In the output, note that the `nr_nodes` line refers to the number of sockets. The Exadata database server where the command is run is a 2-socket 6 cores per socket processor, resulting in 24 physical CPU threads (2 sockets x 6 cores/socket = 12 cores. 12 cores x 2 threads per core = 24 CPU threads).

- Use the following command to determine the current setting of vCPUs configured and online for a user domain:

```
xm list DomainName -l | grep vcpus

(vcpus 4)
(online_vcpus 2)
```

In the preceding command, *DomainName* is the name of the user domain. The output from the command indicates the maximum number of vCPUs for the user domain is 4, and the current number of online vCPUs is 2. This user domain may have the number of online vCPUs adjusted to any value not greater than the `vcpus` parameter while the user domain remains online. The user domain must be taken offline to increase the number of online vCPUs to a value higher than the `vcpus` parameter.

2. Reduce or increase the number of vCPUs as follows:

- To reduce the number of vCPUs:
  - a. Determine the currently allocated number of vCPUs for the user domain using the following command:

```
xm list DomainName
```

- b. Reduce the currently allocated number of vCPUs using the following command:

```
xm vcpu-set DomainName vCPUs_preferred
```

In the preceding command, *vCPUs\_preferred* is the value of the preferred number of vCPUs

- To increase the number of vCPUs
  - a. Determine the current settings of the `vcpus` parameter using the following command:

```
xm list DomainName -l | grep vcpus
(vcpus 4)
(online_vcpus 2)
```

- b. If the preferred number of vCPUs is less than or equal to the value of the `vcpus` parameter, then run the following command to increase the number of online vCPUs.

```
xm vcpu-set DomainName vCPUs_preferred
```

In the preceding command, *vCPUs\_preferred* is the value of the preferred number of vCPUs

- c. If the preferred number of vCPUs is greater than the value of the `vcpus` parameter, then the user domain must be taken offline to increase the number of online vCPUs to a value higher than the `vcpus` parameter. Do the following:

- i. Shut down the user domain.

- ii. Create a backup copy of the `/EXAVMIMAGES/GuestImages/DomainName/vm.cfg` file.

- iii. Edit the `/EXAVMIMAGES/GuestImages/DomainName/vm.cfg` file to set the `vcpus` parameter to the desired number of vCPUs.

**Note:** By default a user domain will online the number of vCPUs configured via the `vcpus` parameter. If you want a user domain to start with some vCPUs offline, then add the `maxvcpus` parameter to `vm.cfg`, setting it to the maximum number of vCPUs the user domain is permitted to have online. Set the `vcpus` parameter to the number of vCPUs to online when the user domain starts. For example, to start a user domain with 2 vCPUs

online and to allow an additional 6 vCPUs to be added to the user domain while it remains online, use the following settings in `vm.cfg`:

```
maxvcpus=8
vcpus=2
```

iv. Start the user domain.

## 5.13 Increasing the Disk Space in a User Domain

You can increase the size of Logical Volume Manager (LVM) partitions, swap space, and file systems in a user domain.

- [Adding a New LVM Disk to a User Domain](#)
- [Increasing the Size of the root File System](#)  
This procedure describes how to increase the size of the system partition and / (root) file system.
- [Increasing the Size of the /u01 File System](#)  
This procedure describes how to increase the size of the /u01 file system.
- [Increasing the Size of the Grid Infrastructure Home or Database Home File System](#)  
You can increase the size of the Oracle Grid Infrastructure or Oracle Database home file system in a user domain.
- [Increasing the Size of the Swap Area](#)  
This procedure describes how to increase the amount of swap configured in a user domain.

### 5.13.1 Adding a New LVM Disk to a User Domain

This procedure describes how to add a new LVM disk to a user domain to increase the amount of usable LVM disk space in a user domain. This procedure is done so that the size of a file system or swap LVM partition can be increased. This procedure is performed while the system remains online.

#### Note:

This procedure requires steps be run in the management domain (Domain-0), and in the user domain.

Run all steps in this procedure as the `root` user.

1. In the management domain, verify the free disk space in `/EXAVMIMAGES` using the following command:

```
df -h /EXAVMIMAGES
```

The following is an example of the output from the command:

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/sda3	721G	111G	611G	16%	/EXAVMIMAGES

2. In the management domain, select a name for the new disk image, and verify that the name is not already used in the user domain.

```
ls -l /EXAVMIMAGES/GuestImages/DomainName/new_disk_image_name

ls: /EXAVMIMAGES/GuestImages/DomainName/new_disk_image_name: No such file
or \
directory
```

In the preceding command, *DomainName* is the name of the domain, and *new\_disk\_image\_name* is the new disk image name.

3. In the management domain, create a new disk image.

```
qemu-img create /EXAVMIMAGES/GuestImages/DomainName/new_disk_image_name
size
```

In the following example of the command, the new disk image name is *pv2\_vgexadb.img*, and the image size is 10 GB.

```
qemu-img create /EXAVMIMAGES/GuestImages/DomainName/pv2_vgexadb.img 10G
```

4. In the user domain, determine an available disk name. In the following example, disk names *xvda* through *xvdd* are used, and disk name *xvde* is unused.

```
lsblk -id
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
xvda 202:0 0 13G 0 disk
xvdb 202:16 0 20G 0 disk /u01/app/12.1.0.2/grid
xvdc 202:32 0 20G 0 disk /u01/app/oracle/product/12.1.0.2/dbhome_1
xvdd 202:48 0 41G 0 disk
```

5. In the management domain, attach the new disk image to the user domain in read/write mode. In the following example, the new disk image is presented in the user domain as device */dev/xvde*.

```
xm block-attach DomainName \
file:/EXAVMIMAGES/GuestImages/DomainName/new_disk_image_name /dev/xvde w
```

6. In the user domain, verify the disk device is available. In the following example, disk name *xvde* is available in the user domain.

```
lsblk -id
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
xvda 202:0 0 13G 0 disk
xvdb 202:16 0 20G 0 disk /u01/app/12.1.0.2/grid
xvdc 202:32 0 20G 0 disk /u01/app/oracle/product/12.1.0.2/dbhome_1
xvdd 202:48 0 41G 0 disk
xvde 202:64 0 10G 0 disk
```

7. In the user domain, partition the new disk device. In the following example, disk device */dev/xvde* is partitioned.

```
parted /dev/xvde mklabel gpt
parted -s /dev/xvde mkpart primary 0 100%
parted -s /dev/xvde set 1 lvm on
```

The `parted mkpart` command may report the following message. This message can be ignored:

```
Warning: The resulting partition is not properly aligned for best
performance.
```

8. In the user domain, create an LVM physical volume on the new disk partition. In the following example, an LVM physical volume is created on disk partition */dev/xvde1*.



```
pvcreate /dev/xvde1
```

9. In the user domain, extend the volume group and verify the additional space in the volume group. In the following example, disk name `xvde` is now available in the user domain.

```
vgextend VGExaDb /dev/xvde1
vdisplay -s
```

10. In the management domain, make a backup of the user domain configuration file `vm.cfg`.

```
cp /EXAVMIMAGES/GuestImages/DomainName/vm.cfg \
 /EXAVMIMAGES/GuestImages/DomainName/vm.cfg.backup
```

11. In the management domain, obtain the UUID of the user domain using the following command:

```
grep ^uuid /EXAVMIMAGES/GuestImages/DomainName/vm.cfg
```

In the following example, the user domain UUID is `49ffddce4efe43f5910d0c61c87bba58`.

```
grep ^uuid /EXAVMIMAGES/GuestImages/dm01db01vm01/vm.cfg
uuid = '49ffddce4efe43f5910d0c61c87bba58'
```

12. In the management domain, generate a UUID for the new disk image using the following command:

```
uuidgen | tr -d '-'
```

In the following example, the new disk UUID is `0d56da6a5013428c97e73266f81c3404`.

```
uuidgen | tr -d '-'
0d56da6a5013428c97e73266f81c3404
```

13. In the management domain, create a symbolic link from `/OVS/Repositories` to the new disk image using the following command:

```
ln -s /EXAVMIMAGES/GuestImages/DomainName/newDiskImage.img \
 /OVS/Repositories/user_domain_uuid/VirtualDisks/new_disk_uuid.img
```

In the following example, a symbolic link is created to the new disk image file `pv2_vgexadb.img` for user domain `dm01db01vm01`. The UUID for user domain `dm01db01vm01` is `49ffddce4efe43f5910d0c61c87bba58`. The UUID for the new disk image is `0d56da6a5013428c97e73266f81c3404`.

```
ln -s /EXAVMIMAGES/GuestImages/dm01db01vm01/pv2_vgexadb.img \
 /OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/VirtualDisks/ \
 0d56da6a5013428c97e73266f81c3404.img
```

14. In the management domain, append an entry for the new disk to the disk parameter in the user domain configuration file `vm.cfg`. This makes the new disk image attach automatically to the user domain during the next startup. The new entry matches the following format:

```
'file:/OVS/Repositories/user_domain_uuid/VirtualDisks/
new_disk_uuid.img,disk_device,w'
```

The following is an example of an original disk parameter entry in the `vm.cfg` file:

```
disk=['file:/OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/VirtualDisks/ \
76197586bc914d3d9fa9d4f092c95be2.img,xvda,w', \
'file:/OVS/Repositories/49ffddce4efe43f591 0d0c61c87bba58/VirtualDisks/ \
78470933af6b4253b9ce27814ceddbbd.img,xvdb,w', \
'file:/OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/VirtualDisks/ \
```

```
20d5528f5f9e4fd8a96f151a13d2006b.img,xvdc,w',
\
'file:/OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/
VirtualDisks/ \
058af368db2c4f27971bbe1f19286681.img,xvdd,w']
```

The following example shows an entry appended to the disk parameter for a new disk image that is accessible within the user domain as disk device `/dev/xvde`:

```
disk=['file:/OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/
VirtualDisks/ \
76197586bc914d3d9fa9d4f092c95be2.img,xvda,w',
\
'file:/OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/
VirtualDisks/ \
78470933af6b4253b9ce27814ceddbbd.img,xvdb,w',
\
'file:/OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/
VirtualDisks/ \
20d5528f5f9e4fd8a96f151a13d2006b.img,xvdc,w',
\
'file:/OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/
VirtualDisks/ \
058af368db2c4f27971bbe1f19286681.img,xvdd,w',
\
'file:/OVS/Repositories/49ffddce4efe43f5910d0c61c87bba58/
VirtualDisks/ \
0d56da6a5013428c97e73266f81c3404.img,xvde,w']
```

## 5.13.2 Increasing the Size of the root File System

This procedure describes how to increase the size of the system partition and `/` (root) file system.

This procedure is performed while the file system remains online.

### Note:

There are two system partitions, `LVDbSys1` and `LVDbSys2`. One partition is active and mounted. The other partition is inactive and used as a backup location during upgrade. The size of both system partitions must be equal.

Keep at least 1 GB of free space in the `VGExaDb` volume group. The free space is used for the LVM snapshot created by the `dbnodeupdate.sh` utility during software maintenance. If you make snapshot-based backups of the `/` (root) and `/u01` directories as described in "[Creating a Snapshot-Based Backup of Oracle Linux Database Server](#)," then keep at least 6 GB of free space in the `VGExaDb` volume group.

1. Collect information about the current environment.
  - a. Use the `df` command to identify the amount of free and used space in the root partition (`/`)

```
df -h /
```

The following is an example of the output from the command:

```
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbSys1
 12G 5.1G 6.2G 46% /
```

 **Note:**

The active root partition may be either `LVDbSys1` or `LVDbSys2`, depending on previous maintenance activities.

- b. Use the `lvs` command to display the current logical volume configuration.

```
lvs -o lv_name,lv_path,vg_name,lv_size
```

The following is an example of the output from the command:

```
LV Path VG LSize
LVDbOra1 /dev/VGExaDb/LVDbOra1 VGExaDb 10.00g
LVDbSwap1 /dev/VGExaDb/LVDbSwap1 VGExaDb 8.00g
LVDbSys1 /dev/VGExaDb/LVDbSys1 VGExaDb 12.00g
LVDbSys2 /dev/VGExaDb/LVDbSys2 VGExaDb 12.00g
```

2. Verify there is available space in the volume group `VGExaDb` using the `vgdisplay` command.

```
vgdisplay VGExaDb -s
```

The following is an example of the output from the command:

```
"VGExaDb" 53.49 GiB [42.00 GiB used / 11.49 GiB free]
```

The volume group must contain enough free space to increase the size of both system partitions, and maintain at least 1 GB of free space for the LVM snapshot created by the `dbnodeupdate.sh` utility during upgrade. If there is not sufficient free space in the volume group, then add a new disk to LVM.

3. Resize both `LVDbSys1` and `LVDbSys2` logical volumes using the `lvextend` command.

```
lvextend -L +size /dev/VGExaDb/LVDbSys1
lvextend -L +size /dev/VGExaDb/LVDbSys2
```

In the preceding command, *size* is the amount of space to be added to the logical volume. The amount of space added to each system partition must be the same.

The following example extends the logical volumes by 10 GB:

```
lvextend -L +10G /dev/VGExaDb/LVDbSys1
lvextend -L +10G /dev/VGExaDb/LVDbSys2
```

4. Resize the file system within the logical volume using the `resize2fs` command.

```
resize2fs /dev/VGExaDb/LVDbSys1
resize2fs /dev/VGExaDb/LVDbSys2
```

5. Verify the space was extended for the active system partition using the `df` command.

```
df -h /
```

### Related Topics

- [Adding a New LVM Disk to a User Domain](#)

## 5.13.3 Increasing the Size of the /u01 File System

This procedure describes how to increase the size of the /u01 file system.

This procedure is performed while the file system remains online.

### Note:

Keep at least 1 GB of free space in the `VGExaDb` volume group. The free space is used for the LVM snapshot created by the `dbnodeupdate.sh` utility during software maintenance. If you make snapshot-based backups of the / (root) and /u01 directories as described in "[Creating a Snapshot-Based Backup of Oracle Linux Database Server](#)," then keep at least 6 GB of free space in the `VGExaDb` volume group

1. Collect information about the current environment.
  - a. Use the `df` command to identify the amount of free and used space in the /u01 partition.

```
df -h /u01
```

The following is an example of the output from the command:

```
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbOral
 9.9G 1.7G 7.8G 18% /u01
```

- b. Use the `lvs` command to display the current logical volume configuration used by the /u01 file system.

```
lvs -o lv_name,lv_path,vg_name,lv_size /dev/VGExaDb/LVDbOral
```

The following is an example of the output from the command:

```
LV Path VG LSize
LVDbOra1 /dev/VGExaDb/LVDbOra1 VGExaDb 10.00g
```

2. Verify there is available space in the volume group `VGExaDb` using the `vgdisplay` command.

```
vgdisplay VGExaDb -s
```

The following is an example of the output from the command:

```
"VGExaDb" 53.49 GiB [42.00 GiB used / 11.49 GiB free]
```

If the output shows there is less than 1 GB of free space, then neither the logical volume nor file system should be extended. Maintain at least 1 GB of free space in the `VGExaDb` volume group for the LVM snapshot created by the `dbnodeupdate.sh` utility during an upgrade. If there is not sufficient free space in the volume group, then add a new disk to LVM.

3. Resize the logical volume using the `lvextend` command.

```
lvextend -L +sizeG /dev/VGExaDb/LVDbOra1
```

In the preceding command, *size* is the amount of space to be added to the logical volume.

The following example extends the logical volume by 10 GB:

```
lvextend -L +10G /dev/VGExaDb/LVDbOra1
```

4. Resize the file system within the logical volume using the `resize2fs` command.

```
resize2fs /dev/VGExaDb/LVDbOra1
```

5. Verify the space was extended using the `df` command.

```
df -h /u01
```

### Related Topics

- [Adding a New LVM Disk to a User Domain](#)

## 5.13.4 Increasing the Size of the Grid Infrastructure Home or Database Home File System

You can increase the size of the Oracle Grid Infrastructure or Oracle Database home file system in a user domain.

The Oracle Grid Infrastructure software home and the Oracle Database software home are created as separate disk image files in the management domain. The disk image files are located in the `/EXAVMIMAGES/GuestImages/DomainName/` directory. The disk image files

are attached to the user domain automatically during virtual machine startup, and mounted as separate, non-LVM file systems in the user domain.

1. Connect to the user domain, and check the file system size using the `df` command, where `$ORACLE_HOME` is an environment variable that points to the Oracle Database home directory, for example, `/u01/app/oracle/product/12.1.0.2/dbhome_1`.

```
df -h $ORACLE_HOME
```

The following is an example of the output from the command:

```
Filesystem Size Used Avail Use% Mounted on
/dev/xvdc 20G 6.5G 13G 35% /u01/app/oracle/product/12.1.0.2/
dbhome_1
```

2. Connect to the management domain, and then shut down the user domain using the `xm` command and specifying the name of the domain.

```
xm shutdown DomainName
```

3. Create an OCFS reflink to serve as a backup of the disk image that will be increased, where *version* is the release number for example, 12.1.0.2.1-3.

```
cd /EXAVMIMAGES/GuestImages/DomainName
reflink dbversion.img before_resize.dbversion.img
```

4. Create an empty disk image using the `qemu-img` command, and append it to the database home disk image.

The empty disk image size is the size to extend the file system. The last command removes the empty disk image after appending to the database home disk image.

```
qemu-img create emptyfile 10G
cat emptyfile >> dbversion.img
rm emptyfile
```

5. Check the file system using the `e2fsck` command.

```
e2fsck -f dbversion.img
```

6. Resize the file system using the `resize2fs` command.

```
resize2fs dbversion.img
```

7. Start the user domain.

```
xm create /EXAVMIMAGES/GuestImages/DomainName/vm.cfg
```

8. Connect to the user domain, and verify the file system size was increased.

```
df -h $ORACLE_HOME
```

The following is an example of the output from the command:

```
Filesystem Size Used Avail Use% Mounted on
/dev/xvdc 30G 6.5G 22G 23% /u01/app/oracle/product/12.1.0.2/
dbhome_1
```

9. Connect to the management domain, and remove the backup image.

Use a command similar to the following where *back\_up\_image.img* is the name of the backup image file:

```
cd /EXAVMIMAGES/GuestImages/DomainName
rm back_up_image.img
```

## 5.13.5 Increasing the Size of the Swap Area

This procedure describes how to increase the amount of swap configured in a user domain.

1. Verify there is available space in the volume group *VGExaDb* using the `vgdisplay` command.

```
vgdisplay VGExaDb -s
```

The following is an example of the output from the command:

```
"VGExaDb" 53.49 GiB [42.00 GiB used / 11.49 GiB free]
```

If the command shows that there is less than 1 GB of free space, then neither the logical volume nor file system should be extended. Maintain at least 1 GB of free space in the *VGExaDb* volume group for the LVM snapshot created by the `dbnodeupdate.sh` utility during an upgrade. If there is not sufficient free space in the volume group, then add a new disk to LVM.

2. Create a new logical volume of the size to increase swap space using the `lvcreate` command.

In the following example, a new 8 GB logical volume named *LVDdbSwap2* is created.

```
lvcreate -L 8G -n LVDdbSwap2 VGExaDb
```

3. Setup the new logical volume as a swap device with a unique label, such as *SWAP2*, using the `mkswap` command. The unique label is a device LABEL entry that is currently unused in the `/etc/fstab` file.

```
mkswap -L SWAP2 /dev/VGExaDb/LVDdbSwap2
```

4. Enable the new swap device using the `swapon` command.

```
swapon -L SWAP2
```

5. Verify the new swap device is enabled using the `swapon` command.

```
swapon -s
```

The following is an example of the output from the command:

Filename	Type	Size	Used	Priority
/dev/dm-3	partition	8388604	306108	-1
/dev/dm-4	partition	8388604	0	-2

6. Edit the `/etc/fstab` file to add the new swap device by copying the existing swap entry, and then changing the LABEL value in the new entry to the label used to create the new swap device. In the following example, the new swap device was added to the `/etc/fstab` file as `LABEL=SWAP2`.

```
cat /etc/fstab
LABEL=DBSYS / ext4 defaults 1 1
LABEL=BOOT /boot ext4 defaults,nodev
1 1
tmpfs /dev/shm tmpfs defaults,size=7998m 0
devpts /dev/pts devpts gid=5,mode=620 0 0
sysfs /sys sysfs defaults 0 0
proc /proc proc defaults 0 0
LABEL=SWAP swap swap defaults 0 0
LABEL=SWAP2 swap swap defaults 0 0
LABEL=DBORA /u01 ext4 defaults 1 1
/dev/xvdb /u01/app/12.1.0.2/grid ext4 defaults 1 1
/dev/xvdc /u01/app/oracle/product/12.1.0.2/dbhome_1
ext4 defaults 1 1
```

### Related Topics

- [Adding a New LVM Disk to a User Domain](#)

## 5.14 Expanding /EXAVMIMAGES After Adding the Database Server Disk Expansion Kit

With the addition of a disk expansion kit to the database server, it is important to follow proper procedures to add this additional space to the `/EXAVMIMAGES` file system.



### Note:

The disk expansion kit is supported only on 2-socket Oracle Exadata Database Machine systems, X5-2 and later.

- [Expanding /EXAVMIMAGES on Management Domain on Release 18.1.x or Later](#)  
If you are using a release of Oracle Exadata System Software release 18c (18.1.0) or later, then use this procedure to expand the `/EXAVMIMAGES` file system on the management domain following the addition of a disk expansion kit.



- [Expanding /EXAVMIMAGES on Management Domain on Release 12.2.x](#)  
If you are using Oracle Exadata System Software release 12.2.x, then use this procedure to expand the /EXAVMIMAGES file system on the management domain following the addition of a disk expansion kit.
- [Expanding /EXAVMIMAGES on Management Domain on Releases Earlier than 12.2.x](#)  
If you are using a release of Oracle Exadata System Software starting with release 12.1.2.3.0 but earlier than release 12.2.x, then use this procedure to expand the /EXAVMIMAGES directory on the management domain following the addition of a disk expansion kit.

## 5.14.1 Expanding /EXAVMIMAGES on Management Domain on Release 18.1.x or Later

If you are using a release of Oracle Exadata System Software release 18c (18.1.0) or later, then use this procedure to expand the /EXAVMIMAGES file system on the management domain following the addition of a disk expansion kit.

During deployment, all available disk space on a database server will be allocated in the management domain (dom0) with the majority of the space allocated to /EXAVMIMAGES for user domain storage. The /EXAVMIMAGES file system is created on /dev/VGExaDb/LVDbExaVMImages.

In the example below, dm01db01 is the name of the management domain, and dm01db01vm01 is a user domain.

1. Ensure `reclaimdisks.sh` has been run in the management domain by using the `-check` option.

Note that the last line reads "Layout: DOM0". If `reclaimdisks.sh` was not run, it would read "Layout: DOM0 + Linux".

```
[root@dm01db01 ~]# /opt/oracle.SupportTools/reclaimdisks.sh -check
Model is ORACLE SERVER X6-2
Number of LSI controllers: 1
Physical disks found: 4 (252:0 252:1 252:2 252:3)
Logical drives found: 1
Linux logical drive: 0
RAID Level for the Linux logical drive: 5
Physical disks in the Linux logical drive: 4 (252:0 252:1 252:2 252:3)
Dedicated Hot Spares for the Linux logical drive: 0
Global Hot Spares: 0
Valid. Disks configuration: RAID5 from 4 disks with no global and
dedicated hot spare disks.
Valid. Booted: DOM0. Layout: DOM0.
```

2. Add the disk expansion kit to the database server.

The kit consists of 4 additional hard drives to be installed in the 4 available slots. Remove the filler panels and install the drives. The drives may be installed in any order.

3. Verify that the RAID reconstruction is completed by seeing the warning and clear messages in the alert history.

This may take several hours to complete. The example below shows that it took approximately 7 hours. Once the clear message (message 1\_2 below) is present, the reconstruction is completed and it is safe to proceed.

```
[root@dm01db01 ~]# dbmcli -e list alerthistory

 1_1 2016-02-15T14:01:00-08:00 warning "A
disk
expansion kit was installed. The additional physical drives were
automatically
added to the existing RAID5 configuration, and reconstruction of
the
corresponding virtual drive was automatically started."

 1_2 2016-02-15T21:01:01-08:00 clear
"Virtual drive
reconstruction due to disk expansion was completed."
```

#### 4. Collect information about the current environment.

```
[root@dm01db01 ~]# df -h /EXAVMIMAGES
Filesystem Size Used Avail Use% Mounted on
/dev/sda3 1.6T 44G 1.5T 3% /EXAVMIMAGES
```

```
[root@dm01db01 ~]# xm list
Name ID Mem VCPUs
State Time(s)
Domain-0 0 8192 4
r----- 94039.1
dm01db01vm01.example.com 4 16384 2 -
b----- 3597.3
```

#### 5. Stop all user domains by running the command `xm shutdown -a -w` from the management domain.

After all user domains are shut down, only Domain-0 (the management domain) should be listed.

```
[root@dm01db01 ~]# xm shutdown -a -w
Domain dm01db01vm01.example.com terminated
All domains terminated
```

```
[root@dm01db01 ~]# xm list
Name ID Mem VCPUs
State Time(s)
Domain-0 0 8192 4
r----- 94073.4
```

#### 6. Run `parted` to view the sector start and end values.

Check the size of the disk against the end of the third partition. If you see a request to fix the GPT, respond with `n`.

```
root@dm01db01 ~]# parted /dev/sda
GNU Parted 2.1Using /dev/sda
```

```
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) unit s
(parted) print
Warning: Not all of the space available to /dev/sda appears to be used,
you can
fix the GPT to use all of the space (an extra 4679680000 blocks) or
continue
with the current setting? Fix/Ignore? F

Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt

Number Start End Size File system Name
Flags
1 64s 1046591s 1046528s ext3 primary boot
4 1046592s 1048639s 2048s primary
bios_grub
2 1048640s 240132159s 239083520s primary lvm

(parted) q
```

The partition table shown above lists partition 2 as ending at sector 240132159 and disk size as 8189440000 sectors. You will use these values in step 7.

**7. Create a fourth partition.**

The start sector is the end of the third partition from step 6 plus 1 sector (240132159+1=240132160). The end sector of the fourth partition is the size of the disk minus 34 (8189440000-34=8189439966).

```
[root@dm01db01 ~]# parted -s /dev/sda mkpart primary 240132160s
8189439966s
```

This command produces no output.

**8. Set the LVM flag for the fourth partition.**

```
[root@dm01db01 ~]# parted -s /dev/sda set 3 lvm on
Warning: WARNING: the kernel failed to re-read the partition table
on /dev/sda (Device or
resource busy). As a result, it may not reflect all of your changes
until after reboot.
```

**9. Review the updated partition table.**

```
[root@dm01db01 ~]# parted -s /dev/sda unit s print
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt

Number Start End Size File system Name Flags
1 64s 1046591s 1046528s ext4 primary boot
4 1046592s 1048639s 2048s primary
```

```

bios_grub
2 1048640s 240132159s 239083520s
primary lvm
3 240132160s 8189439966s 7949307807s
primary lvm

```

**10. Restart the Exadata server.**

```
[root@dm01db01 ~]# shutdown -r now
```

**11. Check the size of the disk against the end of the fourth partition.**

```

[root@dm01db01 ~]# parted -s /dev/sda unit s print
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Number Start End Size File system
Name Flags
1 64s 1048639s 1048576s ext4
primary boot
4 1048640s 3509759966s 3508711327s
primary lvm
2 3509759967s 8189439966s 4679680000s
primary lvm
3

```

**12. Create a LVM physical volume (PV) on the newly created fourth partition.**

```

[root@dm01db01 ~]# lvm pvcreate --force /dev/sda3
Physical volume "/dev/sda3" successfully created

```

**13. Extend the LVM volume group VGExaDb to the newly created third partition.**

```

[root@dm01db01 ~]# lvm vgextend VGExaDb /dev/sda3
Volume group "VGExaDb" successfully extended

```

**14. Dismount the /EXAVMIMAGES OCFS2 partition.**

```
[root@dm01db01 ~]# umount /EXAVMIMAGES/
```

**15. Extend the logical volume that contains the OCFS2 partition to include the rest of the free space.**

```

[root@dm01db01 ~]# lvm lvextend -l +100%FREE /dev/VGExaDb/
LVDbExaVMImages
Size of logical volume VGExaDb/LVDbExaVMImages changed from 1.55
TiB (406549 extents) to
3.73 TiB (977798 extents).
Logical volume LVDbExaVMImages successfully resized.

```

**16. Resize the OCFS2 file system to the rest of the logical volume.**

The `tunefs.ocfs2` command typically runs very quickly and does not produce output.

```
[root@dm01db01 ~]# tunefs.ocfs2 -S /dev/VGExaDb/LVDbExaVMImages
```

17. Mount the OCFS2 partition and then view the file system disk space usage for this partition.

```
[root@dm01db01 ~]# mount -a
```

```
[root@dm01db01 ~]# ls -al /EXAVMIMAGES/
total 4518924
drwxr-xr-x 3 root root 3896 Jul 18 18:01 .
drwxr-xr-x 26 root root 4096 Jul 24 14:50 ..
drwxr-xr-x 2 root root 3896 Jul 18 17:51 lost+found
-rw-r----- 1 root root 26843545600 Jul 18 18:01
System.first.boot.12.2.1.1.8.180510.1.img
```

```
[root@dm01db01 ~]# df -h /EXAVMIMAGES/
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbExaVMImages 3.8T 9.0G 3.8T 1% /EXAVMIMAGES
```

18. Restart the user domains.

#### Related Topics

- [Shutting Down a User Domain From Within the Management Domain](#)
- [Starting a User Domain](#)

## 5.14.2 Expanding /EXAVMIMAGES on Management Domain on Release 12.2.x

If you are using Oracle Exadata System Software release 12.2.x, then use this procedure to expand the /EXAVMIMAGES file system on the management domain following the addition of a disk expansion kit.

During deployment, all available disk space on a database server will be allocated in the management domain (dom0) with the majority of the space allocated to /EXAVMIMAGES for user domain storage. The /EXAVMIMAGES file system is created on /dev/VGExaDb/LVDbExaVMImages.

In the example below, dm01db01 is the name of the management domain, and dm01db01vm01 is a user domain.

1. Ensure `reclaimdisks.sh` has been run in the management domain by using the `-check` option.

Note that the last line reads "Layout: DOM0". If `reclaimdisks.sh` was not run, it would read "Layout: DOM0 + Linux".

```
[root@dm01db01 ~]# /opt/oracle.SupportTools/reclaimdisks.sh -check
Model is ORACLE SERVER X5-2
Number of LSI controllers: 1
Physical disks found: 4 (252:0 252:1 252:2 252:3)
Logical drives found: 1
Linux logical drive: 0
```

```
RAID Level for the Linux logical drive: 5
Physical disks in the Linux logical drive: 4 (252:0 252:1 252:2
252:3)
Dedicated Hot Spares for the Linux logical drive: 0
Global Hot Spares: 0
Valid. Disks configuration: RAID5 from 4 disks with no global and
dedicated hot spare disks.
Valid. Booted: DOM0. Layout: DOM0.
```

**2. Add the disk expansion kit to the database server.**

The kit consists of 4 additional hard drives to be installed in the 4 available slots. Remove the filler panels and install the drives. The drives may be installed in any order.

**3. Verify that the RAID reconstruction is completed by seeing the warning and clear messages in the alert history.**

This may take several hours to complete. The example below shows that it took approximately 7 hours. Once the clear message (message 1\_2 below) is present, the reconstruction is completed and it is safe to proceed.

```
[root@dm01db01 ~]# dbmcli -e list alerthistory

 1_1 2016-02-15T14:01:00-08:00 warning "A
disk
expansion kit was installed. The additional physical drives were
automatically
added to the existing RAID5 configuration, and reconstruction of
the
corresponding virtual drive was automatically started."

 1_2 2016-02-15T21:01:01-08:00 clear
"Virtual drive
reconstruction due to disk expansion was completed."
```

**4. Collect information about the current environment.**

```
[root@dm01db01 ~]# df -h /EXAVMIMAGES
Filesystem Size Used Avail Use% Mounted on
/dev/sda3 1.6T 44G 1.5T 3% /EXAVMIMAGES
```

```
[root@dm01db01 ~]# xm list
Name ID Mem VCPUs
State Time(s)
Domain-0 0 8192 4
r----- 94039.1
dm01db01vm01.example.com 4 16384 2 -
b----- 3597.3
```

**5. Stop all user domains by running the command `xm shutdown -a -w` from the management domain.**

After all user domains are shut down, only Domain-0 (the management domain) should be listed.

```
[root@dm01db01 ~]# xm shutdown -a -w
Domain dm01db01vm01.example.com terminated
All domains terminated

[root@dm01db01 ~]# xm list
Name ID Mem VCPUs State
Time(s)
Domain-0 0 8192 4 r-----
94073.4
```

## 6. Run parted to view the sector start and end values.

Check the size of the disk against the end of the second partition. If you see a request to fix the GPT, respond with **F**.

```
root@dm01db01 ~]# parted /dev/sda
GNU Parted 2.1Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) unit s
(parted) print
Warning: Not all of the space available to /dev/sda appears to be used,
you can
fix the GPT to use all of the space (an extra 4679680000 blocks) or
continue
with the current setting? Fix/Ignore? F

Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt

Number Start End Size File system Name
Flags
1 64s 1048639s 1048576s ext3 primary boot
2 1048640s 3509759966s 3508711327s primary lvm

(parted) q
```

The partition table shown above lists partition 2 as ending at sector 3509759966s and disk size as 8189440000. You will use these values in step 7.

## 7. Create a third partition.

The start sector is the end of the second partition from step 6 plus 1 sector (3509759966+1=3509759967). The end sector of the third partition is the size of the disk minus 34 (8189440000-34=8189439966).

```
[root@dm01db01 ~]# parted -s /dev/sda mkpart primary 3509759967s
8189439966s
```

This command produces no output.

**8. Set the LVM flag for the third partition.**

```
[root@dm01db01 ~]# parted -s /dev/sda set 3 lvm on
Warning: WARNING: the kernel failed to re-read the partition table
on /dev/sda (Device or
resource busy). As a result, it may not reflect all of your
changes until after reboot.
```

**9. Review the updated partition table.**

```
[root@dm01db01 ~]# parted -s /dev/sda unit s print
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Number Start End Size File system
Name Flags
1 64s 1048639s 1048576s ext4
primary boot
2 1048640s 3509759966s 3508711327s
primary lvm
3 3509759967s 8189439966s 4679680000s
primary lvm
```

**10. Reboot the Exadata server.**

```
[root@dm01db01 ~]# shutdown -r now
```

**11. Check the size of the disk against the end of the third partition.**

```
[root@dm01db01 ~]# parted -s /dev/sda unit s print
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Number Start End Size File system
Name Flags
1 64s 1048639s 1048576s ext4
primary boot
2 1048640s 3509759966s 3508711327s
primary lvm
3 3509759967s 8189439966s 4679680000s
primary lvm
```

**12. Create a LVM physical volume (PV) on the newly created third partition.**

```
[root@dm01db01 ~]# lvm pvcreate --force /dev/sda3
Physical volume "/dev/sda3" successfully created
```

**13. Extend the LVM volume group VGExaDb to the newly created third partition.**

```
[root@dm01db01 ~]# lvm vgextend VGExaDb /dev/sda3
Volume group "VGExaDb" successfully extended
```



**14. Dismount the /EXAVMIMAGES OCFS2 partition.**

```
[root@dm01db01 ~]# umount /EXAVMIMAGES/
```

**15. Extend the logical volume that contains the OCFS2 partition to include the rest of the free space.**

```
[root@dm01db01 ~]# lvm lvextend -l +100%FREE /dev/VGExaDb/LVDbExaVMImages
Size of logical volume VGExaDb/LVDbExaVMImages changed from 1.55 TiB
(406549 extents) to
3.73 TiB (977798 extents).
Logical volume LVDbExaVMImages successfully resized.
```

**16. Resize the OCFS2 file system to the rest of the logical volume.**

The `tunefs.ocfs2` command typically runs very quickly and does not produce output.

```
[root@dm01db01 ~]# tunefs.ocfs2 -S /dev/VGExaDb/LVDbExaVMImages
```

**17. Mount the OCFS2 partition and then view the file system disk space usage for this partition.**

```
[root@dm01db01 ~]# mount -a

[root@dm01db01 ~]# ls -al /EXAVMIMAGES/
total 4518924
drwxr-xr-x 3 root root 3896 Jul 18 18:01 .
drwxr-xr-x 26 root root 4096 Jul 24 14:50 ..
drwxr-xr-x 2 root root 3896 Jul 18 17:51 lost+found
-rw-r----- 1 root root 26843545600 Jul 18 18:01
System.first.boot.12.2.1.1.8.180510.1.img

[root@dm01db01 ~]# df -h /EXAVMIMAGES/
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbExaVMImages 3.8T 9.0G 3.8T 1% /EXAVMIMAGES
```

**18. Restart the user domains.****Related Topics**

- [Shutting Down a User Domain From Within the Management Domain](#)
- [Starting a User Domain](#)

## 5.14.3 Expanding /EXAVMIMAGES on Management Domain on Releases Earlier than 12.2.x

If you are using a release of Oracle Exadata System Software starting with release 12.1.2.3.0 but earlier than release 12.2.x, then use this procedure to expand the /EXAVMIMAGES directory on the management domain following the addition of a disk expansion kit.

During deployment, all available disk space on a database server will be allocated in the management domain with the majority of the space allocated to /EXAVMIMAGES for user domain storage. The /EXAVMIMAGES file system is created on /dev/sda3.

In the example below, dm01db01 is the name of the management domain, and dm01db01vm01 is a user domain.

1. Ensure `reclaimdisks.sh` has been run by using the `-check` option.

Note that the last line reads "Layout: DOM0". If `reclaimdisks.sh` was not run, it would read "Layout: DOM0 + Linux".

```
[root@dm01db01 ~]# /opt/oracle.SupportTools/reclaimdisks.sh -check
Model is ORACLE SERVER X5-2
Number of LSI controllers: 1
Physical disks found: 4 (252:0 252:1 252:2 252:3)
Logical drives found: 1
Linux logical drive: 0
RAID Level for the Linux logical drive: 5
Physical disks in the Linux logical drive: 4 (252:0 252:1 252:2
252:3)
Dedicated Hot Spares for the Linux logical drive: 0
Global Hot Spares: 0
Valid. Disks configuration: RAID5 from 4 disks with no global and
dedicated hot spare disks.
Valid. Booted: DOM0. Layout: DOM0.
```

2. Add the disk expansion kit to the database server.

The kit consists of 4 additional hard drives to be installed in the 4 available slots. Remove the filler panels and install the drives. The drives may be installed in any order.

3. Verify that the RAID reconstruction is completed by seeing the warning and clear messages in the alert history.

This may take several hours to complete. The example below shows that it took approximately 7 hours. Once the clear message (message 1\_2 below) is present, the reconstruction is completed and it is safe to proceed.

```
[root@dm01db01 ~]# dbmcli -e list alerthistory

 1_1 2016-02-15T14:01:00-08:00 warning "A
disk
expansion kit was installed. The additional physical drives were
automatically
added to the existing RAID5 configuration, and reconstruction of
the
corresponding virtual drive was automatically started."

 1_2 2016-02-15T21:01:01-08:00 clear
"Virtual drive
reconstruction due to disk expansion was completed."
```

4. Collect information about the current environment.

```
[root@dm01db01 ~]# cat /proc/partitions |grep sda
 8 0 4094720000 sda
 8 1 524288 sda1
 8 2 119541760 sda2
```

```
8 3 1634813903 sda3
```

```
[root@dm01db01 ~]# df -h /EXAVMIMAGES
Filesystem Size Used Avail Use% Mounted on
/dev/sda3 1.6T 44G 1.5T 3% /EXAVMIMAGES
```

```
[root@dm01db01 ~]# xm list
Name ID Mem VCPUs State
Time(s)
Domain-0 0 8192 4 r-----
94039.1
dm01db01vm01.example.com 4 16384 2 -b-----
3597.3
```

5. Stop all user domain guests by running the command `xm shutdown -a -w` from the management domain.

After all user domain guests are shut down, only Domain-0 (dom0) should be listed.

```
[root@dm01db01 ~]# xm shutdown -a -w
Domain dm01db01vm01.example.com terminated
All domains terminated
```

```
[root@dm01db01 ~]# xm list
Name ID Mem VCPUs State
Time(s)
Domain-0 0 8192 4 r-----
94073.4
```

6. Run `parted` to verify the partition size.

If you see a request to fix the GPT, respond with **F**.

```
root@dm01db01 ~]# parted /dev/sda
GNU Parted 2.1Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) print
Warning: Not all of the space available to /dev/sda appears to be used,
you can
fix the GPT to use all of the space (an extra 4679680000 blocks) or
continue
with the current setting? Fix/Ignore? F
```

```
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 4193GB
Sector size (logical/physical): 512B/512B
Partition Table: gpt
```

Number	Start	End	Size	File system	Name	Flags
1	32.8kB	537MB	537MB	ext3	primary	boot
2	537MB	123GB	122GB		primary	lvm
3	<b>123GB</b>	1797GB	<b>1674GB</b>		primary	

```
(parted) q
```

The partition table shown above lists partition 3 as 1674 GB. The size of the disk that contains this partition (/dev/sda) is 4193 GB.

7. Run parted to view the sector start and end values.

```
root@dm01db01 ~]# parted -s /dev/sda unit s print
Model: LSI MR9361-8i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
```

Number	Start	End	Size	File system	Name
1	64s	1048639s	1048576s	ext3	primary
2	1048640s	240132159s	239083520s		primary
3	<b>240132160s</b>	3509759965s	3269627806s		primary

The partition table shown above lists partition 3 as starting at sector 240132160 and disk size as 8189440000. You will use these values in step 10.

8. Dismount the /EXAVMIMAGES file system on the management domain.

```
[root@dm01db01 ~]# umount /EXAVMIMAGES
```

9. Remove partition 3.

```
[root@dm01db01 ~]# parted -s /dev/sda rm 3
```

This command produces no output.

10. Re-create the partition specifying the same starting sector and the new end of the partition sector. The new end of partition sector is calculated by subtracting 34 from the disk size, for example:  $8189440000 - 34 = 8189439966$

```
[root@dm01db01 ~]# parted -s /dev/sda mkpart primary 240132160s
8189439966s
```

You might encounter the following warning:

```
Warning: WARNING: the kernel failed to re-read the partition table
on /dev/sda
(Device or resource busy). As a result, it may not reflect all of
your changes until after
reboot.
```

If you encounter this error, restart the Exadata database server to apply the changes in the partition table.

11. Mount the `/EXAVMIMAGES` partition again and view the file system disk space usage for this partition.

```
[root@dm01db01 ~]# mount /EXAVMIMAGES

[root@dm01db01 ~]# df -h /EXAVMIMAGES
Filesystem Size Used Avail Use% Mounted on
/dev/sda3 1.6T 44G 1.5T 3% /EXAVMIMAGES
```

Note that the size of the file system is still the same, 1.6 TB, as in step 4.

12. Verify that the partition table as seen by the kernel shows the updated size for partition 3. The output for `sda3` should now be larger compared to the output observed earlier in step 4.

```
[root@dm01db01 ~]# cat /proc/partitions |grep sda
 8 0 4094720000 sda
 8 1 524288 sda1
 8 2 119541760 sda2
 8 3 3974653903 sda3
```

13. Expand the file system.

You can do this while the file system is mounted and processes are running. Note the updated file system size, compared to the value in step 4. The `tunefs.ocfs2` command typically runs very quickly and should have no output normally.

```
[root@dm01db01 ~]# tunefs.ocfs2 -S /dev/sda3

[root@dm01db01 ~]# df -h /EXAVMIMAGES
Filesystem Size Used Avail Use% Mounted on
/dev/sda3 3.8T 44G 3.7T 2% /EXAVMIMAGES
```

14. Restart the user domains.

#### Related Topics

- [Shutting Down a User Domain From Within the Management Domain](#)
- [Starting a User Domain](#)

## 5.15 Adding an Oracle VM Cluster

You can use Oracle Exadata Deployment Assistant (OEDA) to create a new Oracle VM cluster on an existing Exadata Database Machine.

See [Adding a VM Cluster to Oracle Exadata Database Machine Using OEDA](#).

## 5.16 Expanding an Oracle RAC Cluster on Oracle VM Using OEDACLI

You can expand an existing Oracle RAC cluster on Oracle VM by adding user domains using the Oracle Exadata Deployment Assistant command-line interface (OEDACLI).

OEDACLI is the preferred method if you have a known, good version of the OEDA XML file for your cluster.



### Note:

During the execution of this procedure, the existing Oracle RAC cluster nodes along with their database instances incur zero downtime.

Use cases for this procedure include:

- You have an existing Oracle RAC cluster that uses only a subset of the database servers of an Oracle Exadata Rack, and now the nodes not being used by the cluster have become candidates for use.
- You have an existing Oracle RAC cluster on Exadata Database Machine that was recently extended with additional database servers.
- You have an existing Oracle RAC cluster that had a complete node failure and the node was removed and replaced with a newly re-imaged node.

Before performing the steps in this section, the new database servers should have been set up as detailed in [Adding a New Database Server to the Cluster](#), including the following:

- The new database server is installed and configured on the network with a management domain.
- Download the latest Oracle Exadata Deployment Assistant (OEDA); ensure the version you download is the July 2019 release, or later.
- You have an OEDA configuration XML file that accurately reflects the existing cluster configuration. You can validate the XML file by generating an installation template from it and comparing it to the current configuration. See the OEDACLI command `SAVE FILES`.
- Review the OEDA Installation Template report for the current system configuration to obtain node names and IP addresses for existing nodes. You will need to have new host names and IP addresses for the new nodes being added. The new host names and IP addresses required are:
  - Administration host names and IP addresses (referred to as ADMINNET) for the management domain and the user domains.
  - Private host names and IP addresses (referred to as PRIVNET) for the management domain and the user domains.
  - Integrated Lights Out Manager (ILOM) host names and IP addresses for the management domain.

- Client host names and IP addresses (referred to as CLIENTNET) for the user domains.
- Virtual IP (VIP) host names and IP addresses (referred to as VIPNET) for the user domains.
- Physical rack number and location of the new node in the rack (in terms of U number)
- Each management domain has been imaged or patched to the same image in use on the existing database servers. The current system image must match the version of the /EXAVMIMAGES/System.first.boot\*.img file on the new management domain node.

 **Note:**

The ~/dom0\_group file referenced below is a text file that contains the host names of the management domains for all existing and new nodes being added.

Check the image version across all management domains are the same.

```
dcli -g ~/dom0_group -l root "imageinfo -ver"

exa01adm01: 19.2.0.0.0.190225
exa01adm02: 19.2.0.0.0.190225
exa01adm03: 19.2.0.0.0.190225
```

If any image versions differ, you must upgrade the nodes as needed so that they match.

Ensure that the System.first.boot version across all management domains matches the image version retrieved in the previous step.

```
dcli -g ~/dom0_group -l root "ls -l /EXAVMIMAGES/System.first.boot*.img"
exa01adm01: /EXAVMIMAGES/System.first.boot.19.2.0.0.0.190225.img
exa01adm02: /EXAVMIMAGES/System.first.boot.19.2.0.0.0.190225.img
exa01adm03: /EXAVMIMAGES/System.first.boot.19.2.0.0.0.190225.img
```

If **any** nodes are missing the System.first.boot.img file that corresponds to the current image, then obtain the required file. See the “Supplemental README note” for your Exadata release in My Oracle Support Doc ID 888828.1 and look for the patch file corresponding to this description, “DomU System.img OS image for V.V.0.0.0 VM creation on upgraded dom0s”

- Place the kclone.zip files (gi-kclone\*.zip and db-kclone\*.zip) in the /EXAVMIMAGES location on the freshly imaged management domain node you are adding to the cluster. These files can be found in the /EXAVMIMAGES directory on the management domain node from where the system was initially deployed.

The steps here show how to add a new management domain node called exa01adm03 that will have a new user domain called exa01adm03vm01. The steps show how to extend an existing Oracle RAC cluster onto the user domain using OEDACLI commands. The existing cluster has management domain nodes named exa01adm01 and exa01adm02 and user domain nodes named exa01adm01vm01 and exa01adm02vm01.

1. Add the management domain information to the OEDA XML file using the `CLONE COMPUTE` command.

In the examples below, the OEDA XML file is assumed to be in:  
*unzipped\_OEDA\_location/ExadataConfigurations.*

```
OEDACLI> LOAD FILE NAME=exa01_original_deployment.xml

OEDACLI> CLONE COMPUTE SRCNAME=exa01adm01 TGTNAME=exa01adm03
OEDACLI> SET ADMINNET NAME=exa01adm03,IP=xx.xx.xx.xx
OEDACLI> SET PRIVNET NAME1=exa01adm03-
priv1,IP1=xx.xx.xx.xx,NAME2=exa01adm03-priv2,IP2=xx.xx.xx.xx
OEDACLI> SET ILOMNET NAME=exa01adm03-c,IP=xx.xx.xx.xx
OEDACLI> SET RACK NUM=NN,ULOC=XX

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS FORCE
OEDACLI> SAVE FILE NAME=exa01_plus_adm03_node.xml
```

At this point we have a new XML file that has the new compute node management domain in the configuration. This file will be used by the subsequent steps.

2. Add the new user domain information to the OEDA XML file using the `CLONE GUEST` command and deploy the guest.

```
OEDACLI> LOAD FILE NAME=exa01_plus_adm03_node.xml

OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01
WHERE STEPNAME=CREATE_GUEST
OEDACLI> SET PARENT NAME=exa01adm03
OEDACLI> SET ADMINNET NAME=exa01adm03vm01,IP=xx.xx.xx.xx
OEDACLI> SET PRIVNET NAME1=exa01db03vm01-
priv1,IP1=xx.xx.xx.xx,NAME2=exa01db03vm01-priv2,IP2=xx.xx.xx.xx
OEDACLI> SET CLIENTNET NAME=exa01client03vm01,IP=xx.xx.xx.xx
OEDACLI> SET VIPNET NAME=exa01client03vm01-vip,IP=xx.xx.xx.xx

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS
```

If you prefer that OEDACLI runs all steps automatically, omit the `WHERE STEPNAME=CREATE_GUEST` clause from the `CLONE GUEST` command and skip step 3 below.

3. Use OEDACLI to extend the cluster to the new user domain.



 **Note:**

Continue using the same XML file, `exa01_plus_adm03_node.xml` in this example. You will continue to update this file as you proceed through these steps. At the very end of the procedure, this XML file will properly reflect the new state of the clusters.

```
OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=CREATE_USERS
```

```
OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS
```

```
OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=CELL_CONNECTIVITY
```

```
OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS
```

```
OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=ADD_NODE
```

```
OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS
```

```
OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=EXTEND_DBHOME
```

```
OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS
```

```
OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=ADD_INSTANCE
```

```
OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS
```

OEDACLI prints out messages similar to the following as each step completes:

```
Deploying Action ID : 39 CLONE GUEST SRCNAME=exa01adm01vm01
TGTNAME=exa01adm03vm01 where STEPNAME=ADD_INSTANCE
Deploying CLONE GUEST
Cloning Guest
Cloning Guest : exa01adm03vm01.example.com_id
Adding new instance for database [dbm] on exa01adm03vm01.example.com
Setting up Huge Pages for Database..[dbm]
Adding instance dbm3 on host exa01adm03vm01.example.com
```

```
Successfully completed adding database instance on the new node
[elapsed Time [Elapsed =
249561 mS [4.0 minutes] Fri Jun 28 13:35:52 PDT 2019]]
Done...
Done
```

4. Save the current state of the configuration and generate configuration information.

```
OEDACLI> SAVE FILES LOCATION=/tmp/exa01_plus_adm03_config
```

The above command writes all the configuration files to the directory `/tmp/exa01_plus_adm03_config`. Save a copy of these files in a safe place since they now reflect the changes made to your cluster.

5. Gather an Oracle EXAchk report and examine it to ensure the cluster is in good health.

#### Related Topics

- OEDA Command Line Interface
- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

## 5.17 Creating a User Domain Without Oracle Grid Infrastructure and Oracle Database

A user domain can be created without Oracle Grid Infrastructure and Oracle Database installed on the system. The new user domain has the following characteristics:

- Operating system image is Oracle Linux
- Access to the management, client, and InfiniBand networks
- No Oracle Grid Infrastructure and Oracle Database is installed

The following procedure creates a user domain without Oracle Grid Infrastructure and Oracle Database installed:

1. Allocate new, unused, IP addresses and host names for the new user domain. IP addresses and host names are needed for the management network, client (SCAN) network, and the private InfiniBand network.

#### Note:

Ensure the intended InfiniBand network IP addresses are unused by using the `ping` command for each address. The `ibhosts` command cannot be used to determine all InfiniBand network IP addresses in use because it does not contain entries for user domains.

2. If necessary, obtain an updated user domain (domU) system image file.

The `exadata.img.domu_maker` command that you will run later in this procedure to create a user domain requires the user domain (domU) system image file `System.first.boot.version.img` in `/EXAVMIMAGES`, where `version` matches the

management domain Exadata software version as determined by running the "imageinfo -ver" command in the management domain.

For example, when `exadata.img.domu_maker` is run to create a new user domain and the management domain Exadata software version is 12.1.2.1.1.150316.2, the user domain (domU) system image file `/EXAVMIMAGES/System.first.boot.12.1.2.1.1.150316.2.img` must exist.

```
imageinfo -ver
12.1.2.1.1.150316.2

ls -l /EXAVMIMAGES/System.first.boot.12.1.2.1.1.150316.2.img
-rw-r--r-- 1 root root 13958643712 Mar 23 12:25 /EXAVMIMAGES/
System.first.boot.12.1.2.1.1.150316.2.img
```

If the user domain (domU) system image file does not exist, then it must be obtained from My Oracle Support and placed in `/EXAVMIMAGES` in the management domain. See My Oracle Support note 888828.1 for additional information.

3. In the management domain, copy an existing XML configuration file from a deployed user domain to a new file name using the following command:

```
cp /EXAVMIMAGES/conf/existingDomainName-vm.xml /EXAVMIMAGES/conf/newDomainName-
vm.xml
```

In the preceding command, `existingDomainName-vm.xml` is the XML configuration file of the deployed user domain, and `newDomainName-vm.xml` is the name of the new file.

In the following example, the configuration file for user domain "dm01db01vm01" is copied to `nondbdomain-vm.xml`.

```
cp /EXAVMIMAGES/conf/dm01db01vm01-vm.xml /EXAVMIMAGES/conf/nondbdomain-vm.xml
```

4. In the management domain, edit the new XML file as follows:
  - a. Change all `<Hostname>` tags to match the new host names for the respective networks.
  - b. Change all `<IP_address>` tags to match the new IP addresses for the respective networks.
  - c. Change the `<virtualMachine>` tag to contain the new host name.
  - d. Change the `<hostName>` tag to contain the new host name.
  - e. Delete the entire `<disk id="disk_2">` and `<disk id="disk_3">` elements, including all their sub-elements. You must delete the entire entry between the starting `<disk>` tag to the corresponding closing `</disk>`.
5. In the management domain, allocate InfiniBand network GUIDs for the new user domain using the `/opt/exadata_ovm/exadata.img.domu_maker` command.

```
/opt/exadata_ovm/exadata.img.domu_maker allocate-guids \
 /EXAVMIMAGES/conf/newDomainName-vm.xml \
 /EXAVMIMAGES/conf/final-newDomainName-vm.xml
```

6. In the management domain, create the new user domain using the `/opt/exadata_ovm/exadata.img.domu_maker` command.

```
/opt/exadata_ovm/exadata.img.domu_maker start-domain \
 /EXAVMIMAGES/conf/final-newDomainName-vm.xml
```

## 5.18 Moving a User Domain to a Different Database Server

User domains can move to different database servers.

The target Oracle Exadata Database Server must meet the following requirements:

- The target database server must have the same Oracle Exadata System Software release installed with Oracle VM.
- The target database server must have the same network visibility.
- The target database server must have access to the same Oracle Exadata Storage Servers.
- The target database server must have sufficient free resources (CPU, memory, and local disk storage) to operate the user domain.
  - It is possible to over-commit virtual CPUs such that the total number of virtual CPUs assigned to all domains exceeds the number of physical CPUs on the system. Over-committing CPUs can be done only when the competing workloads for over-subscribed resources are well understood and the concurrent demand does not exceed physical capacity.
  - It is not possible to over-commit memory.
  - Copying disk images to the target database server may increase space allocation of the disk image files because the copied files are no longer able to benefit from the disk space savings gained by using OCFS2 reflinks.
- The user domain name must not be already in use on the target database server.

The following procedure moves a user domain to a new database server in the same Oracle Exadata System Software configuration. All steps in this procedure are performed in the management domain.

1. Shut down the user domain.

```
xm shutdown DomainName -w
```

2. Copy the user domain disk image and configuration files to the target database server.

In the following examples, replace *DomainName* with the name of the domain.

```
scp -r /EXAVMIMAGES/GuestImages/DomainName/ target:/EXAVMIMAGES/
GuestImages
```

3. Obtain the UUID of the user domain.

```
grep ^uuid /EXAVMIMAGES/GuestImages/DomainName/vm.cfg
```

An example of the user domain UUID is 49ffddce4efe43f5910d0c61c87bba58.

- Using the UUID of the user domain, copy the user domain symbolic links from `/OVS/Repositories` to the target database server.

```
tar cpvf - /OVS/Repositories/UUID/ | ssh target_db_server "tar xpvf - -
C /"
```

- Start the user domain on the target database server.

```
xm create /EXAVMIMAGES/GuestImages/DomainName/xm.cfg
```

## 5.19 Backing up the Management Domain and User Domains in an Oracle VM Deployment

In an Oracle VM deployment, you need to back up the management domain (dom0) and the user domains (domU):

- [Backing up the Management Domain dom0 Using Snapshot-Based Backup](#)  
This procedure describes how to take a snapshot-based backup of the management domain, dom0.
- [Backing up the User Domains](#)  
You can create a backup of all the user domains on a host, or of individual user domains.

### 5.19.1 Backing up the Management Domain dom0 Using Snapshot-Based Backup

This procedure describes how to take a snapshot-based backup of the management domain, dom0.

The logical volume `/dev/VGExaDb/LVDoNotRemoveOrUse` is a placeholder to make sure there is always free space available to create a snapshot. If you run `dbserver_backup.sh`, then the placeholder LVM is removed by the script, the free space is used for a snapshot, and the LVM is re-created after the snapshot is created. If you follow the manual procedure described here, then you have to perform all these tasks manually.

The values shown in the steps below are examples. All steps must be performed as the `root` user.

- Prepare a destination to hold the backup.

The destination should reside outside of the local machine, such as a writable NFS location, and be large enough to hold the backup tar file(s). For non-customized partitions, the space needed for holding the backup is around 60 GB.

The following commands may be used to prepare the backup destination.

```
mkdir -p /remote_FS

mount -t nfs -o rw,intr,soft,proto=tcp,nolock ip_address:/
nfs_location/ /remote_FS
```

`ip_address` is the IP address of the NFS server, and `nfs_location` is the NFS location holding the backups.

**2. Take a snapshot-based backup of the file system hosting the / (root) directory.**

**a. Check for the existence of the `LVDoNotRemoveOrUse` logical volume.**

If this volume is present, then remove the volume to make space for the snapshot. Execute the script below to check for the existence of the `LVDoNotRemoveOrUse` logical volume and remove it if present.

```
lvm lvdisplay --ignorelockingfailure /dev/VGExaDb/
LVDoNotRemoveOrUse
if [$? -eq 0]; then
 # LVDoNotRemoveOrUse logical volume exists.
 lvm lvremove -f /dev/VGExaDb/LVDoNotRemoveOrUse
 if [$? -ne 0]; then
 echo "Unable to remove logical volume:
LVDoNotRemoveOrUse. Unable to proceed with backup"
 fi
fi
```

If the `LVDoNotRemoveOrUse` logical volume does not exist, then investigate the reason and do not proceed with the steps below.

**b. Create a snapshot named `LVDbSys3_snap` for the file system hosting the / (root) directory.**

This example assumes `LVDbSys3` is the active partition.

```
lvcreate -L1G -s -n LVDbSys3_snap /dev/VGExaDb/LVDbSys3
```

**c. Label the snapshot.**

```
e2label /dev/VGExaDb/LVDbSys3_snap DBSYSOVS_SNAP
```

**d. Mount the snapshot.**

```
mkdir /root/mnt
mount /dev/VGExaDb/LVDbSys3_snap /root/mnt -t ext4
```

**e. Change to the directory for the backup.**

```
cd /root/mnt
```

**f. Create the backup file.**

```
tar -pjcvf /remote_FS/mybackup.tar.bz2 * /boot > /tmp/
backup_tar.stdout 2> /tmp/backup_tar.stderr
```

**g. Check the `/tmp/backup_tar.stderr` file for any significant errors.**

Errors about failing to tar open sockets, and other similar errors, can be ignored.

3. Unmount the snapshot and remove the snapshot for the root directory.

```
cd /
umount /root/mnt
/bin/rmdir /root/mnt
lvremove /dev/VGExaDb/LVDbSys3_snap
```

4. Unmount the NFS share.

```
umount /remote_FS
```

5. Recreate the `/dev/VGExaDb/LVDoNotRemoveOrUse` logical volume.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L1G VGExaDb
```

## 5.19.2 Backing up the User Domains

You can create a backup of all the user domains on a host, or of individual user domains.

There are three ways to back up the user domains:

- Method 1: Back up all user domains in the storage repository using Oracle Cluster File System (OCFS) reflinks to get a consistent backup

This method backs up the storage repository that is the `/EXAVMIMAGES` OCFS2 file system. This method provides a more robust and a comprehensive backup than method 2 or 3. Method 3 provides a quicker and an easier backup method, especially in role separated environments.

Method 1 is best-suited for when a management domain (dom0) administrator is responsible for user domain backups.

- Method 2: Back up individual user domains in the storage repository using Oracle Cluster File System (OCFS) reflinks to get a consistent backup.

You select which user domains you want to back up from the `/EXAVMIMAGES` OCFS2 file system. The user domains are located in the `/EXAVMIMAGES/GuestImages/user` directories.

Method 2 is best-suited for when a management domain (dom0) administrator is responsible for user domain backups.

- Method 3: Back up a user domain using snapshot-based backup

This method backs up a single user domain using snapshot-based backup from inside the user domain.

Method 3 is ideal where a user domain administrator is responsible for the user domain backups.

- [Method 1: Back up All the User Domains](#)

You can back up all the user domains by backing up the storage repository that is the `/EXAVMIMAGES` OCFS2 file system.

- [Method 2: Back up Individual User Domains](#)

You can back up an individual user domain by backing up its specific folder in `/EXAVMIMAGES` file system.

- **Method 3: Back up a User Domain from Inside the User Domain**  
You can take a snapshot-based backup of a user domain from inside the user domain, which can then be used to restore the user domain to a workable state.

### 5.19.2.1 Method 1: Back up All the User Domains

You can back up all the user domains by backing up the storage repository that is the /EXAVMIMAGES OCFS2 file system.

The backup destination should reside outside of the local machine, such as a writable NFS location, and be large enough to hold the backup. The space needed for the backup is proportional to the number of Oracle VMs deployed on the system, up to a maximum space of about 1.6 TB.

This procedure assumes there are 15 or less user domains per management domain.

1. Use the following script to prepare the backup destination and prepare the user domains for backup.

```
ScriptStarttime=$(date +%s)
printf "This script is going to remove the directory /EXAVMIMAGES/
Backup.
If that is not acceptable, exit the script by typing n, manually
remove /EXAVMIMAGES/Backup and come back to rerun the script.
Otherwise,
press y to continue :"
read proceed

if [[${proceed} == "n"]] || [[${proceed} == "N"]]
then
 exit 0
fi

rm -rf /EXAVMIMAGES/Backup

Create the Backup Directory

mkdirStartTime=$(date +%s)
find /EXAVMIMAGES -type d|grep -v 'lost+found'|awk '{print "mkdir -
p /EXAVMIMAGES/Backup"$1}'|sh
mkdirEndTime=$(date +%s)
mkdirTime=$(expr ${mkdirEndTime} - ${mkdirStartTime})
echo "Backup Directory creation time : " ${mkdirTime}" seconds"

Create reflinks for files not in /EXAVMIMAGES/GuestImages
relinkothesStartTime=$(date +%s)
find /EXAVMIMAGES/ -not -path "/EXAVMIMAGES/GuestImages/*" -not -
path "/EXAVMIMAGES/Backup/*" -type f|awk '{print "reflink", $0, "/
EXAVMIMAGES/Backup"$0}'|sh
relinkothesEndTime=$(date +%s)
relinkothesTime=$(expr ${relinkothesEndTime} - $
{relinkothesStartTime})
echo "Reflink creation time for files other than in /EXAVMIMAGES/
GuestImages : " ${relinkothesTime}" seconds"
```



```

Pause the user domains
for hostName in $(xm list|egrep -v '^Domain-0|^Name'|awk '{print $1}')
do
PauseStartTime=$(date +%s)
xm pause ${hostName}
PauseEndTime=$(date +%s)
PauseTime=$(expr ${PauseEndTime} - ${PauseStartTime})
echo "PauseTime for guest - ${hostName} : " ${PauseTime}" seconds"

Create reflinks for all the files in /EXAVMIMAGES/GuestImages
relinkStartTime=$(date +%s)
find /EXAVMIMAGES/GuestImages/${hostName} -type f|awk '{print
"relink", $0, "/EXAVMIMAGES/Backup"$0}'|sh
relinkEndTime=$(date +%s)
relinkTime=$(expr ${relinkEndTime} - ${relinkStartTime})
echo "Reflink creation time for guest - ${hostName} : " ${relinkTime}"
seconds"

Unpause the user domains
unPauseStartTime=$(date +%s)
xm unpause ${hostName}
unPauseEndTime=$(date +%s)
unPauseTime=$(expr ${unPauseEndTime} - ${unPauseStartTime})
echo "unPauseTime for guest - ${hostName} : " ${unPauseTime}" seconds"

ScriptEndtime=$(date +%s)
ScriptRunTime=$(expr ${ScriptEndtime} - ${ScriptStarttime})
echo "ScriptRunTime ${ScriptRunTime} seconds"

```

2. Create a backup of the snapshot.

Backup the reflink files in the `/EXAVMIMAGES/Backup` directory that was created with the script in Step 1 to a remote location. For example:

- a. Create a tarball file comprising of all files under `/EXAVMIMAGES/Backup`.
- b. Copy the tarball to a remote location.

This allows for restore operations if the management domain (Dom0) is permanently lost or damaged.

3. Remove the reflinks created by the script.

## 5.19.2.2 Method 2: Back up Individual User Domains

You can back up an individual user domain by backing up its specific folder in `/EXAVMIMAGES` file system.

The backup destination should reside outside of the local machine, such as a writable NFS location, and be large enough to hold the backup. The space needed for the backup is proportional to the number of Oracle VMs deployed on the system, up to a maximum space of about 1.6 TB.

1. Use the following script to prepare the backup destination and prepare the user domain for backup.

```

ScriptStarttime=$(date +%s)
printf "This script is going to remove the directory /EXAVMIMAGES/Backup.

```

```

If that is not acceptable, exit the script by typing n, manually
remove /EXAVMIMAGES/Backup and come back to rerun the script.
Otherwise,
press y to continue : "
read proceed

if [[${proceed} == "n"]] || [[${proceed} == "N"]]
then
 exit 0
fi

rm -rf /EXAVMIMAGES/Backup

printf "Enter the name of the user domains to be backed up : "
read userDomainName

Create the Backup Directory

mkdirStartTime=$(date +%s)
find /EXAVMIMAGES/GuestImages/${userDomainName} -type d | grep -v
'lost+found' | awk '{print "mkdir -p /EXAVMIMAGES/Backup"$1}' | sh
mkdirEndTime=$(date +%s)
mkdirTime=$(expr ${mkdirEndTime} - ${mkdirStartTime})
echo "Backup Directory creation time : " ${mkdirTime} " seconds"

Pause the user domain
PauseStartTime=$(date +%s)
xm pause ${userDomainName}
PauseEndTime=$(date +%s)
PauseTime=$(expr ${PauseEndTime} - ${PauseStartTime})
echo "PauseTime for guest - ${userDomainName} : " ${PauseTime} "
seconds"

Create reflinks for all the files in /EXAVMIMAGES/GuestImages/$
{userDomainName}
relinkStartTime=$(date +%s)
find /EXAVMIMAGES/GuestImages/${userDomainName} -type f | awk '{print
"relink", $0, "/EXAVMIMAGES/Backup"$0}' | sh
relinkEndTime=$(date +%s)
relinkTime=$(expr ${relinkEndTime} - ${relinkStartTime})
echo "Reflink creation time for guest - ${userDomainName} : " $
{relinkTime} " seconds"

Unpause the user domain
unPauseStartTime=$(date +%s)
xm unpause ${userDomainName}
unPauseEndTime=$(date +%s)
unPauseTime=$(expr ${unPauseEndTime} - ${unPauseStartTime})
echo "unPauseTime for guest - ${userDomainName} : " ${unPauseTime} "
seconds"

ScriptEndtime=$(date +%s)
ScriptRunTime=$(expr ${ScriptEndtime} - ${ScriptStarttime})
echo "ScriptRunTime ${ScriptRunTime} seconds"

```

2. Create a backup of the snapshot.

Backup the reflink files in the `/EXAVMIMAGES/Backup` directory that was created with the script in Step 1 to a remote location. For example:

- a. Create a tarball file comprising of all files under `/EXAVMIMAGES/Backup`.
- b. Copy the tarball to a remote location.

This allows for restore operations if the management domain (Dom0) is permanently lost or damaged.

3. Remove the reflinks created by the script.

### 5.19.2.3 Method 3: Back up a User Domain from Inside the User Domain

You can take a snapshot-based backup of a user domain from inside the user domain, which can then be used to restore the user domain to a workable state.

All steps are performed from inside the user domain.

#### Note:

This method of backing up a user domain from inside the user domain using LVM snapshots will have limited usage in terms of recovery. Such a backup can only be used for recovery purposes when the user domain is still bootable and allows login as the root user. This means the damage is such that some files have been lost or damaged but can be restored from the tar backup after the user domain has booted up and the `/` (root) partition and the boot partitions are mounted. If that is not the case and the damage is such that the user domain does not boot, then you need a backup taken using methods 1 or 2 above to recover the user domain, and you need to perform the recovery procedure at the user domain level using the recovery procedure described below.

This procedure backs up the following:

- LVDbSys1
- LVDbOra1
- `/boot` partition
- Grid Infrastructure home
- RDBMS home

All steps must be performed as the root user.

1. Prepare a destination to hold the backup.

In the following example, `ip_address` is the IP address of the NFS server, and `nfs_location` is the NFS location holding the backups.

```
mkdir -p /remote_FS

mount -t nfs -o rw,intr,soft,proto=tcp,nolock ip_address:/
nfs_location/ /remote_FS
```

2. Take a snapshot-based backup of the file systems containing / (root) and the /u01 directories, as follows:

- a. Create a snapshot named `LVDbSys1_snap` for the file system containing the root directory.

The volume group must have at least 1 GB of free space for the command to succeed.

```
lvcreate -L1G -s -n LVDbSys1_snap /dev/VGExaDb/LVDbSys1
```

- b. Label the snapshot.

```
e2label /dev/VGExaDb/LVDbSys1_snap DBSYS_SNAP
```

- c. Mount the snapshot.

```
mkdir /root/mnt
```

```
mount /dev/VGExaDb/LVDbSys1_snap /root/mnt -t ext4
```

- d. Create a snapshot named `u01_snap` for the /u01 directory.

```
lvcreate -L256M -s -n u01_snap /dev/VGExaDb/LVDbOra1
```

- e. Label the snapshot.

```
e2label /dev/VGExaDb/u01_snap DBORA_SNAP
```

- f. Mount the snapshot.

```
mkdir -p /root/mnt/u01
```

```
mount /dev/VGExaDb/u01_snap /root/mnt/u01 -t ext4
```

- g. Change to the directory for the backup.

```
cd /root/mnt
```

- h. Create the backup file to back up the two snapshots taken above, the /boot partition, the Oracle Database home directory, and the Oracle Grid Infrastructure home directory.

In the following example: *Grid\_home* is the location of the Oracle Grid Infrastructure home, for example, `/u01/app/18.1.0/grid`; *DB\_home* is the location of the Oracle Database home, for example, `/u01/app/oracle/product/18.1.0/dbhome_1`.

```
tar -pjcvf /remote_FS/mybackup.tar.bz2 * /boot Grid_home DB_home > /tmp/backup_tar.stdout 2> /tmp/backup_tar.stderr
```

- i. Check the `/tmp/backup_tar.stderr` file for any significant errors.

Errors about failing to `tar` open sockets, and other similar errors, can be ignored.

3. Unmount and remove the snapshots for the file system containing the root directories.

```
cd /
umount /root/mnt/u01
umount /root/mnt
/bin/rmdir /root/mnt
lvremove /dev/VGExaDb/u01_snap
lvremove /dev/VGExaDb/LVDbSys1_snap
```

4. Unmount the NFS share.

```
umount /remote_FS
```

## 5.20 Recovering an Oracle VM Deployment

You can recover an Oracle VM from a snapshot-based backup when severe disaster conditions damage the Oracle VM, or when the server hardware is replaced to such an extent that it amounts to new hardware.

For example, replacing all hard disks leaves no trace of original software on the system. This is similar to replacing the complete system as far as the software is concerned. In addition, it provides a method for disaster recovery of the database servers using an LVM snapshot-based backup taken when the database server was healthy before the disaster condition.

The recovery procedures described in this section do not include backup or recovery of storage servers or the data in an Oracle Database. Oracle recommends testing the backup and recovery procedures on a regular basis.

- [Overview of Snapshot-Based Recovery of Database Servers](#)  
The recovery of the Oracle VM consists of a series of tasks.
- [Scenario 1: Recovering a Management Domain and Its User Domains from Backup](#)  
You can recover the management domain and all its user domains from a backup.
- [Scenario 2: Re-imaging the Management Domain and Restoring User Domains from Backups](#)  
This procedure re-images the management domain and reconstructs all the user domains.
- [Scenario 3: Restoring and Recovering User Domains from Snapshot Backups](#)  
Use this procedure to restore lost or damaged files of a user domain using a snapshot-based user domain backup taken from inside a user domain.

### 5.20.1 Overview of Snapshot-Based Recovery of Database Servers

The recovery of the Oracle VM consists of a series of tasks.

The recovery procedures use the `diagnostics.iso` image as a virtual CD-ROM to restart the Oracle VM in rescue mode using the Integrated Lights Out Manager (ILOM). At a high-level, the steps look like this:

1. Re-create the following:
  - Boot partitions
  - Physical volumes
  - Volume groups

- Logical volumes
  - File system
  - Swap partition
2. Activate the swap partition.
  3. Ensure the `/boot` partition is the active boot partition.
  4. Restore the data.
  5. Reconfigure GRUB.
  6. Restart the server.

## 5.20.2 Scenario 1: Recovering a Management Domain and Its User Domains from Backup

You can recover the management domain and all its user domains from a backup.

The following procedures step you through the recovery process. Chose one of the following procedures, based on the version of Oracle Exadata System Software that is installed on your system.

### **WARNING:**

All existing data on the disks is lost during these procedures.

- [Recovering a Management Domain and Its User Domains \(Releases Prior to 12.2.1.1.0\)](#)  
You can recover a management domain from a snapshot-based backup when severe disaster conditions damage the dom0, or when the server hardware is replaced to such an extent that it amounts to new hardware.
- [Recovering a Management Domain and Its User Domains \(Releases 12.2.1.1.0 and Later\)](#)  
You can recover a management domain from a snapshot-based backup when severe disaster conditions damage the management domain, or when the server hardware is replaced to such an extent that it amounts to new hardware.
- [Recovering a Management Domain and Its User Domains \(Release 18.1 and X7 and Later\)](#)  
You can recover a management domain from a snapshot-based backup when severe disaster conditions damage the management domain, or when the server hardware is replaced to such an extent that it amounts to new hardware.

### Related Topics

- [Backing up the Management Domain dom0 Using Snapshot-Based Backup](#)  
This procedure describes how to take a snapshot-based backup of the management domain, `dom0`.

## 5.20.2.1 Recovering a Management Domain and Its User Domains (Releases Prior to 12.2.1.1.0)

You can recover a management domain from a snapshot-based backup when severe disaster conditions damage the dom0, or when the server hardware is replaced to such an extent that it amounts to new hardware.

To use this recovery method, it is assumed that you have previously completed the steps in [Backing up the Management Domain dom0 Using Snapshot-Based Backup](#).

1. Prepare an NFS server to host the backup archive `mybackup.tar.bz2`.

The NFS server must be accessible by IP address. For example, on an NFS server with the IP address `nfs_ip`, where the directory `/export` is exported from NFS mounts, put the `mybackup.tar.bz2` file in the `/export` directory

2. Restart the recovery target system using the `diagnostics.iso` file.

See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.

3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. If required, use `/opt/MegaRAID/MegaCli/MegaCli64` to configure the disk controller to set up the disks.
5. Remove the logical volumes, the volume group, and the physical volume, in case they still exist after the disaster.

```
lvm vgremove VGExaDb --force
lvm pvremove /dev/sda2 --force
```

6. Remove the existing partitions and clean up the drive.

```
parted
GNU Parted 2.1
Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) rm 1
sda: sda2 sda3
(parted) rm 2
sda: sda3
```

```
(parted) rm 3
sda:
(parted) q

dd if=/dev/zero of=/dev/sda bs=64M count=2
```

**7. Create the three partitions on /dev/sda.**

- a. Get the end sector for the disk /dev/sda from a surviving dom0 and store it in a variable:**

```
end_sector=$(parted -s /dev/sda unit s print|perl -ne '/
^Disk\s+\S+:\s+(\d+)s/ and print $1')
```

- b. Create the boot partition, /dev/sda1.**

```
parted -s /dev/sda mklabel gpt mkpart primary 64s 1048639s set
1 boot on
```

- c. Create the partition that will hold the LVMS, /dev/sda2.**

```
parted -s /dev/sda mkpart primary 1048640s 240132159s set 2
lvm on
```

- d. Create the OCFS2 storage repository partition, /dev/sda3.**

```
parted -s /dev/sda mkpart primary 240132160s ${end_sector}s
set 3
```

**8. Use the /sbin/lvm command to re-create the logical volumes and mkfs to create file systems.**

- a. Create the physical volume and the volume group.**

```
lvm pvcreate /dev/sda2
lvm vgcreate VGExaDb /dev/sda2
```

- b. Create the logical volume for the file system that will contain the / (root) directory and label it.**

```
lvm lvcreate -n LVDbSys3 -L30G VGExaDb
mkfs.ext4 /dev/VGExaDb/LVDbSys3
e2label /dev/VGExaDb/LVDbSys3 DBSYSOVS
```

- c. Create the logical volume for the swap directory, and label it.**

```
lvm lvcreate -n LVDbSwap1 -L24G VGExaDb
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

- d. Create the logical volume for the backup partition, and build a file system on top of it.**

```
lvm lvcreate -n LVDbSys2 -L30G VGExaDb
mkfs.ext4 /dev/VGExaDb/LVDbSys2
```



- e. Create the logical volume for the reserved partition.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L1G VGExaDb
```

 **Note:**

Do not create any file system on this logical volume.

- f. Create a file system on the `/dev/sda1` partition, and label it.

In the `mkfs.ext3` command below, the `-I 128` option is needed to set the inode size to 128.

```
mkfs.ext3 -I 128 /dev/sda1
tune2fs -c 0 -i 0 /dev/sda1
e2label /dev/sda1 BOOT
```

9. Create mount points for all the partitions, and mount the respective partitions.

For example, if `/mnt` is used as the top level directory, the mounted list of partitions may look like:

- `/dev/VGExaDb/LVDbSys3` on `/mnt`
- `/dev/sda1` on `/mnt/boot`

The following example mounts the root file system, and creates two mount points:

```
mount /dev/VGExaDb/LVDbSys3 /mnt -t ext4
mkdir /mnt/boot
mount /dev/sda1 /mnt/boot -t ext3
```

10. Bring up the network on `eth0` and assign the host's IP address and netmask to it.

```
ifconfig eth0 ip_address_for_eth0 netmask netmask_for_eth0 up
route add -net 0.0.0.0 netmask 0.0.0.0 gw gateway_ip_address
```

11. Mount the NFS server holding the backups.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

12. From the backup which was created in [Backing up the Management Domain dom0 Using Snapshot-Based Backup](#), restore the root directory (`/`) and the boot file system.

```
tar -pxvf /root/mnt/backup-of-root-and-boot.tar -C /mnt
```

13. Unmount the restored `/dev/sda1` partition, and remount it on `/boot`.

```
umount /mnt/boot
mkdir /boot
mount /dev/sda1 /boot -t ext3
```

14. Set up the `grub` boot loader using the command below:

```
grub --device-map=/boot/grub/device.map << DOM0_GRUB_INSTALL
root (hd0,0)
setup (hd0)
quit
DOM0_GRUB_INSTALL
```

15. Unmount the `/boot` partition.

```
umount /boot
```

16. Detach the `diagnostics.iso` file.

Using the ILOM Web interface, navigate to the Storage Devices dialog and click **Disconnect**.

The Storage Devices dialog is the interface that you earlier used to attach the `diagnostics.iso` image. See [Booting a Server using the Diagnostic ISO File in Oracle Exadata System Software User's Guide](#).

17. Check the restored `/etc/fstab` file and comment out any reference to `/EXAVMIMAGES` and `/dev/sda3`.

```
cd /mnt/etc
```

Comment out any line that references `/EXAVMIMAGES` or `/dev/sda3`.

18. Restart the system.

```
shutdown -r now
```

This completes the restoration procedure for the management domain (`dom0`).

19. Convert to eighth rack, if required.

If the recovery is on an Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

20. When the server comes back up, build an OCFS2 file system on the `/dev/sda3` partition.

```
mkfs -t ocfs2 -L ocfs2 -T vmstore --fs-features=local /dev/sda3 --
force
```

21. Mount the OCFS2 partition `/dev/sda3` on `/EXAVMIMAGES`.

```
mount -t ocfs2 /dev/sda3 /EXAVMIMAGES
```

22. In `/etc/fstab`, uncomment the references to `/EXAVMIMAGES` and `/dev/sda3`, which you commented out earlier.

23. Mount the backup NFS server that holds the storage repository (`/EXAVMIMAGES`) backup to restore the `/EXAVMIMAGES` file system which holds all the user domain images.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

24. Restore the `/EXAVMIMAGES` file system.

To restore all user domains, use this command:

```
tar -Spxvf /root/mnt/backup-of-exavmimages.tar -C /EXAVMIMAGES
```

To restore a single user domain from the backup, use the following command instead:

```
tar -Spxvf /root/mnt/backup-of-exavmimages.tar -C /EXAVMIMAGES
EXAVMIMAGES/<user-domain-name-to-be-restored>
```

25. Bring up each user domain.

```
xm create /EXAVMIMAGES/GuestImages/user_domain_hostname/vm.cfg
```

At this point all the user domains should come up along with Oracle Grid Infrastructure and the Oracle Database instances. The database instances should join the Oracle Real Application Clusters (Oracle RAC) cluster formed by the other surviving management domain nodes.

## 5.20.2.2 Recovering a Management Domain and Its User Domains (Releases 12.2.1.1.0 and Later)

You can recover a management domain from a snapshot-based backup when severe disaster conditions damage the management domain, or when the server hardware is replaced to such an extent that it amounts to new hardware.

To use this recovery method, it is assumed that you have previously completed the steps in [Backing up the Management Domain dom0 Using Snapshot-Based Backup](#).

1. Prepare an NFS server to host the backup archive `mybackup.tar.bz2`.

The NFS server must be accessible by IP address. For example, on an NFS server with the IP address `nfs_ip`, where the directory `/export` is exported from NFS mounts, put the `mybackup.tar.bz2` file in the `/export` directory

2. Restart the recovery target system using the `diagnostics.iso` file.

See [Booting a Server using the Diagnostic ISO File](#) in *Oracle Exadata System Software User's Guide*.

3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
```

```

from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.

```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. If required, use `/opt/MegaRAID/storcli/storcli64` (or `/opt/MegaRAID/MegaCli/MegaCli64` for releases earlier than Oracle Exadata System Software 19c) to configure the disk controller to set up the disks.
5. Remove the logical volumes, the volume group, and the physical volume, in case they still exist after the disaster.

```

lvm vgremove VGExaDb --force
lvm pvremove /dev/sda2 --force

```

6. Remove the existing partitions and clean up the drive.

```

parted
GNU Parted 2.1
Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) rm 1
[12064.253824] sda: sda2
(parted) rm 2
[12070.579094] sda:
(parted) q

dd if=/dev/zero of=/dev/sda bs=64M count=2

```

7. Create the two partitions on `/dev/sda`.
  - a. Get the end sector for the disk `/dev/sda` from a surviving `dom0` and store it in a variable:

```

end_sector_logical=$(parted -s /dev/sda unit s print|perl -ne
'^Disk\s+\S+:\s+(\d+)s/ and print $1')
end_sector=$(expr $end_sector_logical - 34)

```

The values for the start and end sectors in the commands below were taken from a surviving management domain. Because these values can change over time, it is recommended that these values are checked from a surviving management domain using the following command:

```

parted -s /dev/sda unit S print

```

- b. Create the boot partition, `/dev/sda1`.

```

parted -s /dev/sda mklabel gpt mkpart primary 64s 1048639s set
1 boot on

```

- c. Create the partition that will hold the LVMS, /dev/sda2.

```
parted -s /dev/sda mkpart primary 1048640s 3509759966s set 2 lvm on
```

8. Use the /sbin/lvm command to re-create the logical volumes and mkfs to create file systems.

- a. Create the physical volume and the volume group.

```
lvm pvcreate /dev/sda2
lvm vgcreate VGExaDb /dev/sda2
```

- b. Create the logical volume for the file system that will contain the / (root) directory and label it.

```
lvm lvcreate -n LVDbSys3 -L30G VGExaDb
mkfs -t ext4 -b 4096 /dev/VGExaDb/LVDbSys3
e2label /dev/VGExaDb/LVDbSys3 DBSYSOVS
```

- c. Create the logical volume for the swap directory, and label it.

```
lvm lvcreate -n LVDbSwap1 -L24G VGExaDb
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

- d. Create the logical volume for the backup partition, and build a file system on top of it.

```
lvm lvcreate -n LVDbSys2 -L30G VGExaDb
mkfs -t ext4 -b 4096 /dev/VGExaDb/LVDbSys2
```

- e. Create the logical volume for the reserved partition.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L1G VGExaDb
```

 **Note:**

Do not create any file system on this logical volume.

- f. Create the logical volume for the guest storage repository.

```
lvm lvcreate -l 100%FREE -n LVDbExaVMImages VGExaDb
```

- g. Create a file system on the /dev/sda1 partition, and label it.

In the `mkfs.ext3` command below, the `-I 128` option is needed to set the inode size to 128.

```
mkfs.ext3 -I 128 /dev/sda1
tune2fs -c 0 -i 0 /dev/sda1
e2label /dev/sda1 BOOT
```

9. Create mount points for all the partitions, and mount the respective partitions.

For example, if `/mnt` is used as the top-level directory, the mounted list of partitions might look like:

- `/dev/VGExaDb/LVDbSys3 on /mnt`
- `/dev/sda1 on /mnt/boot`

The following example mounts the root file system, and creates two mount points:

```
mount /dev/VGExaDb/LVDbSys3 /mnt -t ext4
mkdir /mnt/boot
mount /dev/sda1 /mnt/boot -t ext3
```

- 10.** Bring up the network on `eth0` and assign the host's IP address and netmask to it.

```
ifconfig eth0 ip_address_for_eth0 netmask netmask_for_eth0 up
route add -net 0.0.0.0 netmask 0.0.0.0 gw gateway_ip_address
```

- 11.** Mount the NFS server holding the backups.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

- 12.** From the backup which was created in [Backing up the Management Domain dom0 Using Snapshot-Based Backup](#), restore the root directory (`/`) and the boot file system.

```
tar -pxvf /root/mnt/backup-of-root-and-boot.tar -C /mnt
```

- 13.** Unmount the restored `/dev/sda1` partition, and remount it on `/boot`.

```
umount /mnt/boot
mkdir -p /boot
mount /dev/sda1 /boot -t ext3
```

- 14.** Set up the `grub` boot loader using the command below:

```
grub --device-map=/boot/grub/device.map << DOM0_GRUB_INSTALL
root (hd0,0)
setup (hd0)
quit
DOM0_GRUB_INSTALL
```

- 15.** Unmount the `/boot` partition.

```
umount /boot
```

- 16.** Detach the `diagnostics.iso` file.

Using the ILOM Web interface, navigate to the Storage Devices dialog and click **Disconnect**.

The Storage Devices dialog is the interface that you earlier used to attach the `diagnostics.iso` image. See [Booting a Server using the Diagnostic ISO File in Oracle Exadata System Software User's Guide](#).

17. Check the restored `/etc/fstab` file and comment out any reference to `/EXAVMIMAGES`.

```
cd /mnt/etc
```

Comment out any line that references `/EXAVMIMAGES`.

18. Restart the system.

```
shutdown -r now
```

This completes the restoration procedure for the management domain (dom0).

19. Convert to Eighth Rack, if required.

If the recovery is on an Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#)".

20. When the server comes back up, build an OCFS2 file system on the `LVDbExaVMImages` logical volume.

```
mkfs -t ocfs2 -L ocfs2 -T vmstore --fs-features=local /dev/VGExaDb/
LVDbExaVMImages --force
```

21. Mount the OCFS2 partition on `/EXAVMIMAGES`.

```
mount -t ocfs2 /dev/VGExaDb/LVDbExaVMImages /EXAVMIMAGES
```

22. In `/etc/fstab`, uncomment the references to `/EXAVMIMAGES` and `/dev/mapper/VGExaDb-LVDbExaVMImages`, which you commented out earlier.

23. Mount the backup NFS server that holds the storage repository (`/EXAVMIMAGES`) backup to restore the `/EXAVMIMAGES` file system.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

24. Restore the `/EXAVMIMAGES` file system.

To restore all user domains, use this command:

```
tar -Spxvf /root/mnt/backup-of-exavmimages.tar -C /EXAVMIMAGES
```

To restore a single user domain from the backup, use the following command instead:

```
tar -Spxvf /root/mnt/backup-of-exavmimages.tar -C /EXAVMIMAGES
EXAVMIMAGES/<user-domain-name-to-be-restored>
```

25. Bring up each user domain.

```
xm create /EXAVMIMAGES/GuestImages/user_domain_hostname/vm.cfg
```

At this point all the user domains should come up along with Oracle Grid Infrastructure and the Oracle Database instances. The database instances should join the Oracle Real

Application Clusters (Oracle RAC) cluster formed by the other surviving management domain nodes.

### 5.20.2.3 Recovering a Management Domain and Its User Domains (Release 18.1 and X7 and Later)

You can recover a management domain from a snapshot-based backup when severe disaster conditions damage the management domain, or when the server hardware is replaced to such an extent that it amounts to new hardware.

1. Prepare an NFS server to host the backup archive `mybackup.tar.bz2`.

The NFS server must be accessible by IP address. For example, on an NFS server with the IP address `nfs_ip`, where the directory `/export` is exported from NFS mounts, put the `mybackup.tar.bz2` file in the `/export` directory

2. Restart the recovery target system using the `diagnostics.iso` file.

See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.

3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. If required, use `/opt/MegaRaid/storcli/storcli64`(or `/opt/MegaRAID/MegaCli/MegaCli64` for releases earlier than Oracle Exadata System Software 19c) to configure the disk controller to set up the disks.
5. Remove the logical volumes, the volume group, and the physical volume, in case they still exist after the disaster.

```
lvm vgremove VGExaDb --force
lvm pvremove /dev/sda3 --force
```

6. Remove the existing partitions, then verify all partitions were removed.

```
parted
GNU Parted 2.1
Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) print
Model: AVAGO MR9361-16i (scsi)
Disk /dev/sda: 4193GB
```



```
Sector size (logical/physical): 512B/512B
Partition Table: gpt
```

Number	Start	End	Size	File system	Name	Flags
1	32.8kB	537MB	537MB	ext4	primary	boot
2	537MB	805MB	268MB	fat32	primary	boot
3	805MB	4193GB	4192GB		primary	lvm

```
(parted) rm 1
[1730.498593] sda: sda2 sda3
(parted) rm 2
[1736.203794] sda: sda3
```

```
(parted) rm 3
[1738.546845] sda:
(parted) print
Model: AVAGO MR9361-16i (scsi)
Disk /dev/sda: 4193GB
Sector size (logical/physical): 512B/512B
Partition Table: gpt
```

Number	Start	End	Size	File system	Name	Flags
--------	-------	-----	------	-------------	------	-------

```
(parted) q
Information: You may need to update /etc/fstab.
```

## 7. Create the three partitions on /dev/sda.

- a. Get the end sector for the disk /dev/sda from a surviving management domain (dom0) and store it in a variable:

```
end_sector_logical=$(parted -s /dev/sda unit s print|perl -ne '/
^Disk\s+\S+:\s+(\d+)s/ and print $1')
end_sector=$(expr $end_sector_logical - 34)
echo $end_sector
```

The values for the start and end sectors in the commands below were taken from a surviving management domain. Because these values can change over time, it is recommended that these values are checked from a surviving dom0. For example, for an Exadata Database Machine X7-2 database server with 8 hard disk drives, you might see the following:

```
parted -s /dev/sda unit s print
Model: AVAGO MR9361-16i (scsi)
Disk /dev/sda: 8189440000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
```

Number	Start	End	Size	File system	Name	Flags
1	64s	1048639s	1048576s	ext4	primary	boot
2	1048640s	1572927s	524288s	fat32	primary	boot
3	1572928s	8189439966s	8187867039s		primary	lvm

 **Note:**

The `s` (sector) value for the following sub-steps are based on a system with 8 hard disk drives. If you have 4 hard disk drives, then you need to view the partition table from the management domain on a surviving node and adjust the sector values accordingly.

- b.** Create the boot partition, `/dev/sda1`.

```
parted -s /dev/sda mklabel gpt mkpart primary 64s 1048639s set
1 boot on
```

- c.** Create the partition that will hold the LVMs, `/dev/sda2`.

```
parted -s /dev/sda mkpart primary fat32 1048640s 1572927s set
2 boot on
```

- d.** Create the partition that will hold the LVMs, `/dev/sda3`.

```
parted -s /dev/sda mkpart primary 1572928s 8189439966s set 3
lvm on
```

- 8.** Use the `/sbin/lvm` command to re-create the logical volumes and `mkfs` to create the file systems.

- a.** Create the physical volume and the volume group.

```
lvm pvcreate /dev/sda3
lvm vgcreate VGExaDb /dev/sda3
```

- b.** Create the logical volume for the file system that will contain the `/` (root) directory and label it.

```
lvm lvcreate -n LVDbSys3 -L30G VGExaDb
mkfs -t ext4 /dev/VGExaDb/LVDbSys3
e2label /dev/VGExaDb/LVDbSys3 DBSYSOVS
```

- c.** Create the logical volume for the swap directory, and label it.

```
lvm lvcreate -n LVDbSwap1 -L24G VGExaDb
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

- d.** Create the logical volume for the backup partition, and build a file system on top of it.

```
lvm lvcreate -n LVDbSys2 -L30G VGExaDb
mkfs -t ext4 /dev/VGExaDb/LVDbSys2
```

- e.** Create the logical volume for the reserved partition, which is needed for creating snapshots.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L 1G VGExaDb
```

 **Note:**

Do not create any file system on this logical volume.

- f. Create the logical volume for the guest storage repository.

```
lvm lvcreate -l 100%FREE -n LVDbExaVMImages VGExaDb
```

- g. Create a file system on the `/dev/sda1` partition, and label it.

```
mkfs.ext4 /dev/sda1
e2label /dev/sda1 BOOT
tune2fs -l /dev/sda1
```

- h. Create a file system on the `/dev/sda2` partition, and label it.

```
mkfs.vfat -v -c -F 32 -s 2 /dev/sda2
dosfslabel /dev/sda2 ESP
```

9. Create mount points for all the partitions, and mount the respective partitions.

For example, if `/mnt` is used as the top-level directory, the mounted list of partitions might look like:

- `/dev/VGExaDb/LVDbSys3` on `/mnt`
- `/dev/sda1` on `/mnt/boot`
- `/dev/sda2` on `/mnt/boot/efi`

The following example mounts the root (`/`) file system, and creates three mount points:

```
mount /dev/VGExaDb/LVDbSys3 /mnt -t ext4
mkdir /mnt/boot
mount /dev/sda1 /mnt/boot -t ext4
mkdir /mnt/boot/efi
mount /dev/sda2 /mnt/boot/efi -t vfat
```

10. Bring up the network on `eth0` and (if not using DHCP) assign the host's IP address and netmask to it.

If you are using DHCP then you do not have to manually configure the IP address for the host.

```
ip address add ip_address_for_eth0/netmask_for_eth0 dev eth0
ip link set up eth0
ip route add default via gateway_ip_address dev eth0
```

11. Mount the NFS server holding the backups.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/location_of_backup /root/mnt
```

12. From the backup which was created in [Backing up the Management Domain dom0 Using Snapshot-Based Backup](#), restore the root directory (/) and the boot file system.

```
tar -pjxvf /root/mnt/backup-of-root-and-boot.tar -C /mnt
```

13. Use the `efibootmgr` command to set the boot device.

- a. Disable and delete the Oracle Linux boot device. If you see the entry `ExadataLinux_1`, then remove this entry and recreate it.

For example:

```
efibootmgr
BootCurrent: 000F
Timeout: 1 seconds
BootOrder:
000F,0000,0001,0002,0003,0004,0005,0007,0008,0009,000A,000D,000E
Boot0000* ExadataLinux_1
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet
Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet
Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000D* Oracle Linux
Boot000E* UEFI OS
Boot000F* USB:SUN
```

In this example, you would disable and remove Oracle Linux (`Boot00D`) and `ExadataLinux_1` (`Boot000`). Use commands similar to the following to disable and delete the boot devices:

Disable 'Oracle Linux':

```
efibootmgr -b 000D -A
```

Delete 'Oracle Linux':

```
efibootmgr -b 000D -B
```

Disable old 'ExadataLinux\_1':

```
efibootmgr -b 0000 -A
```

Delete old 'ExadataLinux\_1':

```
efibootmgr -b 0000 -B
```

- b. Recreate the boot entry for `ExadataLinux_1` and then view the boot order entries.

```
efibootmgr -c -d /dev/sda -p 2 -l '\EFI\XEN\XEN.EFI' -L
'ExadataLinux_1'
```

```
efibootmgr
BootCurrent: 000F
```

```
Timeout: 1 seconds
BootOrder: 0000,0001,0002,0003,0004,0005,0007,0008,0009,000A,000E,000F
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000E* UEFI OS
Boot000F* USB:SUN
Boot0000* ExadataLinux_1
```

In the output from the `efibootmgr` command, make note of the boot order number for `ExadataLinux_1` and use that value in the following commands:

```
efibootmgr -b (entry number) -A
efibootmgr -b (entry number) -a
```

For example, in the previous output shown in step 13.a, `ExadataLinux_1` was listed as `(Boot000)`. So you would use the following commands:

```
efibootmgr -b 0000 -A
efibootmgr -b 0000 -a
```

**c.** Set the correct boot order.

Set `ExadataLinux_1` as the first boot device. The remaining devices should stay in the same boot order, except for `USB:SUN`, which should be last.

```
efibootmgr -o
0000,0001,0002,0003,0004,0005,0007,0008,0009,000A,000E,000F
```

The boot order should now look like the following:

```
efibootmgr
BootCurrent: 000F
Timeout: 1 seconds
BootOrder: 0000,0001,0002,0003,0004,0005,0007,0008,0009,000A,000E,000F
Boot0000* ExadataLinux_1
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0002* NET1:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0003* NET2:PXE IP4 Oracle Dual Port 10GBase-T Ethernet Controller
Boot0004* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle Dual Port 25Gb Ethernet Adapter
Boot0007* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
Boot000A* PCIE3:PXE IP4 Oracle Quad Port 10GBase-T Adapter
```

```
Boot000E* UEFI OS
Boot000F* USB:SUN
```

- d. Check the boot order using the `ubiosconfig` command.

```
ubiosconfig export all -x /tmp/ubiosconfig.xml
```

Make sure the `ExadataLinux_1` entry is the first child element of `boot_order`.

```
<boot_order>
 <boot_device>
 <description>ExadataLinux_1</description>
 <instance>1</instance>
 </boot_device>
 <boot_device>
 <description>NET0:PXE IP4 Intel(R) I210 Gigabit Network
Connection</description>
 <instance>1</instance>
 </boot_device>
 ...
```

14. Check the restored `/etc/fstab` file and comment out any reference to `/EXAVMIMAGES`.

```
cd /mnt/etc
```

Comment out any line that references `/EXAVMIMAGES`.

15. Detach the `diagnostics.iso` file.

Using the ILOM Web interface, navigate to the Storage Devices dialog and click **Disconnect**.

The Storage Devices dialog is the interface that you earlier used to attach the `diagnostics.iso` image. See *Booting a Server using the Diagnostic ISO File in Oracle Exadata System Software User's Guide*.

16. Unmount the restored `/dev/sda1` partitions so `/dev/sda1` can be remounted on `/boot`.

```
umount /mnt/boot/efi
umount /mnt/boot
umount /mnt
umount /root/mnt
```

17. Restart the system.

```
shutdown -r now
```

This completes the restoration procedure for the management domain (dom0).

18. Convert to Eighth Rack, if required.

If the recovery is on an Exadata Database Machine Eighth Rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

19. When the server comes back up, build an OCFS2 file system on the `LVDbExaVMImages` logical volume.

```
mkfs -t ocfs2 -L ocfs2 -T vmstore --fs-features=local /dev/VGExaDb/
LVDbExaVMImages --force
```

20. Mount the OCFS2 partition on `/EXAVMIMAGES`.

```
mount -t ocfs2 /dev/VGExaDb/LVDbExaVMImages /EXAVMIMAGES
```

21. In `/etc/fstab`, uncomment the references to `/EXAVMIMAGES` and `/dev/mapper/VGExaDb-LVDbExaVMImages`, which you commented out earlier.

22. Mount the backup NFS server that holds the storage repository (`/EXAVMIMAGES`) backup to restore the `/EXAVMIMAGES` file system.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

23. Restore the `/EXAVMIMAGES` file system.

To restore all user domains, use this command:

```
tar -Spxvf /root/mnt/backup-of-exavmimages.tar -C /EXAVMIMAGES
```

To restore a single user domain from the backup, use the following command instead:

```
tar -Spxvf /root/mnt/backup-of-exavmimages.tar -C /EXAVMIMAGES
EXAVMIMAGES/<user-domain-name-to-be-restored>
```

24. Bring up each user domain.

```
xm create /EXAVMIMAGES/GuestImages/user_domain_hostname/vm.cfg
```

At this point all the user domains should come up along with Oracle Grid Infrastructure and the Oracle Database instances. The database instances should join the Oracle RAC cluster formed by the other surviving management domain nodes.

## 5.20.3 Scenario 2: Re-imaging the Management Domain and Restoring User Domains from Backups

This procedure re-images the management domain and reconstructs all the user domains.

The following procedure can be used when the management domain is damaged beyond repair and no backup exists for the management domain, but there is a backup available of the storage repository (`/EXAVMIMAGES` file system) housing all the user domains.

1. Re-image the management domain with the image used in the other management domains in the rack using the procedure described in [Re-Imaging the Oracle Exadata Database Server](#).

2. Run the following commands:

```
/opt/oracle.SupportTools/switch_to_ovm.sh

/opt/oracle.SupportTools/reclaimdisks.sh -free -reclaim
```

3. If the recovery is on Oracle Exadata Database Machine eighth rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

4. Rebuild the OCFS2 file system on the `/dev/sda3` partition.

```
umount /EXAVMIMAGES

mkfs -t ocfs2 -L ocfs2 -T vmstore --fs-features=local /dev/sda3 --force
```

5. Mount the OCFS2 partition `/dev/sda3` on `/EXAVMIMAGES`.

```
mount -t ocfs2 /dev/sda3 /EXAVMIMAGES
```

6. Mount the backup NFS server to restore the `/EXAVMIMAGES` file system which holds the user domain images.

```
mkdir -p /remote_FS

mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/location_of_backup /remote_FS
```

7. Restore the `/EXAVMIMAGES` file system.

```
tar -Spxvf /remote_FS/backup-of-exavmimages.tar -C /EXAVMIMAGES
```

 **Note:**

The restore process of storage repository restores the user domain specific files (files under `/EXAVMIMAGES/GuestImages/user_domain/`) as regular files and not as OCFS2 reflinks, which is what these files in the storage repository were originally at the time of the user domain creation. Consequently, the space usage in `/EXAVMIMAGES` may go up after the restoration process when compared to the original space usage at the time of the backup.

8. Manually configure the network bridges.
  - a. Determine the version of the `ovmutils` RPM.

```
rpm -qa|grep ovmutils
```



- b. If the version of the `ovmutils` RPM is earlier than 12.1.2.2.0, perform these steps:
- i. Back up `/opt/exadata_ovm/exadata.img.domu_maker`. You will need the backup copy later.

```
cp /opt/exadata_ovm/exadata.img.domu_maker /opt/exadata_ovm/exadata.img.domu_maker-orig
```

- ii. Open the `/opt/exadata_ovm/exadata.img.domu_maker` file in a text editor such as `vi`, and search for `g_do_not_set_bridge=yes`. This string should be located a few lines below the case statement option `network-discovery`).

Change the string to `g_do_not_set_bridge=no`.

Save and exit `/opt/exadata_ovm/exadata.img.domu_maker`.

- iii. Run `/opt/exadata_ovm/exadata.img.domu_maker` manually for every XML file in the `/EXAVMIMAGES/conf` directory.

```
cd /EXAVMIMAGES/conf
ls -l|while read file; do /opt/exadata_ovm/exadata.img.domu_maker network-discovery $file /tmp/netdisc-$file; done
```

- iv. Restore `/opt/exadata_ovm/exadata.img.domu_maker` from the backup copy.

```
cp /opt/exadata_ovm/exadata.img.domu_maker-orig /opt/exadata_ovm/exadata.img.domu_maker
```

- c. If the version of the `ovmutils` RPM is 12.1.2.2.0 or later, then run the following command:

```
/opt/exadata_ovm/exadata.img.domu_maker add-bonded-bridge-dom0 vmbondeth0 eth4 eth5
```

9. For each user domain directory in the `/EXAVMIMAGES/GuestImages` directory, perform the following steps:

- a. Get the UUID of the user domain.

```
grep ^uuid /EXAVMIMAGES/GuestImages/user_domain_hostname/vm.cfg |awk -F"=" '{print $2}'|sed s/"//g|sed s/" //g
```

The command returns the `uuid` value, which is used in the commands below.

- b. Create a sub-directory for the UUID.

```
mkdir -p /OVS/Repositories/uuid
```

- c. Create a symbolic link for the `vm.cfg` file for the `user_domain_hostname` in the new UUID directory.

```
ln -s /EXAVMIMAGES/GuestImages/user_domain_hostname/vm.cfg /OVS/Repositories/uuid/vm.cfg
```

- d. Configure autostart for the `user_domain_hostname`.

```
ln -s /OVS/Repositories/uuid/vm.cfg /etc/xen/auto/
user_domain_hostname.cfg
```

- e. Create the `VirtualDisks` sub-directory.

```
mkdir VirtualDisks
```

- f. Enter the `VirtualDisks` directory.

```
cd VirtualDisks
```

- g. Create four symbolic links in this directory using the four disk image names in the `vm.cfg` file, pointing to the four `*.img` files in `/EXAVMIMAGES/GuestImages/user_domain_hostname` directory.

For example, the following is a sample disk entry in a sample `vm.cfg` file in a `/OVS/Repositories/uuid` directory:

```
disk = ['file:/OVS/Repositories/6e7c7109c1bc4ebba279f8
4e595e0b27/VirtualDisks/dfd641a1c6a84bd69643da704ff98594.img,xv
da,w','file:/OVS/Repositories/6e7c7109c1bc4ebba279f84e595e0b27/
VirtualDisks/d349fd420a1e49459118e6a6fcdbc2a4.img,xvdb,w','file
:/OVS/Repositories/6e7c7109c1bc4ebba279f84e595e0b27/VirtualDisk
s/8ac470eeb8704aab9a8b3adedf1c3b04.img,xvdc,w','file:/OVS/Repos
itories/6e7c7109c1bc4ebba279f84e595e0b27/VirtualDisks/333e7ed28
50a441ca4d2461044dd0f7c.img,xvdd,w']
```

You can list the four `*.img` files in the `/EXAVMIMAGES/GuestImages/user_domain_hostname` directory:

```
ls /EXAVMIMAGES/GuestImages/user_domain_name/*.img
/EXAVMIMAGES/GuestImages/user_domain_name/System.img
/EXAVMIMAGES/GuestImages/user_domain_name/grid12.1.0.2.2.img
/EXAVMIMAGES/GuestImages/user_domain_name/db12.1.0.2.2-3.img
/EXAVMIMAGES/GuestImages/user_domain_name/pv1_vgexadb.img
```

In this example, the following commands can be used to create the four symbolic links where `dbm01db08vm01` is the user domain host name:

```
ln -s /EXAVMIMAGES/GuestImages/dbm01db08vm01/System.img
$(grep ^disk /EXAVMIMAGES/GuestImages/dbm01db08vm01/vm.cfg|awk
-F":" '{print $2}'|awk -F"," '{print $1}'|awk -F"/" '{print $6}')
```

```
ln -s /EXAVMIMAGES/GuestImages/dbm01db08vm01/grid12.1.0.2.2.img
$(grep ^disk /EXAVMIMAGES/GuestImages/dbm01db08vm01/vm.cfg|awk
-F":" '{print $3}'|awk -F"," '{print $1}'|awk -F"/" '{print $6}')
```

```
ln -s /EXAVMIMAGES/GuestImages/dbm01db08vm01/db12.1.0.2.2-3.img
$(grep ^disk /EXAVMIMAGES/GuestImages/dbm01db08vm01/vm.cfg|awk
-F":" '{print $4}'|awk -F"," '{print $1}'|awk -F"/" '{print $6}')
```

```
ln -s /EXAVMIMAGES/GuestImages/dbm01db08vm01/pv1_vgexadb.img
$(grep ^disk /EXAVMIMAGES/GuestImages/dbm01db08vm01/vm.cfg|awk
-F":" '{print $5}'|awk -F"," '{print $1}'|awk -F"/" '{print $6}')
```

#### 10. Restart each user domain.

```
xm create /EXAVMIMAGES/GuestImages/user_domain_hostname/vm.cfg
```

At this point all the user domains should start along with the Oracle Grid Infrastructure and the database instances. The node should join the Oracle RAC cluster formed by the other surviving management domain nodes.

## 5.20.4 Scenario 3: Restoring and Recovering User Domains from Snapshot Backups

Use this procedure to restore lost or damaged files of a user domain using a snapshot-based user domain backup taken from inside a user domain.

To use this procedure, the user domain backup must have been created using the procedure described in [Method 3: Back up a User Domain from Inside the User Domain](#).

1. Log in to the user domain as the root user.
2. Mount the backup NFS server to restore the damaged or lost files.

```
mkdir -p /root/mnt

mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

3. Extract the damaged or lost files from the backup to a staging area.

Prepare a staging area to hold the extracted files. The backup LVM `LVDbSys2` can be used for this:

```
mkdir /backup-LVM

mount /dev/mapper/VGExaDb-LVDbSys2 /backup-LVM

mkdir /backup-LVM/tmp_restore

tar -pjxvf /root/mnt/tar_file_name -C /backup-LVM/tmp_restore
absolute_path_of_file_to_be_restored
```

4. Restore the damaged or lost files from the temporary staging area as needed.
5. Restart the user domain.

## 5.21 Removing an Oracle RAC Cluster Running in Oracle VM

You can remove all Oracle RAC nodes of an Oracle VM cluster, including the databases running within the cluster and all data stored on the Oracle Exadata Storage Server used by those databases.

To remove only a subset of user domains of an Oracle VM cluster, refer to the next section.

There are two main steps to remove an Oracle VM cluster:

- Remove the user domain files from the management domain.
- Remove the unused Oracle Exadata grid disks.

 **Note:**

If the Oracle Exadata Deployment Assistant `xml` configuration files are to be reused later, then they will be not synchronized because the definition for the removed user domain still exists in Oracle Exadata Deployment Assistant files.

1. Run the following example script as the grid software owner on any user domain to be removed.

The example shell script generates two scripts, `list_griddisk.sh` and `drop_griddisk.sh`, that are run later in this procedure. Do not run the generated scripts until instructed.

```
#!/bin/bash

Run this script as the Grid Infrastructure software owner.
#
This script identifies griddisks used by this cluster and the
cells to
which they belong, then creates two shell scripts - the list
script to
show the current status, and the drop script to drop the
griddisks.
#
In order for the drop script to succeed, the griddisks must not
be in use,
meaning databases and CRS are down, and the list script returns
no output.
#
The generated scripts are designed to run via dcli -x

ORACLE_SID=$(awk -F: '/^+ASM/{print $1}' /etc/oratab)
ORAENV_ASK=NO . oraenv >/dev/null

listGriddiskScript=list_griddisk.sh
dropGriddiskScript=drop_griddisk.sh

rm -f $listGriddiskScript $dropGriddiskScript

gridDiskList=$(asmcmd lsdsk --suppressheader | awk -F '/'
'{print $NF}')
if [[${PIPESTATUS[0]} != 0]]; then echo "asmcmd failed -
exiting"; exit 1; fi

cellList=$(echo "$gridDiskList" | awk -F_ '{print $NF}' | sort -u)
```

```

for cell in $cellList; do
 myGriddisks=$(echo "$gridDiskList" | grep ${cell}$ | tr '\n' ',')
 echo "[[\$(hostname -s) == ${cell}]] && cellcli -e 'LIST GRIDDISK \
 ${myGriddisks%},} attributes name, asmDiskGroupName, asmModeStatus
 \
 where asmModeStatus != UNKNOWN'" >> $listGriddiskScript
 echo >> $listGriddiskScript
done

chmod +x $listGriddiskScript

echo
echo "Run the following command to list griddisks in use by this cluster:"
echo
echo "# dcli -l celladmin -c ${cellList//'\n'/,} -x $listGriddiskScript"
echo

for cell in $cellList; do
 myGriddisks=$(echo "$gridDiskList" | grep ${cell}$ | tr '\n' ',')
 echo "[[\$(hostname -s) == ${cell}]] && cellcli -e 'DROP GRIDDISK \
 ${myGriddisks%},}'" >> $dropGriddiskScript
 echo >> $dropGriddiskScript
done

chmod +x $dropGriddiskScript

echo
echo "Stop CRS on all nodes in this cluster, then run the following"
echo "command to drop all griddisks used by this cluster:"
echo
echo "# dcli -l celladmin -c ${cellList//'\n'/,} -x $dropGriddiskScript"
echo

exit

```

2. Shut down the databases and Oracle Grid Infrastructure in all user domains that will be removed:

```
Grid_home/bin/crsctl stop crs -f
```

3. Run the `list_griddisk.sh` script generated earlier from any user domain that will be removed.

#### Note:

- Run the script using the `dcli` command to connect as the `celladmin` user to all Oracle Exadata Storage Servers in the configuration.
- Before running the `dcli` command, set up a passwordless SSH connection between the grid software owner on the database server and the `celladmin` user on the cells. Otherwise, the command will keep prompting you to enter the password.

The following is an example of the command:

```
$ dcli -l celladmin -c dm01celadm01,dm01celadm02,dm01celadm03 \
-x list_griddisk.sh
```

The `list_griddisk.sh` script should not output any grid disks. Grid disks returned from the `list_griddisk.sh` script are considered still in use.

Do not proceed until the `list_griddisk.sh` script returns empty output indicating no grid disks are in use. Verify that Oracle Grid Infrastructure and the databases are shut down on all user domains to be dropped.

4. Run the `drop_griddisk.sh` script generated earlier from any user domain that you want to remove.

Run the script using the `dcli` command to connect as the `celladmin` user to all Oracle Exadata Storage Servers in the configuration.

```
$ dcli -l celladmin -c dm01celadm01,dm01celadm02,dm01celadm03 \
-x drop_griddisk.sh
```

5. Run the `exadata.img.domu_maker` command from the management domain of each user domain you want to remove.

This command removes the user domains, where *DomainName* is the name of the user domain.

```
/opt/exadata_ovm/exadata.img.domu_maker remove-domain DomainName
```

In the following example, the commands remove the two user domains for a two-node Oracle VM RAC cluster in which the user domain `dm01db01vm04` runs on the management domain `dm01db01`, and the user domain `dm01db02vm04` runs on the management domain `dm01db02`.

```
[root@dm01db01 ~] # /opt/exadata_ovm/exadata.img.domu_maker \
remove-domain dm01db01vm04
[INFO] Start with command line: /opt/exadata_ovm/
exadata.img.domu_maker \
 remove-domain dm01db01vm04
[INFO] Shutting down DomU dm01db01vm04
[INFO] Autostart link for dm01db01vm04 deleted from /etc/xen/auto
[INFO] Deleted OVM repository /OVS/Repositories/
7bfd49d6bd5a4b2db2e46e8234788067 for DomU dm01db01vm04
[INFO] Deleted guest vm /EXAVMIMAGES/GuestImages/dm01db01vm04 for \
DomU dm01db01vm04
```

```
[root@dm01db02 ~]# /opt/exadata_ovm/exadata.img.domu_maker \
remove-domain dm01db02vm04
[INFO] Start with command line: /opt/exadata_ovm/
exadata.img.domu_maker \
remove-domain dm01db02vm04
[INFO] Shutting down DomU dm01db02vm04
[INFO] Autostart link for dm01db02vm04 deleted from /etc/xen/auto
[INFO] Deleted OVM repository /OVS/Repositories/
1d29719ff26a4a17aca99b2f89fd8032 for DomU dm01db02vm04
```

```
[INFO] Deleted guest vm /EXAVMIMAGES/GuestImages/dm01db02vm04 \
for DomU dm01db02vm04
```

## 5.22 Removing a User Domain

You can remove an user domain in Oracle VM using either OEDACLI or the `domu_maker` utility.

- [Removing a User Domain from an Oracle VM Cluster Using OEDACLI](#)  
You can use OEDACLI to remove a user domain from an Oracle VM cluster.
- [Removing a User Domain from an Oracle VM Cluster Using `domu\_maker`](#)  
You can remove a single Oracle RAC node from an Oracle VM cluster by using the `domu_maker` utility.

### 5.22.1 Removing a User Domain from an Oracle VM Cluster Using OEDACLI

You can use OEDACLI to remove a user domain from an Oracle VM cluster.

The following procedure removes a user domain from a cluster. If the user domain is not part of a cluster, then you can skip the cluster-related commands.

1. Load the XML configuration file (es. `.xml` for example) into OEDACLI.

```
./oedacli -c full_path_to_XML_file
```

2. Use the following OEDACLI commands to delete the database instance from the cluster node.

```
DELETE GUEST WHERE srcname=domain_FQDN stepname=ADD_INSTANCE
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

3. Use the following OEDACLI commands to delete the Oracle Database home from the cluster.

```
DELETE GUEST WHERE srcname=domain_FQDN stepname=EXTEND_DBHOME
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

4. Use the following OEDACLI commands to delete the user domain from the cluster.

```
DELETE GUEST WHERE srcname=domain_FQDN stepname=ADD_NODE
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

5. Use the following OEDACLI commands to remove connectivity to the storage servers and delete the users on the user domain.

```
DELETE GUEST WHERE srcname=domain_FQDN stepname=CELL_CONNECTIVITY
SAVE ACTION
DELETE GUEST WHERE srcname=domain_FQDN stepname=CREATE_USERS
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

6. Use the following OEDACLI commands to remove the user domain.

```
DELETE GUEST WHERE srcname=domain_FQDN stepname=CREATE_GUEST
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

7. Save the updated configuration information.

Specify the full path to a directory where the updated configuration information will be saved.

```
SAVE FILES LOCATION=output_directory
```

8. View the list of user domains on the management domain.

```
xm list
```

The user domain you just removed should not be listed.

## 5.22.2 Removing a User Domain from an Oracle VM Cluster Using `domu_maker`

You can remove a single Oracle RAC node from an Oracle VM cluster by using the `domu_maker` utility.

The Oracle Exadata grid disks remain in use by the remaining nodes in the cluster, and must not be dropped.

### Note:

If Oracle Exadata Deployment Assistant `xml` configuration files are to be reused later, then they will be not synchronized because the definition for the removed user domain still exists in Oracle Exadata Deployment Assistant files.

1. Delete the cluster node.



- Use the following command to shut down and remove the user domain, where *DomainName* is the name of the domain:

```
/opt/exadata_ovm/exadata.img.domu_maker remove-domain DomainName
```

This command removes the user domain files from the management domain.

### Related Topics

- Oracle Clusterware Administration and Deployment Guide*

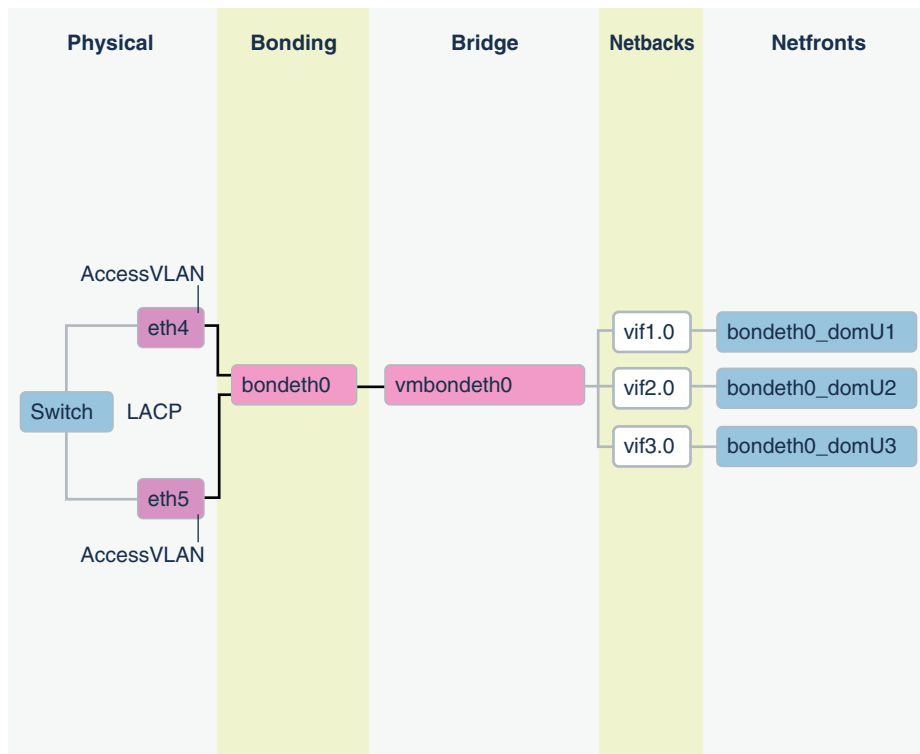
## 5.23 Implementing Tagged VLAN Interfaces

This topic describes the implementation of tagged VLAN interfaces in Oracle VM environments on Exadata.

Oracle databases running in Oracle VM guests on Oracle Exadata Database Machine are accessed through the client Ethernet network defined in the Oracle Exadata Deployment Assistant (OEDA) configuration tool. Client network configuration in both the management domain (dom0) and user domains (domU's) is done automatically when the OEDA installation tool creates the first user domain during initial deployment.

The following figure shows a default bonded client network configuration:

**Figure 5-1 NIC Layout in an Oracle Virtual Environment**



The network has the following configuration:

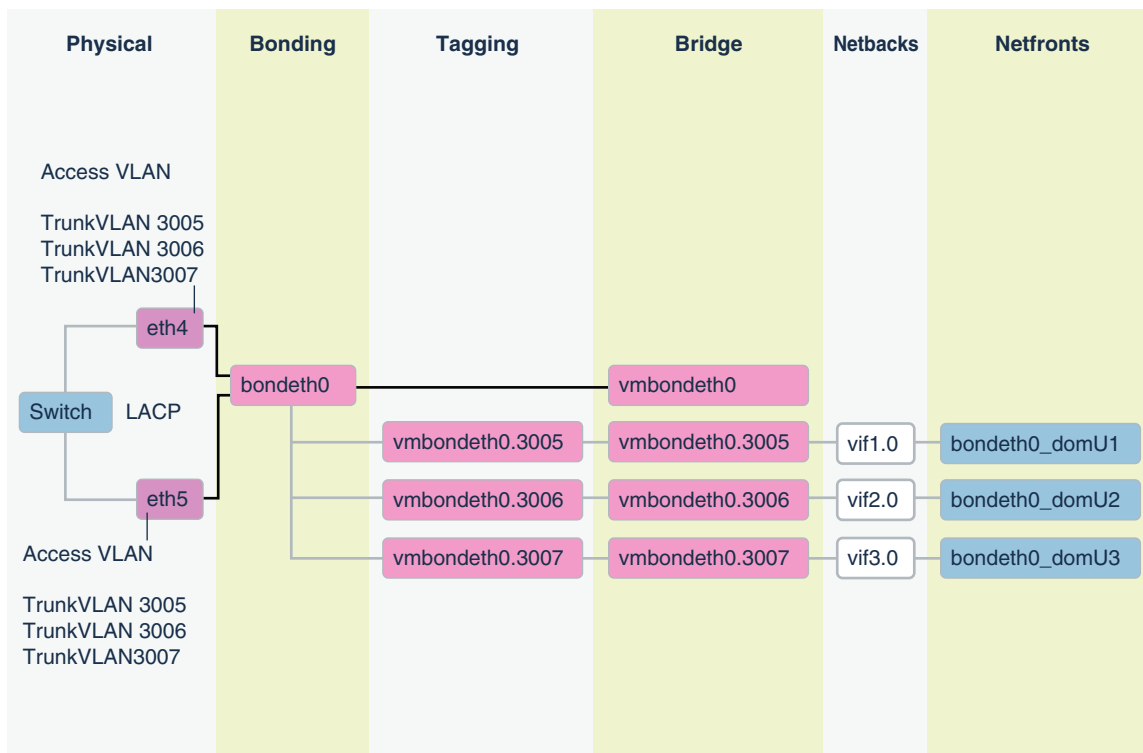
1. In the dom0, eth slave interfaces (for example, eth1 and eth2, or eth4 and eth5) that allow access to the domU client network defined in OEDA are discovered, configured, and brought up, but no IP is assigned.
2. In the dom0, bondeth0 master interface is configured and brought up, but no IP is assigned.
3. In the dom0, bridge interface vmbondeth0 is configured, but no IP is assigned.
4. In the dom0, one virtual backend interface (vif) per domU that maps to that particular domU's bondeth0 interface is configured and brought up, but no IP is assigned. These vifs are configured on top of the bridge interface vmbondeth0, and the mapping between the dom0 vif interface and its corresponding user domain interface bondeth0 is defined in the user domain configuration file called `vm.cfg`, located in `/EXAVMIMAGES/GuestImages/user domain name`.

For default installations, a single bondeth0 and a corresponding vmbondeth0 bridge interface is configured in the dom0 as described above. This bondeth0 interface is based on the default Access Virtual Local Area Network (Access VLAN). The ports on the switch used by the slave interfaces making up bondeth0 are configured for Access VLAN.

### Using VLAN Tagging

If there is a need for virtual deployments on Exadata to access additional VLANs on the client network, such as enabling network isolation across user domains, then 802.1Q-based VLAN tagging is a solution. The following figure shows a client network configuration with VLAN tagging.

**Figure 5-2 NIC Layout for Oracle Virtual Environments with VLAN Tagging**



For instructions on how to configure and use such additional VLAN tagged interfaces on the client network, see My Oracle Support note 2018550.1. The Access VLAN must stay working and configured before and after these instructions are followed. At no time is the Access VLAN to be disabled.

**Related Topics**

- [Implementing Tagged VLAN Interfaces in Oracle VM Environments on Exadata \(My Oracle Support Doc ID 2018550.1\)](#)

## 5.24 Implementing InfiniBand Partitioning across Oracle VM Oracle RAC Clusters on Exadata Database Machine

For Oracle Real Application Clusters (Oracle RAC) clusters running in Oracle VM on Exadata Database Machine, you can isolate the network traffic on the InfiniBand network for each Oracle RAC clusters using custom InfiniBand partitioning, dedicated partition keys, and partitioned tables.

- [About InfiniBand Partitioning Across Oracle RAC Clusters Running in Oracle VM](#)  
An InfiniBand partition defines a group of InfiniBand nodes or members that are allowed to communicate with one another.
- [Requirements for Implementing InfiniBand Partitioning across OVM RAC Clusters](#)
- [About InfiniBand Partitioning Network Configuration](#)  
Plan and allocate sets of IP addresses and netmasks for each Oracle VM RAC cluster that will be used by the cluster pkey interfaces and the storage pkey interfaces when InfiniBand partitioning gets implemented in the cluster.
- [Configuring InfiniBand Partitioning across Oracle VM RAC Clusters](#)  
The steps for configuring InfiniBand Partitioning across Oracle RAC clusters running in Oracle VM are described here.
- [Implementing InfiniBand Partitioning across OVM RAC Clusters: Setting up Limited Membership](#)

**Related Topics**

- [Implementing InfiniBand Partitioning across OVM RAC clusters on Exadata \(Doc ID 2075398.1\)](#)
- [Implementing Tagged VLAN Interfaces in Oracle VM Environments on Exadata \(Doc ID 2018550.1\)](#)

### 5.24.1 About InfiniBand Partitioning Across Oracle RAC Clusters Running in Oracle VM

An InfiniBand partition defines a group of InfiniBand nodes or members that are allowed to communicate with one another.

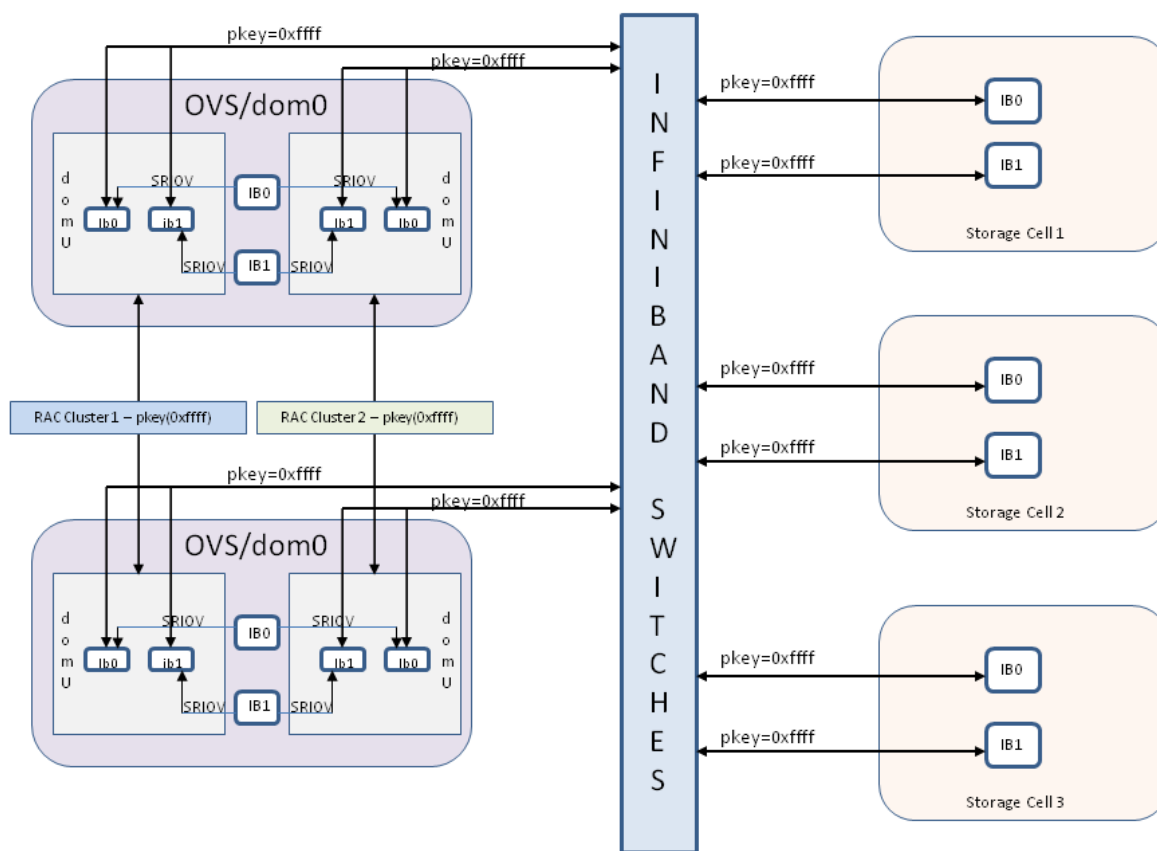
One of the key requirements of consolidated systems from a security standpoint is network isolation across the multiple environments within a consolidated system. For consolidations achieved using Oracle VM Oracle Real Application Clusters (Oracle RAC) clusters on Oracle Exadata, this means isolation across the different Oracle RAC clusters such that network traffic of one Oracle RAC cluster is not accessible to another Oracle RAC cluster. For the Ethernet networks, this is accomplished using VLAN tagging as described in My Oracle

Support document 2018550.1. For the InfiniBand network, this is accomplished using custom InfiniBand partitioning, dedicated partition keys, and partitioned tables.

With InfiniBand partitioning, partitions identified by unique partition keys are created and are managed by the master subnet manager. Members are then assigned to these custom partitions. Members within a partition can only communicate among themselves (depending on the membership as explained in the Appendix 1 of My Oracle Support document 2018550.1). A member of one partition cannot communicate with a member of a different partition regardless of the membership. Continuing along these lines, the Oracle VM Oracle RAC nodes of one particular cluster are assigned one dedicated partition for the clusterware communication and one partition for communication with the storage cells. This way, the nodes of one Oracle RAC cluster will not be able to communicate with the nodes of another Oracle RAC cluster that belong to a different partition. The nodes in each Oracle RAC cluster have different partition keys assigned to them.

By default, the InfiniBand subnet manager provides a single partition that is identified by the partition key 0x7FFF (limited membership) or 0xFFFF (full membership). In Oracle VM deployments on Exadata Database Machine where custom InfiniBand partitioning is not used, the partition key 0xFFFF is used across all the user domains.

**Figure 5-3 Oracle VM Oracle RAC Clusters without InfiniBand Network Isolation Across Clusters**

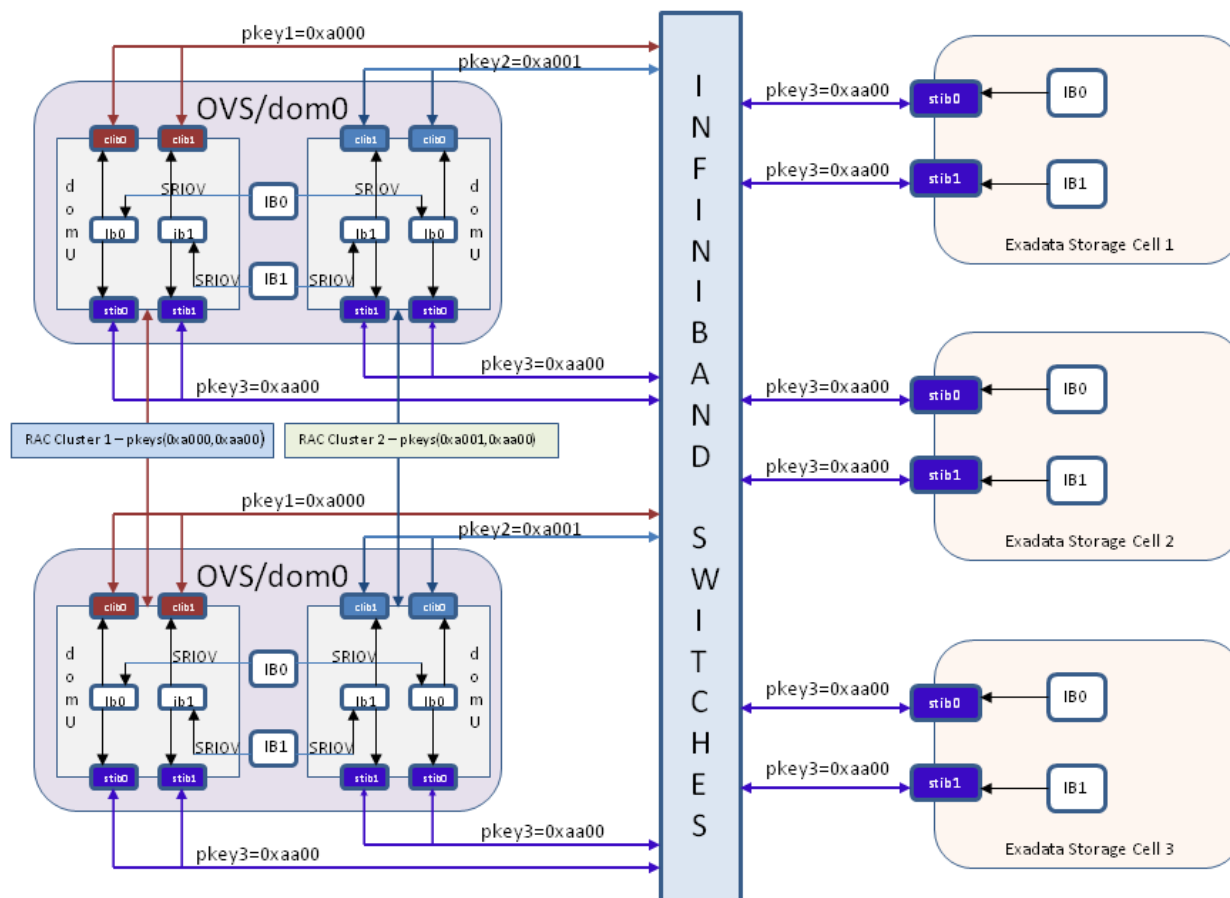


With non-default custom partitions in place for implementing isolation across the Oracle VM Oracle RAC clusters, the configuration changes to what is shown in the

next image. New interfaces `clib0`, `clib1` (for the cluster pkey) and `stib0`, `stib1` (for the storage pkey) exist in each of the user domains (domU's).

There is no change to InfiniBand interfaces in the management domain (dom0).

**Figure 5-4 Oracle VM Oracle RAC Clusters with InfiniBand Network Isolation Across Clusters Using InfiniBand Partitioning**



### Related Topics

- [Implementing InfiniBand Partitioning across OVM RAC clusters on Exadata \(My Oracle Support document 2075398.1\)](#)
- [Implementing Tagged VLAN Interfaces in Oracle VM Environments on Exadata \(My Oracle Support document 2018550.1\)](#)

## 5.24.2 Requirements for Implementing InfiniBand Partitioning across OVM RAC Clusters

Before configuring InfiniBand partitioning, ensure that:

- You have configured OVM on your Exadata system.

- All the user domains and storage cells are using the default partition key 0xFFFF.
- You have set up passwordless secure shell (ssh) access for the root user from one of the management domains (dom0 node) to all the OVM RAC cluster nodes, storage cells, and InfiniBand switches.
- InfiniBand switches are installed with firmware versions 2.0.4 or above.
- You have an understanding of InfiniBand partitioning.

### 5.24.3 About InfiniBand Partitioning Network Configuration

Plan and allocate sets of IP addresses and netmasks for each Oracle VM RAC cluster that will be used by the cluster pkey interfaces and the storage pkey interfaces when InfiniBand partitioning gets implemented in the cluster.

Within an Oracle VM RAC cluster, the cluster pkey IP address and netmask should be on a separate subnet from the storage pkey IP address and netmask.

The tables below can be used as reference for one particular RAC cluster:

**Table 5-2 Existing Configuration**

Interface Name	IP Address	Netmask
ib0	192.168.12.153	255.255.248.0
ib1	192.168.12.154	255.255.248.0

The following table shows the new IP addresses and netmasks required by the pkey interfaces while implementing InfiniBand Partitioning for that one Oracle RAC cluster.

**Table 5-3 New IP Addresses and Netmasks Required by the pkey Interfaces**

Interface Name	IP Address	Netmask
clib0	192.168.112.1	255.255.248.0
clib1	192.168.112.2	255.255.248.0
stib0	192.168.114.1	255.255.240.0
stib1	192.168.114.2	255.255.240.0

### 5.24.4 Configuring InfiniBand Partitioning across Oracle VM RAC Clusters

The steps for configuring InfiniBand Partitioning across Oracle RAC clusters running in Oracle VM are described here.

In this procedure, the Oracle RAC clusters incur a minimal downtime. The downtime occurs when the Oracle RAC cluster is restarted to use the new interfaces.

Before you start this task, download and untar the file `create_pkeys.tar`. This file can be downloaded from [Implementing InfiniBand Partitioning across OVM RAC clusters on Exadata \(My Oracle Support Doc ID 2075398.1\)](#). The file should be downloaded to one of the management domain (dom0) nodes. This is the node that you will use for

running all the scripts in this procedure. This node will be referred to as *driver\_dom0* in this procedure.

When you untar the file, you should get three files:

- `create_pkeys_on_switch.sh`
- `run_create_pkeys.sh`
- `create_pkey_files.sh`

**1.** Allocate IP addresses to be used by the pkey interfaces.

Plan and allocate sets of IP addresses and netmasks for each Oracle VM RAC cluster that will be used by the cluster pkey interfaces and the storage pkey interfaces when InfiniBand partitioning gets implemented in the cluster.

Refer to the topic [About InfiniBand Partitioning Network Configuration](#) for an example.

**2.** On the InfiniBand switches, create a dedicated partition (cluster pkey) for each Oracle RAC cluster to be used by the clusterware and create one partition (storage pkey) to be used by all the Oracle VM RAC clusters and the storage cells for communication between the Oracle RAC cluster nodes and the storage cells.

You assign a pkey to each partition as a simplified means of identifying the partition to the Subnet Manager. Pkeys are 15-bit integers. Values `0x0001` and `0x7fff` are default partitions. Use values between `0x0002` and `0x7ffe` for your pkeys.

**a.** Enable password-less ssh equivalence for the `root` user from the *driver\_dom0* management domain (dom0) node to all the switches on the InfiniBand fabric.

Use a command similar to the following where *ib\_switch\_list* refers to a file that contains the list of all the InfiniBand switches on the fabric, with each switch name on a separate line.

```
dcli -g ib_switch_list -l root -k
```

**b.** Run the script `create_pkeys_on_switch.sh` from *driver\_dom0* to create and configure the partition keys on the InfiniBand switches.

 **Note:**

Each execution of the script `create_pkeys_on_switch.sh` creates exactly one partition. You must run the script once for each partition to be created. For example, an environment that contains two Oracle VM RAC clusters will have a total of three partitions: one storage partition and two cluster partitions (one per Oracle RAC cluster). In this example, you will need to run `create_pkeys_on_switch.sh` three times.

You must run the script on only one node (*driver\_dom0*). The script creates the partitions in all the switches provided as input during the execution of the script.

**c.** After you finish running the script, verify the partitions were created on all the switches.

```
/usr/local/sbin/smpartition list active no-page
```

The following example output shows the default partitions (0x0001 and 0x7fff), and an additional partition, 0x0004. The partition with pkey 0x0004 is configured for IPoIB and has two member ports that are assigned full membership of the partition.

```
Sun DCS IB partition config file
#! version_number : 1
#! version_number : 12
Default=0x7fff, ipoib :
ALL_CAS=full,
ALL_SWITCHES=full,
SELF=full;
SUN_DCS=0x0001, ipoib :
ALL_SWITCHES=full;
 = 0x0004, ipoib:
0x0021280001cf3787=full,
0x0021280001cf205b=full;
```

At this stage ensure that you have created all the required partitions.

3. On the Oracle VM RAC nodes and on the storage cells, generate all the relevant network configuration files for the new IP over InfiniBand (IPoIB) interfaces.

Each partition requires a new IPoIB network interface.

This step makes the following changes on the Oracle RAC cluster nodes:

- Modifies these files:
  - /etc/sysconfig/network-scripts/ifcfg-ib0
  - /etc/sysconfig/network-scripts/ifcfg-ib1
- Removes these files:
  - /etc/sysconfig/network-scripts/rule-ib0
  - /etc/sysconfig/network-scripts/rule-ib1
  - /etc/sysconfig/network-scripts/route-ib0
  - /etc/sysconfig/network-scripts/route-ib1
- Creates the following new files in /etc/sysconfig/network-scripts:
  - ifcfg-clib0, ifcfg-clib1
  - rule-clib0, rule-clib1
  - route-clib0, route-clib1
  - ifcfg-stib0, ifcfg-stib1
  - rule-stib0, rule-stib1
  - route-stib0, route-stib1





FinalHexValue is the value that will be entered in the command here and HexValue is the value entered in step 2 for pkey\_id.

The following table provides an example of the inputs for the four runs for a cluster node:

**Table 5-4 Four Runs for Cluster Nodes**

Run	Interface Name	pkey_id	node_type	pkey_ipaddress	pkey_netmask	pkey_interfaceType
1	ib0	a000	compute	192.168.12.153	/21	cluster
2	ib1	a000	compute	192.168.12.154	/21	cluster
3	ib0	aa00	compute	192.168.114.155	/20	storage
4	ib1	aa00	compute	192.168.114.156	/20	storage

You use these values in each execution of the script, denoted by the Run column, as shown in this example, where vm-guest-1 is the name of the cluster node.

```
./run_create_pkeys.sh vm-guest-1 ib0 a000 compute
192.168.12.153 /21 cluster
```

At this stage all the required networking files listed at the beginning of this step have been created for the new pkey-enabled network interfaces on the Oracle VM RAC cluster nodes.

Oracle Grid Infrastructure has also been modified to make use of the new network interfaces upon restart. The output of the command `$GRID_HOME/bin/oifcfg getif` should list `clib0` and `clib1` in the list of interfaces to be used for the cluster interconnect.

4. Modify Oracle ASM and Oracle RAC `CLUSTER_INTERCONNECTS` parameter.
  - a. Log in to each of the Oracle ASM instances in the Oracle RAC cluster using SQL\*Plus as SYS, and run the following command:

```
ALTER SYSTEM SET
cluster_interconnects='<cluster_pkey_IP_address_of_ib0>:
<cluster_pkey_IP_address_of_ib1>' scope=spfile
sid='<name_of_current_ASM_instance>';
```

For example:

```
ALTER SYSTEM SET
cluster_interconnects='192.168.12.153:192.168.12.154'
scope=spfile sid='+ASM1';
```

- b. Log in to each of the database instances in the Oracle RAC cluster using SQL\*Plus, and run the same command for the Oracle RAC instance:

For example:

```
ALTER SYSTEM SET cluster_interconnects='192.168.12.153:192.168.12.154'
 scope=spfile sid='RACDB1';
```

- c. Shut down and disable CRS auto-start on all the Oracle RAC cluster nodes.

```
Grid_home/bin/crsctl stop crs

Grid_home/bin/crsctl disable crs
```

At this stage Oracle Grid Infrastructure, the Oracle ASM instances, and the Oracle Database instances have been modified to make use of the newly created network interfaces.

5. Modify `cellip.ora` and `cellinit.ora` on all the cluster nodes (user domains).

Perform these steps on any one database server node of the cluster (user domain for an Oracle VM RAC cluster).

- a. Make a backup of the `cellip.ora` and `cellinit.ora` files.

```
cd /etc/oracle/cell/network-config
cp cellip.ora cellip.ora-bak
cp cellinit.ora cellinit.ora-bak
```

- b. Modify the `cellip.ora-bak` file to replace the existing IP address with the two storage pkey IP addresses of every storage cell that will be setup in step 7.

The two IP addresses are separated by a semi-colon (;).

- c. Make sure ssh equivalence is set up for the `root` user to all the cluster nodes from this cluster node.
- d. Replace the `cellip.ora` file on all the cluster nodes.

Use the following commands to backup and then replace the `cellip.ora` file on all the cluster nodes. In this example `cluster_nodes` refers to a file containing the names of all the Oracle RAC cluster nodes of the Oracle VM RAC cluster, with each node on a separate line.

```
/usr/local/bin/dcli -g cluster_nodes -l root
"/bin/cp /etc/oracle/cell/network-config/cellip.ora /e
tc/oracle/cell/network-config/cellip-orig.ora"

/usr/local/bin/dcli -g cluster_nodes -l root -f celli
p.ora-bak -d /etc/oracle/cell/network-config/cellip.ora
```

- e. Manually edit the `/etc/oracle/cell/network-config/cellinit.ora-bak` file to replace the existing IP addresses and netmask with the two storage pkey IP addresses and netmask of the cluster node which was used in step 3.
- f. Make sure ssh equivalence is set up for the `root` user to all the cluster nodes from this cluster node.
- g. Replace the `cellinit.ora` file on all the cluster nodes.

The IP address and netmask were used in the third and fourth run of step 3.

Use the following commands to backup and then replace the `cellinit.ora` file on all the cluster nodes. In this example `cluster_nodes` refers to a file containing the names of all the Oracle RAC cluster nodes of the Oracle VM RAC cluster, with each node on a separate line.

```
/usr/local/bin/dcli -g cluster_nodes -l root
"/bin/cp /etc/oracle/cell/network-config/cellinit.ora /e
tc/oracle/cell/network-config/cellinit-orig.ora"

/usr/local/bin/dcli -g cluster_nodes -l root -f cellini
t.ora-bak -d /etc/oracle/cell/network-config/cellinit.ora
```

6. In the management domains (dom0s), modify the user domain configuration file for each user domain to use the partition key applicable to that user domain.

Modify all the relevant `vm.cfg` files in the management domain. This step is applicable only for Oracle VM environments. Log in to all the management domains and manually edit `/EXAVMIMAGES/GuestImages/user_domain_name/vm.cfg` to include the partition keys created in step 2.

For example, modify the line:

```
ib_pkeys = [{"pf":'40:00.0','port':'1','pkey':['0xffff',],},
{"pf":'40:00.0','port':'2','pkey':['0xffff',],},]
```

to:

```
ib_pkeys = [{"pf":'40:00.0','port':'1','pkey':['0xa000',
'0xaa00',],}, {"pf":'40:00.0','port':'2','pkey':['0xa000',
'0xaa00',],},]
```

In this example, `0xa000` is the cluster partition key derived from the cluster `pkey_id` value entered in step 2, and `0xaa00` is the storage partition key derived from the storage `pkey_id` value.

You can use the following command to derive the partition key values to use in `vm.cfg` from the `pkey_id` values entered in step 2.

```
FinalHexValue=$(echo "obase=16;ibase=2;$(expr 100000000
0000000 + $(echo "obase=2;ibase=16;$(echo $HexValue|tr
[:lower:] [:upper:]")|bc))"|bc|tr [:upper:] [:lower:]
```

`FinalHexValue` is the value that you enter in `vm.cfg` and `HexValue` is the value entered in step 2 for `pkey_id`.

 **Note:**

If your environment has multiple Oracle VM RAC clusters, the next two steps (step 7 and step 8) should be performed only once AFTER steps 3 through step 6 have been executed for all the Oracle VM RAC clusters.

7. Modify the storage cells to use the newly created IPoIB interfaces.



You use these values in each execution of the script, denoted by the Run column, as shown in this example, where `cell101` is the name of the storage server.

```
./run_create_pkeys.sh cell101 ib0 aa00 cell 192.168.114.1 /20
storage
```

 **Note:**

You can ignore the following messages from the script. The restart of the storage cells at the end of this task will take care of these issues.

```
Network configuration altered. Please issue the
following commands
as root to restart the network and open IB stack:
 service openibd restart
 service network restart
A restart of all services is required to put new network
configuration into
effect. MS-CELLSRV communication may be hampered until
restart.
```

At this stage the storage servers (cells) have been modified to use the new network interfaces upon restart.

8. Modify the `/opt/oracle.cellos/cell.conf` file on each storage sever and restart the storage servers.
  - a. Make a backup of the `/opt/oracle.cellos/cell.conf` file.

```
cd /opt/oracle.cellos
cp cell.conf cell.conf-prepkey
```

- b. Change the Pkey configuration lines in `/opt/oracle.cellos/cell.conf`.

Change this line:

```
<Pkeyconfigured>no</Pkeyconfigured>
```

to:

```
<Pkeyconfigured>yes</Pkeyconfigured>
```

Change this line for the 2 private interfaces `ib0` and `ib1`:

```
<IP_enabled>yes</IP_enabled>
```

to:

```
<IP_enabled>no</IP_enabled>
```

- c. Make sure Oracle Grid Infrastructure is stopped on all Oracle VM RAC nodes.
- d. Restart all the storage cell servers.

```
shutdown -r now
```

- e. Verify that the new pkey-enabled network interfaces are in use.

```
cellcli -e list cell detail | egrep 'interconnect|ipaddress'
```

The output should show the new pkey-enabled interfaces (`stib0` and `stib1`) along with the new set of IP addresses.

9. Restart the Oracle RAC clusters.
  - a. Log in to the corresponding management domain of each of the user domain nodes.
  - b. Run the following commands:

```
xm shutdown user_domain_name
```

```
xm create /EXAVMIMAGES/GuestImages/user_domain_name/vm.cfg
```

10. Start and verify the Oracle Grid Infrastructure stack is fully started on all the cluster nodes.
  - a. Start and enable auto-start of the Oracle Grid Infrastructure stack on all the Oracle RAC cluster nodes.

```
$GRID_HOME/bin/crsctl start crs
```

```
$GRID_HOME/bin/crsctl enable crs
```

- b. After Oracle Grid Infrastructure has started on all the nodes, verify the `cluster_interconnects` parameter is set to use the newly configured pkey interfaces.

Log in to a database instance and run the following query:

```
SQL> SELECT inst_id, value FROM gv$parameter
WHERE name = 'cluster_interconnects'
```

- c. Remove the old cluster interconnect interfaces from the Oracle Cluster Registry (OCR).

```
Grid_home/bin/oifcfg delif -global ib0/<old subnet>
```

```
Grid_home/bin/oifcfg delif -global ib1/<old subnet>
```

## 5.24.5 Implementing InfiniBand Partitioning across OVM RAC Clusters: Setting up Limited Membership

The 12.1.0.2 October 2016 Database Bundle Patch introduces a security enhancement feature where the GUIDs of the database nodes can be assigned to the storage pkey with limited membership instead of full membership, as was the case prior to the 12.1.0.2 October

2016 Bundle Patch. This addresses a security concern where one RAC node from one RAC cluster could talk to a RAC node from another RAC cluster using the storage pkey interfaces.

### Full Membership and Limited Membership

An InfiniBand partition defines a group of InfiniBand nodes that are allowed to communicate with one another. With InfiniBand partitioning, you define custom or unique partition keys that are managed by the master subnet manager, and assign members to the custom partition keys. Members with the same partition key can only communicate amongst themselves. A member of one partition key cannot communicate with a member that has a different partition key, regardless of membership type. The OVM RAC cluster nodes of one cluster are assigned one partition key for clusterware communication and another partition key for communication with storage cells. This way, the nodes of one RAC cluster will not be able to communicate with the nodes of another RAC cluster, which have a different partition key assigned to them. This is very similar conceptually to tagged VLANs in the Ethernet world.

Partition keys (pkeys) are 15-bit integers and have a value of 0x1 to 0x7FFF. An additional bit, the membership bit, identifies the membership of a member of the partition. Memberships can be:

- Full: The membership bit is set to 1. Members with full membership can communicate with each other as well as members with limited membership within same the partition key.
- Limited: The membership bit is set to 0. Members with limited membership within a partition cannot communicate with each other. However they can communicate with other members with full membership within the same partition.

Combined together, the pkey and the membership bit comprise a 16-bit integer. The most significant bit is the membership bit.

By default, the InfiniBand subnet manager provides a single partition and it is identified by the partition key 0x7FFF (limited membership) or 0xFFFF (full membership).

An HCA port can participate in a maximum of 128 partitions. Each partition key provides a new IPoIB network interface. For example, InfiniBand port 1 with partition key 0xa001 will result in a new network interface. These interfaces are named with meaningful names through the `ifcfg-<interface>` file parameters.

An InfiniBand node can be a member of multiple partitions. When a packet arrives at a database node, the partition key (pkey) of the packet is matched with the Subnet Manager configuration. This validation prevents a database node from communicating with another database node outside of the partitions of which it is a member.

Every node within the infiniband fabric has a partition key table which you can see in `/sys/class/infiniband/mlx4_0/ports/[1-2]/pkeys`. Every Queue Pair (QP) of the node has an index (pkey) associated with it that maps to an entry in that table. Whenever a packet is sent from the QP's send queue, the indexed pkey is attached with it. Whenever a packet is received on the QP's receive queue, the indexed pkey is compared with that of the incoming packet. If it does not match, the packet is silently discarded. The receiving Channel Adapter does not know it arrived and the sending Channel Adapter gets no acknowledgement as well that it was received. The sent packet simply gets manifested as a lost packet. It is only when the pkey of the incoming packet matches the indexed pkey of the QP's receive queue, a handshake is made and the packet is accepted and an acknowledgment is sent to the sending



channel adapter. This is how only members of the same partition are able to communicate with each other and not with hosts that are not members of that partition (which means those hosts that does not have that pkey in their partition table).

The steps below describe how to set up this enhancement on a pkey-enabled environment that has the 12.1.0.2 October 2016 Database Bundle Patch applied. There are two possible scenarios, as described below:

### Case 1. Implementing the feature on a pkey-enabled environment in a rolling manner

In this case, you have already applied the 12.1.0.2 October 2016 Database Bundle Patch.

Perform the steps below on one node at a time.

1. Shut down the Grid Infrastructure on the node.

```
$GI_HOME/bin/crsctl stop crs
```

2. Determine the two port GUIDs of the dom0 (control domain) which manages this user domain OVM RAC cluster node.

```
/usr/sbin/ibstat | grep Port
```

3. Login to the Infiniband Switch where the SM master is running as root.

4. Run the commands below on the InfiniBand switch.

```
/usr/local/sbin/smpartition start
```

```
/usr/local/sbin/smpartition modify -n <storage pkey name> -port <Port GUID1 of the dom0 from step 2> -m limited
```

```
/usr/local/sbin/smpartition modify -n <storage pkey name> -port <Port GUID2 of the dom0 from step 2> -m limited
```

```
/usr/local/sbin/smpartition commit
```

5. Modify the `vm.cfg` file for this OVM RAC user domain node in the dom0.

- a. Login to the dom0 as root.

- b. Edit `/EXAVMIMAGES/GuestImages/<user domain name>/vm.cfg` and modify the partition keys as shown in the example below.

Modify this line:

```
ib_pkeys = [{ 'pf': '40:00.0', 'port': '1', 'pkey':
['0xclpkey', '0x<stpkey>',] }, { 'pf': '40:00.0', 'port': '2', 'pkey':
['0xclpkey', '0x<stpkey>',] },]
```

to this:

```
ib_pkeys = [{ 'pf': '40:00.0', 'port': '1', 'pkey':
['0xclpkey', '0x<mod_stpkey>',] }, { 'pf': '40:00.0', 'port': '2', 'pkey':
['0xclpkey', '0x<mod_stpkey>',] },]
```

`<mod_stpkey>` is derived from `<stpkey>` using the formula below:

```
mod_stpkey=$(echo "obase=16;ibase=2;$(expr $(echo "obase=2;ibase=16;$(echo $stpkey|tr [:lower:] [:upper:])"|bc) - 1000000000000000)"|bc|tr [:upper:] [:lower:])
```

Note that `<stpkey>` and `<mod_stpkey>` in the formula above are specified without the "0x" prefix.

6. Modify the `/etc/sysconfig/network-scripts/ifcfg-stib*` files on the user domain RAC nodes.

Edit the PKEY\_ID in those files using the formula below:

```
mod_stpkey=$(echo "obase=16;ibase=2;$(expr $(echo "obase=2;ibase=16;$(echo $stpkey|tr [:lower:] [:upper:])"|bc) - 1000000000000000)"|bc|tr [:upper:] [:lower:])
```

`mod_stpkey` is the new PKEY\_ID, and `stpkey` is the old PKEY\_ID.

Note that `<stpkey>` and `<mod_stpkey>` in the formula above are specified without the "0x" prefix.

7. Modify `/opt/oracle.cellos/pkey.conf` on the user domain RAC nodes.

Edit the Pkey for the storage network pkey interfaces (stib\*):

Change:

```
<Pkey>0xstpkey</Pkey>
```

to:

```
<Pkey>0xmod_stpkey</Pkey>
```

`mod_stpkey` is derived from `stpkey` using the formula below:

```
mod_stpkey=$(echo "obase=16;ibase=2;$(expr $(echo "obase=2;ibase=16;$(echo $stpkey|tr [:lower:] [:upper:])"|bc) - 1000000000000000)"|bc|tr [:upper:] [:lower:])
```

`stpkey` and `mod_stpkey` used in the formula above are specified without the "0x" prefix.

8. Restart the OVM RAC user domain node.

- a. Login to the dom0 as root.
- b. Run the following commands:

```
xm shutdown <user domain name>
```

```
xm create /EXAVMIMAGES/GuestImages/<user domain name>/vm.cfg
```

9. Verify the Grid Infrastructure stack is fully up on the cluster node.
10. Repeat the steps on the remaining cluster nodes, one node at a time.

### Case 2. Implementing the feature on a pkey-enabled environment while you apply the 12.1.0.2 October 2016 Database Bundle Patch in a rolling manner

Perform the steps below on one node at a time.

1. Apply the 12.1.0.2 October 2016 Database Bundle Patch on the cluster node.
2. Run the steps 1 through 10 from Case 1 above on the node where the patch was applied.
3. Move on to the next cluster node and repeat steps 1 and 2 above.

 **Note:**

Once the dom0 GUIDs are converted to limited membership, deployment of any new cluster will have the October 2016 Database Bundle Patch as a prerequisite.

## 5.25 Running Oracle EXAchk in Oracle VM Environments

Oracle EXAchk version 12.1.0.2.2 and higher supports virtualization on Exadata Database Machine.

To perform the complete set of Oracle EXAchk audit checks in an Exadata Database Machine Oracle VM environment, Oracle EXAchk must be installed in and run from multiple locations, as follows:

- From one management domain (dom0)
- From one user domain (domU) in each Oracle VM Oracle Real Application Clusters (Oracle RAC) cluster

For example, an Exadata Database Machine Quarter Rack with 2 database servers containing 4 Oracle VM Oracle RAC clusters (2 nodes per cluster for a total of 8 domU's across both database servers) requires running Oracle EXAchk 5 separate times, as follows:

1. Run Oracle EXAchk in the first user domain (domU) for the first cluster.
2. Run Oracle EXAchk in the first user domain (domU) for the second cluster.
3. Run Oracle EXAchk in the first user domain (domU) for the third cluster.
4. Run Oracle EXAchk in the first user domain (domU) for the fourth cluster.
5. Run Oracle EXAchk in the first management domain (dom0).

The audit checks performed by Oracle EXAchk are specified in the following table:

**Table 5-6 Audit Checks Performed by Oracle EXAchk**

Where to Install and Run Oracle EXAchk	Audit Checks Performed
Management domain (dom0)	Hardware and operating system level checks for: <ul style="list-style-type: none"> <li>• Database servers (management domains)</li> <li>• Storage servers</li> <li>• RDMA Network Fabric</li> <li>• RDMA Network Fabric switches</li> </ul>
User domain (domU)	Operating system level checks for user domains, and checks for Oracle Grid Infrastructure and Oracle Database

### Oracle EXAchk Command Line Options

Oracle EXAchk requires no special command line options. It automatically detects that it is running in an Exadata Database Machine Oracle VM environment and whether it is running in

a management domain or user domain and performs the applicable audit checks. For example, in the simplest case, you can run Oracle EXAchk with no command line options:

```
./exachk
```

When Oracle EXAchk is run in the management domain, it performs audit checks on all database servers, storage servers, and RDMA Network Fabric switches accessible through the RDMA Network Fabric.

To run Oracle EXAchk on a subset of servers or switches, use the following command line options:

**Table 5-7 Command Line Options for Oracle EXAchk**

Option	Description
-clusternodes	Specifies a comma-separated list of database servers.
-cells	Specifies a comma-separated list of storage servers.
-ibswitches	Specifies a comma-separated list of RDMA Network Fabric switches.

For example, for an Exadata Database Machine Full Rack where only the first Quarter Rack is configured for virtualization, but all components are accessible through the RDMA Network Fabric, you can run a command similar to the following from the database server `dm01adm01`:

```
./exachk -clusternodes dm01adm01,dm01adm02
-cells dm01celadm01,dm01celadm02,dm01celadm03
-ibswitches dm01swibs0,dm01sw-iba0,dm01sw-ibb0
```

### Related Topics

- Compliance Framework (Oracle ORAchk and Oracle EXAchk) Command-Line Options

# 6

## Managing Oracle Linux KVM Guests

Starting with Exadata Database Machine X8M-2, Oracle Linux KVM is the virtualization technology for systems that use RoCE Network Fabric.

- [Oracle Linux KVM and Exadata Database Machine](#)  
When deploying Exadata Database Machine X8M-2 or later, you can decide to implement Oracle Linux KVM on the database servers.
- [Migrating a Bare Metal Oracle RAC Cluster to an Oracle RAC Cluster in Oracle Linux KVM](#)  
You can move an existing Oracle Real Application Clusters (Oracle RAC) cluster into a virtual environment that is managed by KVM.
- [Showing Running Domains](#)  
Use the `vm_maker` utility to list the running domains.
- [Starting a Guest](#)  
You can start a guest manually, or configure the guest to start automatically when the KVM host is started.
- [Starting a Guest using the Diagnostic ISO File](#)  
Use this procedure to boot a guest using the diagnostic ISO file (`diagnostics.iso`).
- [Monitoring a Guest Console During Startup](#)  
To see Oracle Linux boot messages during guest startup, use the `--console` option with the `vm_maker --start-domain` command.
- [Managing Automatic Startup of Oracle Linux KVM Guests](#)  
By default, when you create a guest, it is configured to automatically start when the KVM host is started. You can enable and disable this feature as needed.
- [Shutting Down a User Domain From Within the User Domain](#)
- [Shutting Down a Guest From Within the KVM host](#)  
You can shut down a guest from within a KVM host.
- [Backing up the KVM host and Guests](#)  
In an Oracle Linux KVM deployment, you need to back up the KVM host and the guests.
- [Backing Up and Restoring Oracle Databases on KVM Guests](#)  
Backing up and restoring Oracle databases on KVM guests is the same as backing up and restoring Oracle databases on physical nodes.
- [Modifying the Memory Allocated to a Guest](#)  
You can modify the memory allocated to a guest using `vm_maker`.
- [Modifying the Number of Virtual CPUs Allocated to a Guest](#)  
You can dynamically modify the number of virtual CPUs allocated to a guest with the `vm_maker --set --vcpu` command.
- [Increasing the Disk Space in a Guest](#)  
The KVM guest local space can be extended after initial deployment by adding local disk images.

- [Expanding /EXAVMIMAGES on the KVM host](#)  
Use this procedure to expand /EXAVMIMAGES using available space on the KVM host.
- [Adding an Oracle Linux KVM Cluster](#)  
You can use Oracle Exadata Deployment Assistant (OEDA) to create a new Oracle Linux KVM cluster on an existing Exadata Database Machine.
- [Expanding an Oracle RAC Cluster in Oracle Linux KVM Using OEDACLI](#)  
You can expand an existing Oracle RAC cluster on Oracle Linux KVM by adding guests using the Oracle Exadata Deployment Assistant command-line interface (OEDACLI).
- [Moving a Guest to a Different Database Server](#)  
Guests can move to different database servers.
- [Recovering a KVM Deployment](#)  
A KVM host can be recovered from backups and guests can be recovered from snapshot backups.
- [Removing a Guest](#)  
You can remove a guest in Oracle Linux KVM using either OEDACLI or the `vm_maker` utility.
- [Using Client Network VLAN Tagging with Oracle Linux KVM](#)  
This topic describes the implementation of tagged VLAN interfaces for the client network in conjunction with Oracle Linux KVM.
- [Using Exadata Secure RDMA Fabric Isolation with Oracle Linux KVM](#)  
This topic describes the implementation of Exadata Secure RDMA Fabric Isolation in conjunction with Oracle Linux KVM.
- [Adding a Bonded Network Interface to an Oracle Linux KVM Guest](#)  
Use this procedure to add a bonded network interface to an existing Oracle Linux KVM guest.
- [Using Oracle EXAchk in Oracle Linux KVM Environments](#)  
Oracle EXAchk version 12.1.0.2.2 and higher supports virtualization on Exadata Database Machine.

## 6.1 Oracle Linux KVM and Exadata Database Machine

When deploying Exadata Database Machine X8M-2 or later, you can decide to implement Oracle Linux KVM on the database servers.

A KVM host and one or more guests are installed on every database server. You can configure Oracle Linux KVM environments on your initial deployment using scripts created by Oracle Exadata Deployment Assistant (OEDA) or you can migrate an existing environment to Oracle Linux KVM.



### Note:

Oracle Linux KVM is not supported on 8-socket servers, such as X8M-8.

- [About Oracle Linux KVM](#)  
Oracle Linux KVM enables you to deploy the Oracle Linux operating system and application software within a supported virtual environment that is managed by KVM.
- [Oracle Linux KVM Deployment Specifications and Limits](#)  
This topic describes the deployment specifications and limits for using Oracle Linux KVM on Oracle Exadata Database Machine.
- [Supported Operations in the KVMHost](#)  
Manually modifying the KVM host can result in configuration issues, which can degrade performance or cause a loss of service.
- [Oracle Linux KVM Resources](#)  
Two fundamental parts of the Oracle Linux KVM infrastructure – networking and storage – are configured outside of Oracle Linux KVM.

### 6.1.1 About Oracle Linux KVM

Oracle Linux KVM enables you to deploy the Oracle Linux operating system and application software within a supported virtual environment that is managed by KVM.

If you use Oracle Linux KVM on Exadata Database Machine, then Oracle Linux KVM provides CPU, memory, operating system, and system administration isolation for your workloads. You can combine virtual machines (VMs) with network and I/O prioritization to achieve full stack isolation. For consolidation, you can create multiple trusted databases or pluggable databases using Oracle Linux KVM, allowing resources to be shared more dynamically.

Starting with Oracle Exadata System Software release 19.3.0, KVM is the virtualization technology used with Exadata Database Machine systems configured with RDMA over Converged Ethernet (RoCE) interconnects. An Oracle Linux KVM environment consists of a management server (the KVM host), virtual machines, and resources. A KVM host is a managed virtual environment providing a lightweight, secure, server platform which runs VMs, also known as guests.

The KVM host is installed on a bare metal computer. The hypervisor on each KVM host is an extremely small-footprint VM manager and scheduler. It is designed so that it is the only fully privileged entity in the system. It controls only the most basic resources of the system, including CPU and memory usage, privilege checks, and hardware interrupts.

The hypervisor securely executes multiple VMs on one host computer. Each VM runs in its own guest and has its own operating system. The KVM host also runs as a guest on top of the hypervisor. The KVM host has privileged access to the hardware and device drivers. This is the environment from where you manage all the guests.

A guest is an unprivileged VM that can access the RoCE interface. The guest is started and managed on the KVM host. Because a guest operates independently of other VMs, a configuration change applied to the virtual resources of a guest does not affect any other guests. A failure of the guest does not impact any other guests.

The terms "guest", "domain", and "virtual machine" are often used interchangeably.

In general, each KVM host supports up to 12 guests. However, the limit is 8 guests on servers that contain 384 GB of RAM and are configured to support Exadata Secure RDMA Fabric Isolation.

Each guest is started alongside the KVM host. The guests never interact with the KVM host directly. Their requirements are handled by the hypervisor itself. The KVM host only provides a means to administer the hypervisor.

You use Oracle Exadata Deployment Assistant (OEDA) to create and configure Oracle Linux KVMs on Exadata Database Machine.

## 6.1.2 Oracle Linux KVM Deployment Specifications and Limits

This topic describes the deployment specifications and limits for using Oracle Linux KVM on Oracle Exadata Database Machine.

**Table 6-1 Oracle Linux KVM Deployment Specifications and Limits**

Attribute	Value for X8M-2	Value for X8M-2 Eighth Rack	Value for X9M-2	Value for X9M-2 Eighth Rack
Maximum number of Oracle Linux KVM guests on each database server	On servers that contain 384 GB of RAM and are configured to support Exadata Secure RDMA Fabric Isolation: 8 Otherwise: 12	On servers that contain 384 GB of RAM and are configured to support Exadata Secure RDMA Fabric Isolation: 8 Otherwise: 12	12	12
Total physical memory on each database server	Default: 384 GB Maximum: 1536 GB	Default: 384 GB Maximum: 768 GB	Default: 512 GB Maximum: 2048 GB	Default: 384 GB Maximum: 1024 GB
Total available memory on each database server for all Oracle Linux KVM guests	Default: 328 GB Maximum: 1390 GB	Default: 328 GB Maximum: 680 GB	Default: 440 GB Maximum: 1870 GB	Default: 328 GB Maximum: 920 GB
Minimum memory limit for each Oracle Linux KVM guest	16 GB	16 GB	16 GB	16 GB
Total CPU cores (vCPUs) on each database server	48 (96)	24 (48)	64 (128)	32 (64)
Over-provisioning limit for CPU cores (vCPUs) on each database server for all Oracle Linux KVM guests <b>Note:</b> CPU over-provisioning may cause performance conflicts.	92 (184)	46 (92)	124 (248)	62 (124)
CPU core (vCPU) limits for each Oracle Linux KVM guest	Minimum: 2 (4) Maximum: 46 (92)	Minimum: 2 (4) Maximum: 23 (46)	Minimum: 2 (4) Maximum: 62 (124)	Minimum: 2 (4) Maximum: 31 (62)



**Table 6-1 (Cont.) Oracle Linux KVM Deployment Specifications and Limits**

Attribute	Value for X8M-2	Value for X8M-2 Eighth Rack	Value for X9M-2	Value for X9M-2 Eighth Rack
Total usable disk storage for Oracle Linux KVM guests on each database server	Default: 3.15 TB	Default: 3.15 TB	Default: 3.15 TB	Default: 3.15 TB

**Table 6-2 Oracle Linux KVM Available Memory Limits**

Database server total physical memory	Total available memory for Oracle Linux KVM guests
384 GB	328 GB
512 GB	440 GB
768 GB	680 GB
1024 GB	920 GB
1536 GB	1390 GB
2048 GB	1870 GB



**Note:**

1 CPU core = 1 OCPU = 2 vCPUs = 2 hyper-threads

### 6.1.3 Supported Operations in the KVMHost

Manually modifying the KVM host can result in configuration issues, which can degrade performance or cause a loss of service.



**Caution:**

Oracle does not support any changes that are made to the KVM host beyond what is documented. Third-party applications can be installed on the KVM host and guests, but if there are issues with the Oracle software, then Oracle Support Services may request the removal of the third-party software while troubleshooting the cause.

If you are in doubt whether an operation on the KVM host is supported, contact Oracle Support Services.

## 6.1.4 Oracle Linux KVM Resources

Two fundamental parts of the Oracle Linux KVM infrastructure – networking and storage – are configured outside of Oracle Linux KVM.

### Networking

When specifying the configuration details for your Oracle Exadata Rack using Oracle Exadata Deployment Assistant (OEDA), you provide input on how the required network IP addresses for Oracle Linux KVM environments should be created. The generated OEDA setup files are transferred to the Oracle Exadata Rack and used to create the network addresses.

### Storage

Oracle Linux KVM always requires a location to store environment resources that are essential to the creation and management of virtual machines (VMs). These resources include ISO files (virtual DVD images), VM configuration files and VM virtual disks. The location of such a group of resources is called a storage repository.

On Exadata Database Machine, storage for the Oracle Linux KVMs uses an XFS file system.

On 2-socket Oracle Exadata Database Machine systems only, you can purchase a disk expansion kit to increase storage capacity. You can use the additional disk space to support more Oracle Linux KVM guests (up to a maximum of 12) by expanding `/EXAVMIMAGES` or to increase the size of the `/u01` partition in each guest.

### Related Topics

- Using Oracle Exadata Deployment Assistant

## 6.2 Migrating a Bare Metal Oracle RAC Cluster to an Oracle RAC Cluster in Oracle Linux KVM

You can move an existing Oracle Real Application Clusters (Oracle RAC) cluster into a virtual environment that is managed by KVM.

### Note:

This topic applies only to two-socket x86 servers. It does not apply to eight-socket servers such as Exadata Database Machine X8M-8.

The migration of a bare metal Oracle RAC cluster to an Oracle RAC cluster in Oracle Linux KVM can be achieved in the following ways:

- Migrate to Oracle RAC cluster in Oracle Linux KVM using the existing bare metal Oracle RAC cluster with zero downtime.
- Migrate to Oracle RAC cluster in Oracle Linux KVM by creating a new Oracle RAC cluster in Oracle Linux KVM with minimal downtime.

- Migrate to Oracle RAC cluster in Oracle Linux KVM using Oracle Data Guard with minimal downtime.
- Migrate to Oracle RAC cluster in Oracle Linux KVM using Oracle Recovery Manager (RMAN) backup and restore with complete downtime.

The conversion of a bare metal Oracle RAC cluster to an Oracle RAC cluster in Oracle Linux KVM has the following implications:

- Each of the database servers will be converted to an Oracle Linux KVM server on which a KVM host is created along with one or more guests, depending on the number of Oracle RAC clusters being deployed. Each guest on a database server will belong to a particular Oracle RAC cluster.
- As part of the conversion procedure, the bare metal Oracle RAC cluster will be converted to one Oracle RAC cluster in Oracle Linux KVM to start with. There will be one guest per database server.
- At the end of the conversion, the cell disk and grid disk configuration of the storage cells are the same as they were at the beginning of the conversion.
- The KVM host uses a small portion of the system resources on each database server. Typically, the KVM host uses 24 GB plus 6% of the server RAM, and 4 virtual CPUs. Take these resource requirements into consideration when sizing the SGA for databases running in conjunction with Oracle Linux KVM.
- Refer to My Oracle Support note 2099488.1 for the complete instructions.

#### Related Topics

- [Migration of a Bare metal RAC cluster to an OVM RAC cluster on Exadata \(My Oracle Support Doc ID 2099488.1\)](#)

## 6.3 Showing Running Domains

Use the `vm_maker` utility to list the running domains.

1. Connect to the management domain.
2. Run the command `/opt/exadata_ovm/vm_maker --list-domains` to list the domains.

```
/opt/exadata_ovm/vm_maker --list-domains
dm01db01vm01.example.com(55) : running
dm01db01vm02.example.com(57) : running
dm01db01vm03.example.com(59) : running
```

To view memory or CPU distribution for the domains, there are separate commands:

- `/opt/exadata_ovm/vm_maker --list --memory`
- `/opt/exadata_ovm/vm_maker --list --vcpu`

## 6.4 Starting a Guest

You can start a guest manually, or configure the guest to start automatically when the KVM host is started.

1. Connect to the KVM host.

2. To manually start a guest, use `vm_maker` to start the guest.

In the following example, `db01_guest01.example.com` is the name of the guest.

```
/opt/exadata_ovm/vm_maker --start-domain db01_guest01.example.com
[INFO] Running 'virsh start db01_guest01.example.com...'
Domain db01_guest01.example.com started

[INFO] The domain has been started but may not have network
connectivity for
several minutes.
```

3. To configure autostart for a guest, use the `vm_maker --autostart` command.

In the following example, `db01_guest01.example.com` is the name of the guest.

```
/opt/exadata_ovm/vm_maker --autostart db01_guest01.example.com --
enable
[INFO] Running 'virsh autostart db01_guest01.example.com'...'
Domain db01_guest01.example.com marked as autostarted
```

## 6.5 Starting a Guest using the Diagnostic ISO File

Use this procedure to boot a guest using the diagnostic ISO file (`diagnostics.iso`).

1. Connect to the KVM host.
2. Download the diagnostic ISO file (`diagnostics.iso`) corresponding to your current Oracle Exadata System Software release.

If required, use the `imageinfo` command to determine your current Oracle Exadata System Software release.

To find the diagnostic ISO file, search the My Oracle Support (MOS) patch repository using "exadata diagnostic iso" as the search term. You can also locate the Diagnostic ISO file in the Supplemental README that is associated with your Oracle Exadata System Software release. The Supplemental README for each Oracle Exadata System Software release is documented in [My Oracle Support document 888828.1](#).

3. To start a guest using the diagnostic ISO file:
  - a. Configure the guest to boot using the diagnostic ISO file.

Run the `vm_maker --boot-from-iso` command:

```
vm_maker --boot-from-iso ISO-file --domain guest-name
```

In the command:

- *ISO-file* specifies the name of the diagnostic ISO that you want to use to boot the specified guest.
- *guest-name* specifies the name of the guest that you want to boot using the specified ISO file.

For example:

```
vm_maker --boot-from-iso /root/home/diagnostics.iso --domain
dm01vm01
[INFO] Running 'virsh undefine dm01vm01.example.com'...
[INFO] Running 'virsh define /var/log/exadatatmp/
dm01vm01.example.com.xml.new.357b'...
[INFO] The domain 'dm01vm01.example.com' is ready for booting.
[INFO] Run the following command to boot from the diagnostic iso:
[INFO]
[INFO] virsh start dm01vm01.example.com --console
[INFO]
[INFO] If network is needed to be setup on the VM, run
[INFO] setup_management.sh from the console after the guest has
booted.
[INFO]
[INFO] When finished, run the following commands to restore
[INFO] the domain to boot from its hard disk:
[INFO]
[INFO] vm_maker --stop-domain dm01vm01.example.com --force
[INFO] vm_maker --boot-from-hd --domain dm01vm01.example.com
```

**b.** Boot the guest using the diagnostic ISO file.

Use the `virsh start` command specified in the output from the previous `vm_maker --boot-from-iso` command.

For example:

```
virsh start dm01vm01.example.com --console
```

The guest now boots using the diagnostic ISO file and the console is displayed in the terminal session.

**c.** If required, start the guest network.

If you need network access to the guest while in diagnostic mode, you can start a network interface and SSH server by running `setup_management.sh` from the guest console and following the prompts to supply the network details.

For example:

```
Welcome to Exadata Shell!
bash-4.2# setup_management.sh
Ethernet interface (eth0,1,2,3) with optional VLAN id (ethX.YYYY)
[eth0]:
IP Address of this host: 192.0.2.132
Netmask of this host: 255.255.255.128
Default gateway: 192.0.2.129
[INFO] 192.0.2.129 added as default gateway.
* sshd.service - OpenSSH server daemon
 Loaded: loaded (/usr/lib/systemd/system/sshd.service; disabled;
vendor preset: enabled)
 Active: inactive (dead)
 Docs: man:sshd(8)
```

```
man:sshd_config(5)
[INFO] Starting sshd service
```

4. When you are finished using the guest in diagnostic mode, stop the domain and reconfigure it to boot using its primary boot device.

Use the commands specified in the output from the previous `vm_maker --boot-from-iso` command.

For example:

```
vm_maker --stop-domain dm01vm01.example.com --force
[INFO] Running 'virsh destroy dm01vm01.example.com --graceful'...
Domain dm01vm01.example.com destroyed
[INFO] Checking for DomU shutdown...
[INFO] DomU successfully shutdown.

vm_maker --boot-from-hd --domain dm01vm01.example.com
[INFO] Running 'virsh undefine dm01vm01.example.com'...
[INFO] Running 'virsh define /var/log/exadatatmp/
dm01vm01.example.com.xml.new.eab9'...
[INFO] The domain is ready to be restarted.
```

The guest is now ready to be restarted by using the `vm_maker --start-domain` command.

## 6.6 Monitoring a Guest Console During Startup

To see Oracle Linux boot messages during guest startup, use the `--console` option with the `vm_maker --start-domain` command.

1. Connect as the `root` user to the KVM host.
2. Obtain the guest name using the `/opt/exadata_ovm/vm_maker --list-domains` command.
3. Use the following command to attach to the guest console, as part of starting the guest:

In the following command, *GuestName* is the name of the guest.

```
vm_maker --start-domain GuestName --console
```

4. Press `CTRL+]` to disconnect from the console.

## 6.7 Managing Automatic Startup of Oracle Linux KVM Guests

By default, when you create a guest, it is configured to automatically start when the KVM host is started. You can enable and disable this feature as needed.

- [Enabling Guest Automatic Start](#)  
You can configure a guest to automatically start when the KVM host is started.

- [Disabling Guest Automatic Start](#)  
You can disable a guest from automatically starting when the KVM host is started.

## 6.7.1 Enabling Guest Automatic Start

You can configure a guest to automatically start when the KVM host is started.

1. Connect to the KVM host.
2. Use `vm_maker` to enable autostart for the guest.

In the following example, `db01_guest01.example.com` is the name of the guest.

```
/opt/exadata_ovm/vm_maker --autostart db01_guest01.example.com --enable
[INFO] Running 'virsh autostart db01_guest01.example.com --enable'...
Domain db01_guest01.example.com marked as autostarted
```

## 6.7.2 Disabling Guest Automatic Start

You can disable a guest from automatically starting when the KVM host is started.

1. Connect to the KVM host.
2. Use `vm_maker` to disable autostart for the guest.

In the following example, `db01_guest01.example.com` is the name of the guest.

```
/opt/exadata_ovm/vm_maker --autostart db01_guest01.example.com --disable
[INFO] Running 'virsh autostart db01_guest01.example.com --disable'...
Domain db01_guest01.example.com unmarked as autostarted
```

## 6.8 Shutting Down a User Domain From Within the User Domain

The following procedure describes how to shut down a user domain from within a user domain:

1. Connect as the `root` user to the user domain.
2. Use the following command to shut down the domain:

```
shutdown -h now
```

## 6.9 Shutting Down a Guest From Within the KVM host

You can shut down a guest from within a KVM host.

1. Connect as the `root` user to the KVM host.
2. Use the following command to shut down the guest, where *GuestName* is the name of the guest:

```
/opt/exadata_ovm/vm_maker --stop-domain GuestName
```

To shut down all guests within the KVM host, use the following command:

```
/opt/exadata_ovm/vm_maker --stop-domain --all
```

The following is an example of the output:

```
[INFO] Running 'virsh shutdown db01_guest01.example.com'...
Domain db01_guest01.example.com is being shutdown
```

## 6.10 Backing up the KVM host and Guests

In an Oracle Linux KVM deployment, you need to back up the KVM host and the guests.

Backups are required to restore and recover from a database server physical or logical data issue where you need to restore database server operating system files.

- [Backing up the KVM host Using Snapshot-Based Backup](#)  
This procedure describes how to take a snapshot-based backup of the KVM host.
- [Backing up the Oracle Linux KVM Guests](#)  
You can back up Oracle Linux KVM guests by using the following procedures.

### 6.10.1 Backing up the KVM host Using Snapshot-Based Backup

This procedure describes how to take a snapshot-based backup of the KVM host.

The values shown in the following steps are examples and you may need to substitute different values to match your situation.

All steps must be performed as the `root` user.

1. Prepare a destination to hold the backup.

The destination should reside outside of the local machine, such as a writable NFS location, and be large enough to hold the backup `tar` files. For non-customized partitions, the space needed for holding the backup is approximately 50 GB.

You can use the following commands to prepare a backup destination using NFS.

```
mkdir -p /root/remote_FS
mount -t nfs -o rw,intr,soft,proto=tcp,noexec ip_address:/
nfs_location/ /root/remote_FS
```

In the `mount` command, `ip_address` is the IP address of the NFS server, and `nfs_location` is the NFS location holding the backups.

2. Remove the `LVDoNotRemoveOrUse` logical volume.

The logical volume `/dev/VGExaDb/LVDoNotRemoveOrUse` is a placeholder to make sure there is always free space available to create a snapshot.



Use the following script to check for the existence of the `LVDoNotRemoveOrUse` logical volume and remove it if present.

```
lvm lvdisplay --ignorelockingfailure /dev/VGExaDb/LVDoNotRemoveOrUse
if [$? -eq 0]; then
 # LVDoNotRemoveOrUse logical volume exists.
 lvm lvremove -f /dev/VGExaDb/LVDoNotRemoveOrUse
 if [$? -ne 0]; then
 echo "Unable to remove logical volume: LVDoNotRemoveOrUse. Do not
 proceed with backup."
 fi
fi
```

If the `LVDoNotRemoveOrUse` logical volume does not exist, then do not proceed with the remaining steps and determine the reason.

### 3. Determine the active system volume.

You can use the `imageinfo` command and examine the device hosting the active system partition.

```
imageinfo

Kernel version: 4.14.35-1902.5.1.4.el7uek.x86_64 #2 SMP Wed Oct 9
19:29:16 PDT 2019 x86_64
Image kernel version: 4.14.35-1902.5.1.4.el7uek
Image version: 19.3.1.0.0.191018
Image activated: 2019-11-04 19:18:32 -0800
Image status: success
Node type: KVMHOST
System partition on device: /dev/mapper/VGExaDb-LVDbSys1
```

In the `imageinfo` output, the system partition device ends with the name of the logical volume supports the active root (`/`) file system. Depending on the system image that is in use, the logical volume name is `LVDbSys1` or `LVDbSys2`. Likewise, the logical volume for the `/var` file system is either `LVDbVar1` or `LVDbVar2`.

You can also confirm the active devices by using the `df -hT` command and examining the output associated with the root (`/`) and `/var` file systems. For example:

```
df -hT
```

Filesystem	Type	Size	Used	Avail	Use%	Mounted on
devtmpfs	devtmpfs	378G	0	378G	0%	/dev
tmpfs	tmpfs	755G	1.0G	754G		
1% /dev/shm						
tmpfs	tmpfs	378G	4.8M	378G	1%	/run
tmpfs	tmpfs	378G	0	378G		
0% /sys/fs/cgroup						
/dev/mapper/VGExaDb-LVDbSys1	xfs	15G	7.7G	7.4G	52%	/
/dev/sda1	xfs	510M	112M	398M	22%	/boot
/dev/sda2	vfat	254M	8.5M	246M	4%	/
boot/efi						
/dev/mapper/VGExaDb-LVDbHome	xfs	4.0G	33M	4.0G	1%	/home

```

/dev/mapper/VGExaDb-LVDbVar1 xfs 2.0G 139M 1.9G
7% /var
/dev/mapper/VGExaDb-LVDbVarLog xfs 18G 403M 18G
3% /var/log
/dev/mapper/VGExaDb-LVDbVarLogAudit xfs 1014M 143M 872M
15% /var/log/audit
/dev/mapper/VGExaDb-LVDbTmp xfs 3.0G 148M 2.9G
5% /tmp
/dev/mapper/VGExaDb-LVDbOral xfs 100G 32G 69G
32% /u01
tmpfs tmpfs 76G 0 76G
0% /run/user/0

```

The remaining examples in the procedure use `LVDbSys1` and `LVDbVar1`, which is consistent with the above `imageinfo` and `df` output. However, if the active image is using `LVDbSys2`, then modify the examples in the following steps to use `LVDbSys2` instead of `LVDbSys1`, and `LVDbVar2` instead of `LVDbVar1`.

#### 4. Take snapshots of the logical volumes on the server.

Depending on the active system partition identified in the previous step, remember to use either `LVDbSys1` or `LVDbSys2` to identify the logical volume for the root (`/`) file system, and likewise use either `LVDbVar1` or `LVDbVar2` to identify the logical volume for the `/var` file system.

```

lvcreate -L1G -s -c 32K -n root_snap /dev/VGExaDb/LVDbSys1
lvcreate -L1G -s -c 32K -n home_snap /dev/VGExaDb/LVDbHome
lvcreate -L1G -s -c 32K -n var_snap /dev/VGExaDb/LVDbVar1
lvcreate -L1G -s -c 32K -n varlog_snap /dev/VGExaDb/LVDbVarLog
lvcreate -L1G -s -c 32K -n audit_snap /dev/VGExaDb/LVDbVarLogAudit
lvcreate -L1G -s -c 32K -n tmp_snap /dev/VGExaDb/LVDbTmp

```

#### 5. Label the snapshots.

```

xfs_admin -L DBSYS_SNAP /dev/VGExaDb/root_snap
xfs_admin -L HOME_SNAP /dev/VGExaDb/home_snap
xfs_admin -L VAR_SNAP /dev/VGExaDb/var_snap
xfs_admin -L VARLOG_SNAP /dev/VGExaDb/varlog_snap
xfs_admin -L AUDIT_SNAP /dev/VGExaDb/audit_snap
xfs_admin -L TMP_SNAP /dev/VGExaDb/tmp_snap

```

#### 6. Mount the snapshots.

Mount all of the snapshots under a common directory location; for example, `/root/mnt`.

```

mkdir -p /root/mnt
mount -t xfs -o nouuid /dev/VGExaDb/root_snap /root/mnt
mkdir -p /root/mnt/home
mount -t xfs -o nouuid /dev/VGExaDb/home_snap /root/mnt/home
mkdir -p /root/mnt/var
mount -t xfs -o nouuid /dev/VGExaDb/var_snap /root/mnt/var
mkdir -p /root/mnt/var/log
mount -t xfs -o nouuid /dev/VGExaDb/varlog_snap /root/mnt/var/log
mkdir -p /root/mnt/var/log/audit

```

```
mount -t xfs -o nouuid /dev/VGExaDb/audit_snap /root/mnt/var/log/audit
mkdir -p /root/mnt/tmp
mount -t xfs -o nouuid /dev/VGExaDb/tmp_snap /root/mnt/tmp
```

7. Back up the snapshots.

Use the following commands to write a backup file as a compressed tar file to your prepared NFS backup destination.

```
cd /root/mnt
tar --acls --xattrs --xattrs-include=* --format=pax -pjcvt /root/
remote_FS/myKVMbackup.tar.bz2 * /boot > /tmp/backup_tar.stdout 2> /tmp/
backup_tar.stderr
```

8. Check the `/tmp/backup_tar.stderr` file for any significant errors.

Errors about failing to tar open sockets, and other similar errors, can be ignored.

9. Unmount and remove all of the snapshots.

```
cd /
umount /root/mnt/tmp
umount /root/mnt/var/log/audit
umount /root/mnt/var/log
umount /root/mnt/var
umount /root/mnt/home
umount /root/mnt
lvremove /dev/VGExaDb/tmp_snap
lvremove /dev/VGExaDb/audit_snap
lvremove /dev/VGExaDb/varlog_snap
lvremove /dev/VGExaDb/var_snap
lvremove /dev/VGExaDb/home_snap
lvremove /dev/VGExaDb/root_snap
```

10. Unmount the NFS backup destination.

```
umount /root/remote_FS
```

11. Remove the mount point directories that you created during this procedure.

```
rm -r /root/mnt
rmdir /root/remote_FS
```

12. Recreate the `/dev/VGExaDb/LVDoNotRemoveOrUse` logical volume.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L2G VGExaDb -y
```

## 6.10.2 Backing up the Oracle Linux KVM Guests

You can back up Oracle Linux KVM guests by using the following procedures.

There are three ways to back up the guests:

**Table 6-3 Oracle Linux KVM guest backup approaches**

Method	Description	Managed By	Best For
<a href="#">Method 1: Back Up All of the KVM Guests</a>	From the KVM host, back up all guests in the /EXAVMIMAGES storage repository using XFS reflinks to get a consistent backup.	KVM host administrator	Recovering all guests after a compute node failure, which renders the guests unbootable.
<a href="#">Method 2: Back Up an Individual Guest</a>	From the KVM host, selectively back up a guest in the /EXAVMIMAGES storage repository using XFS reflinks to get a consistent backup.	KVM host administrator	Selective recovery of a guest after a compute node failure that renders the guest unbootable but does not affect all of the other guests.
<a href="#">Method 3: Back Up a Guest Internally</a>	Back up a guest using a snapshot-based backup procedure that is run inside the guest.	Guest administrator	Recovery of a guest after a failure where the guest is still bootable and allows root login. This method also enables selective recovery of specific files.

- [Method 1: Back Up All of the KVM Guests](#)  
You can back up all of the guests by backing up the storage repository under /EXAVMIMAGES.
- [Method 2: Back Up an Individual Guest](#)  
You can back up an individual guest by backing up its specific folder under /EXAVMIMAGES.
- [Method 3: Back Up a Guest Internally](#)  
You can take a snapshot-based backup of a guest from inside the guest.

### 6.10.2.1 Method 1: Back Up All of the KVM Guests

You can back up all of the guests by backing up the storage repository under /EXAVMIMAGES.

The backup destination should be separate from the KVM host server, such as a writable NFS location, and be large enough to hold the backup. The space needed for the backup is proportional to the number of guests deployed on the system. The space needed for each guest backup is approximately 200 GB.

#### 1. Prepare the guest images.

Use the following script to prepare the guest image backups under /EXAVMIMAGES/Backup.

```
#!/bin/bash

ScriptStarttime=$(date +%s)
```

```
printf "This script is going to remove the directory /EXAVMIMAGES/Backup
if it exists. If that is not acceptable, exit the script by typing n,
manually remove /EXAVMIMAGES/Backup and come back to rerun the script.
Otherwise, press y to continue : "
read proceed
if [[${proceed} == "n"]] || [[${proceed} == "N"]]
then
 exit 0
elif [[${proceed} != "n"]] && [[${proceed} != "N"]] && [[${proceed} !=
"y"]] && [[${proceed} != "Y"]]
then
 echo "Invalid input"
 exit 1
fi
rm -rf /EXAVMIMAGES/Backup

Create the Backup Directory

mkdirStartTime=$(date +%s)
find /EXAVMIMAGES -type d|grep -v 'lost+found'| awk '{print "mkdir -p /
EXAVMIMAGES/Backup"$1}'|sh
mkdir -p /EXAVMIMAGES/Backup/XML
mkdirEndTime=$(date +%s)
mkdirTime=$(expr ${mkdirEndTime} - ${mkdirStartTime})
echo "Backup Directory creation time : " ${mkdirTime}" seconds"

Create reflinks for files not in /EXAVMIMAGES/GuestImages

relinkothesStartTime=$(date +%s)

find /EXAVMIMAGES/ -not -path "/EXAVMIMAGES/GuestImages/*" -not -path "/
EXAVMIMAGES/Backup/*" -type f|awk '{print "cp --reflink", $0, "/EXAVMIMAGES/
Backup"$0}'|sh

relinkothesEndTime=$(date +%s)
relinkothesTime=$(expr ${relinkothesEndTime} - ${relinkothesStartTime})

echo "Reflink creation time for files other than in /EXAVMIMAGES/
GuestImages : " ${relinkothesTime}" seconds"

cp /etc/libvirt/qemu/*.xml /EXAVMIMAGES/Backup/XML

for hostName in $(virsh list|egrep -v 'Id|^-'|awk '{print $2}'|sed '/^$/d')
do

Pause the guests

PauseStartTime=$(date +%s)
virsh suspend ${hostName}
PauseEndTime=$(date +%s)
PauseTime=$(expr ${PauseEndTime} - ${PauseStartTime})
echo "SuspendTime for guest - ${hostName} : " ${PauseTime}" seconds"

Create reflinks for all the files in /EXAVMIMAGES/GuestImages
```

```

relinkStartTime=$(date +%s)

find /EXAVMIMAGES/GuestImages/${hostName} -type f | awk '{print "cp
--relink", $0, "/EXAVMIMAGES/Backup"$0}' | sh

relinkEndTime=$(date +%s)
relinkTime=$(expr ${relinkEndTime} - ${relinkStartTime})
echo "Reflink creation time for guest - ${hostName} :" $
{relinkTime}" seconds"

Unpause the guest

unPauseStartTime=$(date +%s)
virsh resume ${hostName}
unPauseEndTime=$(date +%s)
unPauseTime=$(expr ${unPauseEndTime} - ${unPauseStartTime})
echo "ResumeTime for guest - ${hostName} :" ${unPauseTime}"
seconds"

done

ScriptEndtime=$(date +%s)
ScriptRunTime=$(expr ${ScriptEndtime} - ${ScriptStarttime})
echo ScriptRunTime ${ScriptRunTime}" seconds"

```

## 2. Create a backup of the guest images.

Back up all of the reflink files under `/EXAVMIMAGES/Backup` to a remote location. The backup enables restoration if the KVM host is permanently damaged or lost.

For example:

```

mkdir -p /remote_FS
mount -t nfs -o rw,intr,soft,proto=tcp,nolock ip_address:/
nfs_location/ /remote_FS
cd /EXAVMIMAGES/Backup
tar --acls --xattrs --xattrs-include=* --format=pax -pjcvf /
remote_FS/exavmimages.tar.bz2 * > /tmp/exavmimages_tar.stdout
2> /tmp/exavmimages_tar.stderr

```

In the `mount` command, `ip_address` is the IP address of the NFS server, and `nfs_location` is the NFS location holding the backup.

After the backup completes, check for any significant errors from the `tar` command. In the previous example, the `tar` command writes errors to the file at `/tmp/exavmimages_tar.stderr`.

## 3. Remove the `/EXAVMIMAGES/Backup` directory and its contents.

For example:

```

cd /
rm -rf /EXAVMIMAGES/Backup

```

## 4. Unmount the NFS backup location and remove the mount point directory.

For example:

```
umount /remote_FS
rmdir /remote_FS
```

## 6.10.2.2 Method 2: Back Up an Individual Guest

You can back up an individual guest by backing up its specific folder under `/EXAVMIMAGES`.

The backup destination should be separate from the KVM host server, such as a writable NFS location, and be large enough to hold the backup. The space needed for an individual guest backup is approximately 200 GB.

### 1. Prepare the guest image.

Use the following script to prepare the guest image backup under `/EXAVMIMAGES/Backup`.

```
#!/bin/bash

ScriptStarttime=$(date +%s)

printf "This script is going to remove the directory /EXAVMIMAGES/Backup
if it exists. If that is not acceptable, exit the script by typing n,
manually remove /EXAVMIMAGES/Backup and come back to rerun the script.
Otherwise, press y to continue :"

read proceed

if [[${proceed} == "n"]] || [[${proceed} == "N"]]
then
 exit 0
elif [[${proceed} != "n"]] && [[${proceed} != "N"]] && [[${proceed} !=
"y"]] && [[${proceed} != "Y"]]
then
 echo "Invalid input"
 exit 1
fi

rm -rf /EXAVMIMAGES/Backup

printf "Enter the name of the KVM guest to be backed up :"

read KVMGuestName

Create the Backup Directory

if [! -d /EXAVMIMAGES/GuestImages/${KVMGuestName}]
then
 echo "Guest ${KVMGuestName} does not exist"
 exit 1
fi

mkdirStartTime=$(date +%s)

find /EXAVMIMAGES/GuestImages/${KVMGuestName} -type d |grep -v
```

```
'lost+found'|awk '{print "mkdir -p /EXAVMIMAGES/Backup"$1}'|sh

mkdir -p /EXAVMIMAGES/Backup/XML

mkdirEndTime=$(date +%s)
mkdirTime=$(expr ${mkdirEndTime} - ${mkdirStartTime})
echo "Backup Directory creation time : " ${mkdirTime}" seconds"

cp /etc/libvirt/qemu/${KVMGuestName}.xml /EXAVMIMAGES/Backup/XML

Pause the guest

PauseStartTime=$(date +%s)
virsh suspend ${KVMGuestName}
PauseEndTime=$(date +%s)
PauseTime=$(expr ${PauseEndTime} - ${PauseStartTime})
echo "PauseTime for guest - ${KVMGuestName} : " ${PauseTime}"
seconds"

Create reflinks for all the files in /EXAVMIMAGES/GuestImages/${
KVMGuestName}

relinkStartTime=$(date +%s)

find /EXAVMIMAGES/GuestImages/${KVMGuestName} -type f|awk '{print
"cp --reflink", $0,"/EXAVMIMAGES/Backup"$0}'|sh

relinkEndTime=$(date +%s)
relinkTime=$(expr ${relinkEndTime} - ${relinkStartTime})
echo "Reflink creation time for guest - ${KVMGuestName} : " $
{relinkTime}" seconds"

Unpause the guest

unPauseStartTime=$(date +%s)
virsh resume ${KVMGuestName}
unPauseEndTime=$(date +%s)
unPauseTime=$(expr ${unPauseEndTime} - ${unPauseStartTime})
echo "unPauseTime for guest - ${KVMGuestName} : " ${unPauseTime}"
seconds"

ScriptEndtime=$(date +%s)
ScriptRunTime=$(expr ${ScriptEndtime} - ${ScriptStarttime})
echo ScriptRunTime ${ScriptRunTime}" seconds"
```

## 2. Create a backup of the guest image.

Back up the reflink files under `/EXAVMIMAGES/Backup` to a remote location. The backup enables restoration of the specific guest if the KVM host is permanently damaged or lost.

For example:

```
mkdir -p /remote_FS
mount -t nfs -o rw,intr,soft,proto=tcp,nolock ip_address:/
nfs_location/ /remote_FS
```



```
cd /EXAVMIMAGES/Backup
tar --acls --xattrs --xattrs-include=* --format=pax -pjcvf /remote_FS/
exavmimage.tar.bz2 * > /tmp/exavmimage_tar.stdout 2> /tmp/
exavmimage_tar.stderr
```

In the `mount` command, *ip\_address* is the IP address of the NFS server, and *nfs\_location* is the NFS location holding the backup.

In the example, the backup file is named `exavmimage.tar.bz2`. You may choose another name that identifies the guest being backed up.

After the backup completes, check for any significant errors from the `tar` command. In the previous example, the `tar` command writes errors to the file at `/tmp/exavmimage_tar.stderr`.

3. Remove the `/EXAVMIMAGES/Backup` directory and its contents.

For example:

```
cd /
rm -rf /EXAVMIMAGES/Backup
```

4. Unmount the NFS backup location and remove the mount point directory.

For example:

```
umount /remote_FS
rmdir /remote_FS
```

### 6.10.2.3 Method 3: Back Up a Guest Internally

You can take a snapshot-based backup of a guest from inside the guest.

All steps are performed from inside the guest.

#### Note:

This backup method is performed internally within the guest and uses logical volume snapshots. Compared with other backup methods, this method provides more limited recovery options because the backup is only useful when the guest is bootable and allows `root` user login.

This procedure backs up the contents of all currently active file systems in the guest. Before starting, ensure that all of the file systems that you want to back up are mounted.

The values shown in the following steps are examples and you may need to substitute different values to match your situation.

All steps must be performed as the `root` user.

1. Prepare a destination to hold the backup.

The destination should reside outside of the local machine, such as a writable NFS location, and be large enough to hold the backup `tar` files. For non-customized partitions, the space needed for holding the backup is approximately 60 GB.

You can use the following commands to prepare a backup destination using NFS.

```
mkdir -p /root/remote_FS
mount -t nfs -o rw,intr,soft,proto=tcp,nolock ip_address:/
nfs_location/ /root/remote_FS
```

In the `mount` command, `ip_address` is the IP address of the NFS server, and `nfs_location` is the NFS location holding the backups.

2. Remove the `LVDoNotRemoveOrUse` logical volume.

The logical volume `/dev/VGExaDb/LVDoNotRemoveOrUse` is a placeholder to make sure there is always free space available to create a snapshot.

Use the following script to check for the existence of the `LVDoNotRemoveOrUse` logical volume and remove it if present.

```
lvm lvdisplay --ignorelockingfailure /dev/VGExaDb/LVDoNotRemoveOrUse
if [$? -eq 0]; then
 # LVDoNotRemoveOrUse logical volume exists.
 lvm lvremove -f /dev/VGExaDb/LVDoNotRemoveOrUse
 if [$? -ne 0]; then
 echo "Unable to remove logical volume: LVDoNotRemoveOrUse. Do
not proceed with backup."
 fi
fi
```

If the `LVDoNotRemoveOrUse` logical volume does not exist, then do not proceed with the remaining steps and determine the reason.

3. Gather information about the currently active file systems and logical volumes.

In this step, you must gather information from your guest to use later in the commands that create the logical volume snapshots and backup files.

Run the following command:

```
df -hT | grep VGExa
```

For every entry in your command output, determine the following information and create a table of values to use later:

- The volume group (VG) name and logical volume (LV) name are contained in the file system name as follows:

```
/dev/mapper/VG-name-LV-name
```

For example, in `/dev/mapper/VGExaDb-LVDbHome`, the VG name is `VGExaDb` and the LV name is `LVDbHome`.

- The backup label is a string that identifies the file system and its backup file. Use `root` for the root (`/`) file system. Otherwise, you can use a string that concatenates the directories in the mount point. For example, you can use `varlogaudit` for `/var/log/audit`.

- Define a short label, which contains 12 or fewer characters. You will use the short label to label the snapshot file system.

For example:

```
df -hT | grep VGExa
/dev/mapper/VGExaDb-LVDbSys1 xfs 15G
4.2G 11G 28% /
/dev/mapper/VGExaDb-LVDbHome xfs 4.0G
45M 4.0G 2% /home
/dev/mapper/VGExaDb-LVDbVar1 xfs 2.0G
90M 2.0G 5% /var
/dev/mapper/VGExaDb-LVDbVarLog xfs 18G
135M 18G 1% /var/log
/dev/mapper/VGExaDb-LVDbVarLogAudit xfs 1014M
89M 926M 9% /var/log/audit
/dev/mapper/VGExaDb-LVDbTmp xfs 3.0G
33M 3.0G 2% /tmp
/dev/mapper/VGExaDb-LVDbKdump xfs 20G
33M 20G 1% /crashfiles
/dev/mapper/VGExaDbDisk.u01.15.img-LVDBDisk xfs 5.0G
33M 5.0G 1% /u01
/dev/mapper/VGExaDbDisk.u02.10.img-LVDBDisk xfs 10G
33M 10G 1% /u02
/dev/mapper/VGExaDbDisk.u03.15.img-LVDBDisk xfs 15G
33M 15G 1% /u03
/dev/mapper/VGExaDbDisk.grid19.7.0.0.200414.img-LVDBDisk xfs 20G
6.0G 15G 30% /u01/app/19.0.0.0/grid
```

From the above output, you could derive the following table of information to use later in the commands that create the logical volume snapshots and backup files.

File system	VG Name	LV Name	Backup Label	Short Label
/dev/mapper/ VGExaDb- LVDbSys1	VGExaDb	LVDbSys1	root	root_snap
/dev/mapper/ VGExaDb- LVDbHome	VGExaDb	LVDbHome	home	home_snap
/dev/mapper/ VGExaDb- LVDbVar1	VGExaDb	LVDbVar1	var	var_snap
/dev/mapper/ VGExaDb- LVDbVarLog	VGExaDb	LVDbVarLog	varlog	varlog_snap
/dev/mapper/ VGExaDb- LVDbVarLogAudi t	VGExaDb	LVDbVarLogAudi t	varlogaudit	audit_snap
/dev/mapper/ VGExaDb- LVDbTmp	VGExaDb	LVDbTmp	tmp	tmp_snap

File system	VG Name	LV Name	Backup Label	Short Label
/dev/mapper/ VGExaDb- LVDbKdump	VGExaDb	LVDbKdump	crashfiles	crash_snap
/dev/mapper/ VGExaDbDisk.u0 1.5.img- LVDBDisk	VGExaDbDisk.u0 1.5.img	LVDBDisk	u01	u01_snap
/dev/mapper/ VGExaDbDisk.u0 2.10.img- LVDBDisk	VGExaDbDisk.u0 2.10.img	LVDBDisk	u02	u02_snap
/dev/mapper/ VGExaDbDisk.u0 3.15.img- LVDBDisk	VGExaDbDisk.u0 3.15.img	LVDBDisk	u03	u03_snap
/dev/mapper/ VGExaDbDisk.gr id19.7.0.0.200 414.img- LVDBDisk	VGExaDbDisk.gr id19.7.0.0.200 414.img	LVDBDisk	u01app19000gri d	grid_snap

 **Note:**

- The information gathered from your guest may be significantly different from this example. Ensure that you gather the required information directly from your guest and only use current information.
- Depending on the currently active system volume, the logical volume for the root (/) file system is `LVDbSys1` or `LVDbSys2`. Likewise, the logical volume for the `/var` file system is either `LVDbVar1` or `LVDbVar2`.

4. Create the file system snapshots and backup files.

Use the table of information for your guest, which you gathered in the previous step.

Perform the following for each row in your table, substituting the appropriate values in each command.

a. Create the snapshot.

```
lvcreate -L1G -s -n LV-Name_snap /dev/VG-Name/LV-Name
```

b. Label the snapshot.

```
xfs_admin -L Short-Label /dev/VG-Name/LV-Name_snap
```

- c. Mount the snapshot.

```
mkdir -p /root/mnt/Backup-Label
mount -o nouuid /dev/VG-Name/LV-Name_snap /root/mnt/Backup-Label
```

- d. Change to the directory for the backup.

```
cd /root/mnt/Backup-Label
```

- e. Create the backup file.

- For the root (/) file system only, use the following command to include the contents of /boot in the backup file:

```
tar --acls --xattrs --xattrs-include=* --format=pax -pjcvf /root/remote_FS/rootfs-boot.tar.bz2 * /boot > /tmp/backup_tar.stdout
2> /tmp/backup_tar.stderr
```

- Otherwise, use the following command template:

```
tar --acls --xattrs --xattrs-include=* --format=pax -pjcvf /root/remote_FS/Backup-Label.tar.bz2 * > /tmp/backup_tar.stdout 2> /tmp/backup_tar.stderr
```

- f. Check the /tmp/backup\_tar.stderr file for any significant errors.

You can ignore errors about failing to tar open sockets, and other similar errors.

- g. Unmount and remove the snapshot.

```
cd /
umount /root/mnt/Backup-Label
/bin/rmdir /root/mnt/Backup-Label
lvremove -f /dev/VG-Name/LV-Name_snap
```

5. Unmount the NFS share.

```
umount /root/remote_FS
```

6. Recreate the /dev/VGExaDb/LVDoNotRemoveOrUse logical volume.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L2G VGExaDb -y
```

## 6.11 Backing Up and Restoring Oracle Databases on KVM Guests

Backing up and restoring Oracle databases on KVM guests is the same as backing up and restoring Oracle databases on physical nodes.

- If you are using Exadata storage servers, then each Oracle Real Application Clusters (Oracle RAC) cluster running in guests requires its own Oracle Automatic Storage Management (Oracle ASM) disk group to be designated as the fast recovery area (FRA),

such as `+RECO`. Refer to the [Exadata Database Machine Backup and Restore Configuration and Operational Best Practices](#) technical reference paper for details.

- If you are using Oracle ZFS Storage Appliance, refer to the [Protecting Exadata Database Machine with the Oracle ZFS Storage Appliance: Configuration Best Practices](#) technical reference paper for details.

## 6.12 Modifying the Memory Allocated to a Guest

You can modify the memory allocated to a guest using `vm_maker`.

This operation requires a guest restart. You can let `vm_maker` restart the guest after changing the memory configuration.

1. If you are decreasing the amount of memory used by the guest, then you must first review and adjust Oracle Database memory usage and the operating system huge pages configuration settings in the guest.

- a. Review the SGA size of databases and reduce if necessary.

If you do not first reduce the memory requirements of the databases running in the guest, then the guest might fail to restart because too much memory is reserved for huge pages when the Oracle Linux operating system attempts to boot. See My Oracle Support Doc ID 361468.1 for details.

- b. Review the operating system configuration and reduce the memory allocation for huge pages if necessary.
- c. If you modify the huge pages settings in the operating system kernel configuration file (`/etc/sysctl.conf`), regenerate the `initramfs` file to reflect the system configuration change.

You should backup up the existing `initramfs` file and then regenerate it by using the `dracut` command. For example:

```
ls -l /boot/initramfs-$(uname -r).img
-rw----- 1 root root 55845440 Jan 8 10:34 /boot/
initramfs-4.14.35-2047.508.3.3.el7uek.x86_64.img

cp /boot/initramfs-$(uname -r).img backup_directory

dracut --force
```

2. Connect to the KVM host.

The remainder of this procedure is preformed inside the KVM host.

3. If you are increasing the amount of memory used by the guest, then use the following command to determine the amount of free memory available:

```
/opt/exadata_ovm/vm_maker --list --memory
```

In the output, the lowest value between `Available memory (now)` and `Available memory (delayed)` is the limit for free memory.

 **Note:**

When assigning free memory to a guest, reserve approximately 1% to 2% of free memory for storing metadata and control structures.

4. Modify the guest memory allocation and restart the guest.

For example, to modify `db01_guest01.example.com` and set a memory allocation of 32 GB, use the following command:

```
/opt/exadata_ovm/vm_maker --set --memory 32G --domain
db01_guest01.example.com --restart-domain
```

The command shuts down the guest, modifies the memory setting, and restarts the guest.

**Related Topics**

- [HugePages on Oracle Linux 64-bit \(My Oracle Support Doc ID 361468.1\)](#)

## 6.13 Modifying the Number of Virtual CPUs Allocated to a Guest

You can dynamically modify the number of virtual CPUs allocated to a guest with the `vm_maker --set --vcpu` command.

All actions to modify the number of vCPUs allocated to a guest are performed in the KVM host.

It is possible to over-commit vCPUs such that the total number of vCPUs assigned to all guests exceeds the number of physical CPUs on the system. However, over-committing CPUs should be done only when competing workloads for oversubscribed resources are well understood and concurrent demand does not exceed physical capacity.

1. Determine the number of physical CPUs.

```
/opt/exadata_ovm/vm_maker --list --vcpu --domain
db01_guest01.example.com
```

2. Modify the number of allocated vCPUs.

The number of vCPUs must be a multiple of 2.

For example, if you want to change the number of vCPUs allocated to 4 for the `db01_guest01.example.com` guest, you would use the following command:

```
/opt/exadata_ovm/vm_maker --set --vcpu 4 --domain
db01_guest01.example.com
```

## 6.14 Increasing the Disk Space in a Guest

The KVM guest local space can be extended after initial deployment by adding local disk images.

- [Adding a New LVM Disk to a Guest](#)  
You can add a new LVM disk to an Oracle Linux KVM guest to increase the amount of usable disk space in a guest.
- [Increasing the Size of the root File System](#)  
This procedure describes how to increase the size of the system partition and / (root) file system.
- [Increasing the Size of the /u01 File System](#)  
This procedure describes how to increase the size of the /u01 file system in Oracle Linux KVM.
- [Increasing the Size of the Grid Infrastructure Home or Database Home File System](#)  
You can increase the size of the Oracle Grid Infrastructure or Oracle Database home file system in a Oracle Linux KVM guest.
- [Increasing the Size of the Swap Area](#)  
You can increase the amount of swap configured in a guest.

## 6.14.1 Adding a New LVM Disk to a Guest

You can add a new LVM disk to an Oracle Linux KVM guest to increase the amount of usable disk space in a guest.

You might add an LVM disk to a guest so that the size of a file system or swap space can be increased. The system remains online while you perform this procedure.



### Note:

During this procedure you perform actions in both the KVM host and in the guest.

Run all steps in this procedure as the `root` user.

1. In the KVM host, verify that there is sufficient free disk space in `/EXAVMIMAGES`. For example:

```
df -h /EXAVMIMAGES
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbExaVMImages 1.5T 39G 1.5T 3% /
EXAVMIMAGES
```

2. In the KVM host, create a new disk image and attach it to the guest.

For example, the following command adds a guest-specific disk image named `pv2_vgexadb.img` to the guest `dm01db01vm01.example.com`:

```
/opt/exadata_ovm/vm_maker --create --disk-image /EXAVMIMAGES/
GuestImages/dm01db01vm01.example.com/pv2_vgexadb.img --attach --
domain dm01db01vm01.example.com
[INFO] Allocating an image for /EXAVMIMAGES/GuestImages/
dm01db01vm01.example.com/pv2_vgexadb.img, size 52.000000G...
[INFO] Running 'qemu-img create /EXAVMIMAGES/GuestImages/
```



```

dm01db01vm01.example.com/pv2_vgexadb.img 52.000000G '...
[INFO] Create label gpt on /EXAVMIMAGES/GuestImages/
dm01db01vm01.example.com/pv2_vgexadb.img.
[INFO] Running 'parted -a none -s /EXAVMIMAGES/GuestImages/
dm01db01vm01.example.com/pv2_vgexadb.img mklabel gpt'...
[INFO] Running 'losetup -P -f /EXAVMIMAGES/GuestImages/
dm01db01vm01.example.com/pv2_vgexadb.img'...
[INFO] Finding loop device...
[INFO] loop device is /dev/loop0
[INFO] Finding number of sectors...
[INFO] 109051904 sectors
[INFO] Finding sector size...
[INFO] 512 bytes per sector
[INFO] Creating filesystem on /dev/VGExaDbDisk.pv2_vgexadb.img/LVDBDisk...
[INFO] Running 'mkfs -t xfs -b size=4096 -f /dev/
VGExaDbDisk.pv2_vgexadb.img/LVDBDisk '...
[INFO] Checking that we have a file system on /dev/
VGExaDbDisk.pv2_vgexadb.img/LVDBDisk...
[INFO] Releasing loop device /dev/loop0...
[INFO] Removing device maps for /dev/loop0...
[INFO] Running 'kpartx -d -v /dev/loop0'...
[INFO] Removing loop device /dev/loop0...
[INFO] ##
[INFO] ## Finished .
[INFO] ##
[INFO] Created image /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/
pv2_vgexadb.img
[INFO] Running 'vgscan --cache'...
[INFO] ----- MANUAL STEPS TO BE COMPLETED FOR MOUNTING THE DISK WITHIN
DOMU dm01db01vm01.example.com -----
[INFO] 1. Check a disk with name /dev/VGExaDbDisk.pv2_vgexadb.img/
LVDBDisk exists.
[INFO] - Check for the existence of a disk named: /dev/
VGExaDbDisk.pv2_vgexadb.img/LVDBDisk. Use the 'lvdisplay' command and
check the output.
[INFO] 2. Create a mount directory for the new disk
[INFO] 3. Add the following line to /etc/fstab: /dev/
VGExaDbDisk.pv2_vgexadb.img/LVDBDisk <mount_point_from_step_2> <fstype>
defaults 1 1
[INFO] 4. Mount the new disk. Use the 'mount -a' command.
[INFO] Note: when detaching and re-attaching the same disk multiple
times, run the following command after detaching and before attaching in
the guest domain:
[INFO] 'lvm vgchange VGExaDbDisk.pv2_vgexadb.img -a -n' when re-attaching
the same disk.

```

At this time, do not perform the manual steps described at the end of the output. However, take note of the logical volume path identified in manual step number 1. In general, the logical volume path has the form: `/dev/VolumeGroupName/LogicalVolumeName`. In the example, the logical volume path is `/dev/VGExaDbDisk.pv2_vgexadb.img/LVDBDisk`.

3. On the KVM host, list the available disk images for the guest and verify the creation of the new disk image.

In the example in the previous step, the disk image file is identified as /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/pv2\_vgexadb.img. This image should now appear in the list of disk images for the guest. For example:

```
/opt/exadata_ovm/vm_maker --list --disk-image --domain
dm01db01vm01.example.com
File /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/System.img
File /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/
grid19.2.0.0.0.img
File /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/
db19.2.0.0.0-3.img
File /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/
pv1_vgexadb.img
File /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/
pv2_vgexadb.img
```

4. On the guest, identify the newly added disk.

Use the `lvdisplay` command along with the logical volume path noted earlier.

```
lvdisplay /dev/VGExaDbDisk.pv2_vgexadb.img/LVDBDisk
--- Logical volume ---
LV Path /dev/VGExaDbDisk.pv2_vgexadb.img/LVDBDisk
LV Name LVDBDisk
VG Name VGExaDbDisk.pv2_vgexadb.img
LV UUID ePC0Qe-PfOX-oCoP-Pd5n-2nDj-z0KU-c9IygG
LV Write Access read/write
LV Creation host, time dm01db01vm01.example.com, 2022-01-10
03:06:18 -0800
LV Status available
open 0
LV Size 50.00 GiB
Current LE 12800
Segments 1
Allocation inherit
Read ahead sectors auto
- currently set to 256
Block device 252:11
```

5. In the guest, remove the logical volume and volume group that were created for the added disk.

You must perform this step in order to use the newly created disk to extend an existing volume group.

a. Remove the logical volume.

In this example, the logical volume path is /dev/VGExaDbDisk.pv2\_vgexadb.img/LVDBDisk.

```
lvremove /dev/VGExaDbDisk.pv2_vgexadb.img/LVDBDisk
Do you really want to remove active logical volume
VGExaDbDisk.pv2_vgexadb.img/LVDBDisk? [y/n]: y
Logical volume "LVDBDisk" successfully removed
```

b. Remove the volume group that came with the logical volume.

In this example, the volume group name is `VGExaDbDisk.pv2_vgexadb.img`.

```
vgremove VGExaDbDisk.pv2_vgexadb.img
Volume group "VGExaDbDisk.pv2_vgexadb.img" successfully removed
```

At this point, all that is left is the physical volume with no logical volume and no volume group.

6. In the guest, identify the physical volume device for the newly added disk.

Use the `pvdisplay` command and look for the new physical volume.

In the following example, the output is truncated to highlight the new physical volume:

```
pvdisplay
...
"/dev/sdf1" is a new physical volume of "<50.00 GiB"
--- NEW Physical volume ---
PV Name /dev/sdf1
VG Name
PV Size <50.00 GiB
Allocatable NO
PE Size 0
Total PE 0
Free PE 0
Allocated PE 0
PV UUID tfb81M-eHe9-SPch-8UAu-pkHe-dAYx-ez3Sru
...
```

7. In the guest, use the new physical volume to extend an existing volume group.

In the following example, the new physical volume (`/dev/sdf1`) is used to extend the volume group `VGExaDb`. The `vgdisplay` output is truncated to highlight `VGExaDb`.

```
vgdisplay -s
...
"VGExaDb" 88.00 GiB [88.00 GiB used / 0 free]
...

vgextend VGExaDb /dev/sdf1
Volume group "VGExaDb" successfully extended

vgdisplay -s
...
"VGExaDb" <139.24 GiB [88.00 GiB used / <51.24 GiB free]
...
```

To increase the size of various file systems, using the additional space added to the volume group by this procedure, refer to the following topics:

- [Increasing the Size of the root File System](#)
- [Increasing the Size of the /u01 File System](#)
- [Increasing the Size of the Grid Infrastructure Home or Database Home File System](#)
- [Increasing the Size of the Swap Area](#)

## 6.14.2 Increasing the Size of the root File System

This procedure describes how to increase the size of the system partition and / (root) file system.

This procedure is performed while the file system remains online.

### Note:

There are two system partitions, `LVDbSys1` and `LVDbSys2`. One partition is active and mounted. The other partition is inactive and used as a backup location during upgrade. The size of both system partitions must be equal.

Keep at least 1 GB of free space in the `VGExaDb` volume group. The free space is used for the LVM snapshot created by the `dbnodeupdate.sh` utility during software maintenance. If you make snapshot-based backups of the / (root) and `/u01` directories as described in [Creating a Snapshot-Based Backup of Oracle Linux Database Server](#), then keep at least 6 GB of free space in the `VGExaDb` volume group.

This task assumes that additional disk space is available to be used. If that is not the case, then complete the task [Adding a New LVM Disk to a Guest](#) before starting this procedure.

1. Collect information about the current environment.
  - a. Use the `df` command to identify the amount of free and used space in the root partition (/)

```
df -h /
```

The following is an example of the output from the command:

```
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbSys1 12G 5.1G 6.2G 46% /
```

### Note:

The active root partition may be either `LVDbSys1` or `LVDbSys2`, depending on previous maintenance activities.

- b. Use the `lvs` command to display the current logical volume configuration.

```
lvs -o lv_name,lv_path,vg_name,lv_size
```

The following is an example of the output from the command:

```
LV Path VG LSize
LVDbOra1 /dev/VGExaDb/LVDbOra1 VGExaDb 10.00g
```

```
LVDbSwap1 /dev/VGExaDb/LVDbSwap1 VGExaDb 8.00g
LVDbSys1 /dev/VGExaDb/LVDbSys1 VGExaDb 12.00g
LVDbSys2 /dev/VGExaDb/LVDbSys2 VGExaDb 12.00g
```

2. Verify there is available space in the volume group `VGExaDb` using the `vgdisplay` command.

```
vgdisplay VGExaDb -s
```

The following is an example of the output from the command:

```
"VGExaDb" 53.49 GiB [42.00 GiB used / 11.49 GiB free]
```

The volume group must contain enough free space to increase the size of both system partitions, and maintain at least 1 GB of free space for the LVM snapshot created by the `dbnodeupdate.sh` utility during upgrade. If there is not sufficient free space in the volume group, then add a new disk to LVM.

3. Resize both `LVDbSys1` and `LVDbSys2` logical volumes using the `lvextend` command.

```
lvextend -L +size /dev/VGExaDb/LVDbSys1
lvextend -L +size /dev/VGExaDb/LVDbSys2
```

In the preceding command, *size* is the amount of space to be added to the logical volume. The amount of space added to each system partition must be the same.

The following example extends the logical volumes by 10 GB:

```
lvextend -L +10G /dev/VGExaDb/LVDbSys1
lvextend -L +10G /dev/VGExaDb/LVDbSys2
```

4. Resize the partition using the `xfs_growfs` command.

```
xfs_growfs /
```

5. Verify the space was extended for the active system partition using the `df` command.

```
df -h /
```

### 6.14.3 Increasing the Size of the `/u01` File System

This procedure describes how to increase the size of the `/u01` file system in Oracle Linux KVM.

This procedure is performed while the file system remains online.

 **Note:**

Keep at least 1 GB of free space in the `VGExaDb` volume group. The free space is used for the LVM snapshot created by the `dbnodeupdate.sh` utility during software maintenance. If you make snapshot-based backups of the `/` (root) and `/u01` directories as described in [Creating a Snapshot-Based Backup of Oracle Linux Database Server](#), then keep at least 6 GB of free space in the `VGExaDb` volume group

This task assumes that additional disk space is available to be used. If that is not the case, then complete the task [Adding a New LVM Disk to a Guest](#) before starting this procedure.

1. Collect information about the current environment.
  - a. Use the `df` command to identify the amount of free and used space in the `/u01` partition.

```
df -h /u01
```

The following is an example of the output from the command:

```
Filesystem Size Used Avail Use% Mounted on
/dev/mapper/VGExaDb-LVDbOra1
 9.9G 1.7G 7.8G 18% /u01
```

- b. Use the `lvs` command to display the current logical volume configuration used by the `/u01` file system.

```
lvs -o lv_name,lv_path,vg_name,lv_size /dev/VGExaDb/LVDbOra1
```

The following is an example of the output from the command:

```
LV Path VG LSize
LVDbOra1 /dev/VGExaDb/LVDbOra1 VGExaDb 10.00g
```

2. Verify there is available space in the volume group `VGExaDb` using the `vgdisplay` command.

```
vgdisplay VGExaDb -s
```

The following is an example of the output from the command:

```
"VGExaDb" 53.49 GiB [42.00 GiB used / 11.49 GiB free]
```

If the output shows there is less than 1 GB of free space, then neither the logical volume nor file system should be extended. Maintain at least 1 GB of free space in the `VGExaDb` volume group for the LVM snapshot created by the `dbnodeupdate.sh` utility during an upgrade. If there is not sufficient free space in the volume group, then add a new disk to LVM.

3. Resize the logical volume using the `lvextend` command.

```
lvextend -L +sizeG /dev/VGExaDb/LVDbOral
```

In the preceding command, *size* is the amount of space to be added to the logical volume.

The following example extends the logical volume by 10 GB:

```
lvextend -L +10G /dev/VGExaDb/LVDbOral
```

4. Resize the partition using the `xfs_growfs` command.

```
xfs_growfs /u01
```

5. Verify the space was extended using the `df` command.

```
df -h /u01
```

#### Related Topics

- [Adding a New LVM Disk to a User Domain](#)

## 6.14.4 Increasing the Size of the Grid Infrastructure Home or Database Home File System

You can increase the size of the Oracle Grid Infrastructure or Oracle Database home file system in a Oracle Linux KVM guest.

The Oracle Grid Infrastructure software home and the Oracle Database software home are created as separate disk image files in the KVM host. The disk image files are located in the `/EXAVMIMAGES/GuestImages/DomainName/` directory. The disk image files are attached to the guest automatically during virtual machine startup, and mounted as separate, non-LVM file systems in the guest.

This task assumes that additional disk space is available to be used.

1. Connect to the guest, and check the file system size using the `df` command. For example:

```
df -h
Filesystem
 Size Used Avail Use% Mounted on
...
/dev/mapper/VGExaDbDisk.grid--klone--Linux--x86--64--190000.50.img-
LVDBDisk
 50G 5.9G 45G 12% /u01/app/19.0.0.0/grid
/dev/mapper/VGExaDbDisk.db--klone--Linux--x86--64--190000.50.img-LVDBDisk
 50G 6.5G 44G 13% /u01/app/oracle/product/19.0.0.0/DbHome_3
...
```

2. Connect to the KVM host, and then shut down the guest.

In the example, the guest name is `dm01db01vm01.example.com`.

```
/opt/exadata_ovm/vm_maker --stop-domain dm01db01vm01.example.com
```

3. While still connected to the KVM host, create a reflink copy to serve as a backup of the disk image that will be increased.

The disk image is located under `/EXAVMIMAGES/GuestImages/GuestName`. The image file name contains `grid` for an Oracle Grid Infrastructure image or `db` for an Oracle Database image. For example:

```
cd /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com
```

```
cp --reflink db19.6.0.0.200114-3.img pre_resize.db19.0.0.img
```

4. In the KVM host, create a new disk image and attach it to the guest.

For example, the following command adds the disk image `db03.img` to the guest `dm01db01vm01.example.com`:

```
/opt/exadata_ovm/vm_maker --create --disk-image /EXAVMIMAGES/
db03.img --attach
--domain dm01db01vm01.example.com
[INFO] Allocating an image for /EXAVMIMAGES/db03.img, size
50.000000G...
[INFO] Running 'qemu-img create /EXAVMIMAGES/db03.img 50.000000G
'...
[INFO] Create label gpt on /EXAVMIMAGES/db03.img.
[INFO] Running 'parted -a none -s /EXAVMIMAGES/db03.img mklabel
gpt'...
[INFO] Running 'losetup -P -f /EXAVMIMAGES/rk02.img'...
[INFO] Finding loop device...
[INFO] loop device is /dev/loop0
[INFO] Finding number of sectors...
[INFO] 104857600 sectors
[INFO] Releasing loop device /dev/loop0...
[INFO] Removing device maps for /dev/loop0...
[INFO] Running 'kpartx -d -v /dev/loop0'...
[INFO] Removing loop device /dev/loop0...
[INFO] ##
[INFO] ## Finished .
[INFO] ##
[INFO] Created image /EXAVMIMAGES/db03.img
[INFO] File /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/
db03.img is a reflink from
/EXAVMIMAGES/db03.img and added as disk to domain
dm01db01vm01.example.com
[INFO] ----- MANUAL STEPS TO BE COMPLETED FOR MOUNTING THE DISK
WITHIN DOMU dm01db01vm01
.example.com -----
[INFO] 1. Check a disk with name /dev/VGExaDbDisk.db03.img/LVDBDisk
exists.
[INFO] - Run the command 'lvsdisplay' to verify a disk with name
'/dev/VGExaDbDisk.db03.img/
LVDBDisk' exists.
```



```
[INFO] 2. Create a directory that will to be used for mounting the new
disk.
[INFO] 3. Add the following line to /etc/fstab: /dev/VGExaDbDisk.db03.img/
LVDBDisk <mount_
point_from_step_2> <fstype> defaults 1 1
[INFO] 4. Mount the newly added disk to mount point through the command:
mount -a.
```

Do not perform the manual steps described in the output. However, take note of the logical volume path identified in manual step number 1.

In general, the logical volume path has the form: `/dev/VolumeGroupName/LogicalVolumeName`.

In the example, the logical volume path is `/dev/VGExaDbDisk.db03.img/LVDBDisk`.

5. Restart the guest.

For example:

```
/opt/exadata_ovm/vm_maker --start-domain dm01db01vm01.example.com --
console
```

6. In the guest, confirm the newly added disk device.

Use the `lvdisplay` command along with the logical volume path noted earlier.

```
lvdisplay /dev/VGExaDbDisk.db03.img/LVDBDisk
LV Path /dev/VGExaDbDisk.db03.img/LVDBDisk
LV Name LVDBDisk
VG Name VGExaDbDisk.db03.img
LV UUID u3RBKF-UmCK-JQxc-iFf5-6WqS-GWAw-3nLjdn
LV Write Access read/write
LV Creation host, time dm01db01vm01.example.com, 2019-10-28 04:11:28
-0700
LV Status available
open 0
LV Size <50.00 GiB
Current LE 12799
Segments 1
Allocation inherit
Read ahead sectors auto
- currently set to 256
Block device 252:14
```

7. In the guest, remove the logical volume and volume group that were created for the added disk.

You must perform this step in order to use the newly created disk to extend an existing volume group.

a. Remove the logical volume.

In this example, the logical volume path is `/dev/VGExaDbDisk.db03.img/LVDBDisk`.

```
lvremove /dev/VGExaDbDisk.db03.img/LVDBDisk
Do you really want to remove active logical volume
```

```
VGExaDbDisk.db03.img/LVDBDisk? [y/n]: y
Logical volume "LVDBDisk" successfully removed
```

- b.** Remove the volume group that came with the logical volume.

In this example, the volume group name is `VGExaDbDisk.db03.img`.

```
vgremove VGExaDbDisk.db03.img
Volume group "VGExaDbDisk.db03.img" successfully removed
```

At this point, all that is left is the physical volume with no logical volume and no volume group.

- 8.** In the guest, identify the physical volume device for the newly added disk.

The physical volume identifies itself as a `NEW Physical volume` in `pvdisk` output. For example:

```
pvdisk
...
"/dev/sdf4" is a new physical volume of "<50.00 GiB"
--- NEW Physical volume ---
PV Name /dev/sdf4
VG Name
PV Size <50.00 GiB
Allocatable NO
PE Size 0
Total PE 0
Free PE 0
Allocated PE 0
PV UUID tfb81M-eHe9-SPch-8UAu-pkHe-dAYx-ru3Sez
...
```

- 9.** In the guest, identify the volume group for the file system that you want to extend.

Use the `vgdisplay` command. The volume group name contains `grid` for Oracle Grid Infrastructure or `db` for Oracle Database. For example:

```
vgdisplay -s
...
"VGExaDbDisk.grid-klone-Linux-x86-64-190000.50.img" <50.00 GiB
[<50.00 GiB used / 0 free]
"VGExaDbDisk.db-klone-Linux-x86-64-190000.50.img" <50.00 GiB
[<50.00 GiB used / 0 free]
...
```

- 10.** In the guest, extend the volume group, then verify the additional space in the volume group.

Use the `vgextend` command and specify the volume group name and physical volume device that you identified previously. For example:

```
vgextend VGExaDbDisk.db-klone-Linux-x86-64-190000.50.img /dev/sdf4
Volume group "VGExaDbDisk.db-klone-Linux-x86-64-190000.50.img"
successfully extended
```

Use the `vgdisplay` command to verify that the volume group now contains some free space. For example:

```
vgdisplay -s
...
"VGExaDbDisk.grid-klone-Linux-x86-64-190000.50.img" <50.00 GiB [<50.00
GiB used / 0 free]
"VGExaDbDisk.db-klone-Linux-x86-64-190000.50.img" <101.24 GiB [<50.00
GiB used / <51.24 GiB free]
...
```

11. In the guest, resize the logical volume using the following `lvextend` command:

```
lvextend -L +sizeG LogicalVolumePath
```

The following example extends the logical volume by 10 GB:

```
lvextend -L +10G /dev/VGExaDbDisk.db-klone-Linux-x86-64-190000.50.img/
LVDBDisk
```

12. In the guest, resize the file system partition using the `xfs_growfs` command.

```
xfs_growfs /u01/app/oracle/product/19.0.0.0/DbHome_3
```

13. In the guest, verify the file system size was increased. For example:

```
df -h
Filesystem
 Size Used Avail Use% Mounted on
...
/dev/mapper/VGExaDbDisk.db--klone--Linux--x86--64--190000.50.img-LVDBDisk
 60G 6.5G 53G 10% /u01/app/oracle/product/19.0.0.0/DbHome_3
...
```

14. Connect to the KVM host, and remove the backup image.

Use a command similar to the following where `pre_resize.db19.0.0.img` is the name of the backup image file created in step 3:

```
cd /EXAVMIMAGES/GuestImages/DomainName
rm pre_resize.db19.0.0.img
```

## 6.14.5 Increasing the Size of the Swap Area

You can increase the amount of swap configured in a guest.

1. In the KVM host, create a new disk image and attach it to the guest.

For example, the following command adds the disk image `swap2.img` to the guest `dm01db01vm01.example.com`:

```
/opt/exadata_ovm/vm_maker --create --disk-image /EXAVMIMAGES/swap2.img
--attach --domain dm01db01vm01.example.com
```

```
[INFO] Allocating an image for /EXAVMIMAGES/swap2.img, size
50.000000G...
[INFO] Running 'qemu-img create /EXAVMIMAGES/swap2.img 50.000000G
'...
[INFO] Create label gpt on /EXAVMIMAGES/swap2.img.
[INFO] Running 'parted -a none -s /EXAVMIMAGES/swap2.img mklabel
gpt'...
[INFO] Running 'losetup -P -f /EXAVMIMAGES/swap2.img'...
[INFO] Finding loop device...
[INFO] loop device is /dev/loop0
[INFO] Finding number of sectors...
[INFO] 104857600 sectors
[INFO] Releasing loop device /dev/loop0...
[INFO] Removing device maps for /dev/loop0...
[INFO] Running 'kpartx -d -v /dev/loop0'...
[INFO] Removing loop device /dev/loop0...
[INFO] ##
[INFO] ## Finished .
[INFO] ##
[INFO] Created image /EXAVMIMAGES/swap2.img
[INFO] File /EXAVMIMAGES/GuestImages/dm01db01vm01.example.com/
swap2.img is a reflink from
/EXAVMIMAGES/swap2.img and added as disk to domain
dm01db01vm01.example.com
[INFO] ----- MANUAL STEPS TO BE COMPLETED FOR MOUNTING THE DISK
WITHIN DOMU dm01db01vm01
.example.com -----
[INFO] 1. Check a disk with name /dev/VGExaDbDisk.swap2.img/
LVDBDisk exists.
[INFO] - Run the command 'lvdisplay' to verify a disk with name
'/dev/VGExaDbDisk.swap2.img/
LVDBDisk' exists.
[INFO] 2. Create a directory that will to be used for mounting the
new disk.
[INFO] 3. Add the following line to /etc/fstab: /dev/
VGExaDbDisk.swap2.img/LVDBDisk <mount_
point_from_step_2> <fstype> defaults 1 1
[INFO] 4. Mount the newly added disk to mount point through the
command: mount -a.
```

Do not perform the manual steps described in the output. However, take note of the logical volume path identified in manual step number 1.

In general, the logical volume path has the form: `/dev/VolumeGroupName/LogicalVolumeName`.

In the example, the logical volume path is `/dev/VGExaDbDisk.swap2.img/LVDBDisk`.

2. In the guest, configure the new logical volume as a swap device.

Use the `mkswap` command, and configure the new logical volume with a unique label, which is not currently in use in the `/etc/fstab` file.

In the following example, the swap device label is `SWAP2` and the logical volume path is `/dev/VGExaDbDisk.swap2.img/LVDBDisk`.

```
mkswap -L SWAP2 /dev/VGExaDbDisk.swap2.img/LVDBDisk
```

3. In the guest, enable the new swap device.

Use the `swapon` command with the `-L` option and specify the label of the newly created swap device.

For example:

```
swapon -L SWAP2
```

4. In the guest, verify that the new swap device is enabled by using the `swapon -s` command.

For example:

```
swapon -s
```

Filename	Type	Size	Used	Priority
/dev/dm-3	partition	8388604	306108	-1
/dev/VGExaDb/LVDbSwap2	partition	8388604	0	-2

5. In the guest, edit the `/etc/fstab` file to include the new swap device.

You can copy the existing swap entry, and then change the `LABEL` value in the new entry to the label used to create the new swap device.

In the following example, the new swap device is added to the `/etc/fstab` file using `LABEL=SWAP2`.

```
cat /etc/fstab
```

LABEL=DBSYS	/		ext4
defaults	1 1		
LABEL=BOOT	/boot		ext4
defaults,nodev	1 1		
tmpfs	/dev/shm		tmpfs
defaults,size=7998m	0		
devpts	/dev/pts		devpts
gid=5,mode=620	0 0		
sysfs	/sys		sysfs
defaults	0 0		
proc	/proc		proc
defaults	0 0		
LABEL=SWAP	swap		swap
defaults	0 0		
LABEL=SWAP2	swap		swap
defaults	0 0		
LABEL=DBORA	/u01		ext4
defaults	1 1		
/dev/xvdb	/u01/app/12.1.0.2/grid		ext4
defaults	1 1		
/dev/xvdc	/u01/app/oracle/product/12.1.0.2/dbhome_1		ext4
defaults	1 1		

## 6.15 Expanding /EXAVMIMAGES on the KVM host

Use this procedure to expand /EXAVMIMAGES using available space on the KVM host.

On Exadata database servers, local disk storage is governed by a volume manager, with all of the available storage space allocated to a single volume group. In turn, the volume group contains numerous logical volumes that support various file systems. On KVM hosts, most of the space is allocated to /EXAVMIMAGES, which is used for guest storage.

Typically, a modest amount of free space is preserved in the volume group so that a file system can be easily extended if required. Additional space is also available by adding the disk expansion kit to a database server. The kit consists of 4 hard drives, which are installed in the unused slots in the database server.

### Note:

The disk expansion kit is supported on 2-socket Oracle Exadata Database Machine systems only.

If you installed the disk expansion kit, ensure that you have completed the procedure outlined in [Adding the Disk Expansion Kit to Database Servers: X8M-2 and Prior](#) before you proceed with this procedure:

1. Examine the volume group to confirm the available free space.

```
vgs
VG #PV #LV #SN Attr VSize VFree
VGExaDb 1 11 0 wz--n- 3.27t <1.73t
```

2. Confirm the current space allocation for /EXAVMIMAGES.

```
df /EXAVMIMAGES
Filesystem 1K-blocks Used
Available Use% Mounted on
/dev/mapper/VGExaDb-LVDbExaVMImages 1572096000 250734224
1321361776 16% /EXAVMIMAGES
```

3. Extend the logical volume associated with /EXAVMIMAGES.

Use the `lvextend` command to add space to the logical volume.

In the following example, all of the available free space is added to the logical volume.

```
lvextend -l +100%FREE /dev/VGExaDb/LVDbExaVMImages
Size of logical volume VGExaDb/LVDbExaVMImages changed from 1.46
TiB (384000 extents) to 3.19 TiB (837430 extents).
Logical volume VGExaDb/LVDbExaVMImages successfully resized.
```

If you want to retain some free space for future use, then you can use a subset of the available free space. For example, the following command uses 90% of the available free space:

```
lvextend -l +90%FREE /dev/VGExaDb/LVDbExaVMImages
```

Or, you can specify the amount of space that you want to add. For example, the following command expands the logical volume by 500 GB:

```
lvextend -l +500G /dev/VGExaDb/LVDbExaVMImages
```

#### 4. Extend the file system associated with /EXAVMIMAGES.

Use the `xfs_growfs` command to extend the file system into the expanded logical volume.

```
xfs_growfs /EXAVMIMAGES
meta-data=/dev/mapper/VGExaDb-LVDbExaVMImages isize=512 agcount=32,
agsize=12288000 blks
 = sectsz=512 attr=2, projid32bit=1
 = crc=1 finobt=0 spinodes=0 rmapbt=0
 = reflink=1 data
 = bsize=4096 blocks=393216000,
imaxpct=5
 = sunit=256 swidth=256 blks naming
 =version 2 bsize=4096 ascii-ci=0 ftype=1 log
 =internal bsize=4096 blocks=192000, version=2
 = sectsz=512 sunit=8 blks, lazy-count=1
realtime
 =none extsz=4096 blocks=0, rtextents=0 data
blocks changed from 393216000 to 857528320
```

#### 5. Confirm the expansion of /EXAVMIMAGES.

```
df /EXAVMIMAGES
Filesystem 1K-blocks Used Available
Use% Mounted on
/dev/mapper/VGExaDb-LVDbExaVMImages 3429345280 261835784
3167509496 8% /EXAVMIMAGE
```

### Related Topics

- [Adding the Disk Expansion Kit to Database Servers: X8M-2 and Prior](#)

## 6.16 Adding an Oracle Linux KVM Cluster

You can use Oracle Exadata Deployment Assistant (OEDA) to create a new Oracle Linux KVM cluster on an existing Exadata Database Machine.

See [Adding a VM Cluster to Oracle Exadata Database Machine Using OEDA](#).

## 6.17 Expanding an Oracle RAC Cluster in Oracle Linux KVM Using OEDACLI

You can expand an existing Oracle RAC cluster on Oracle Linux KVM by adding guests using the Oracle Exadata Deployment Assistant command-line interface (OEDACLI).

OEDACLI is the preferred method if you have a known, good version of the OEDA XML file for your cluster.

 **Note:**

During the execution of this procedure, the existing Oracle RAC cluster nodes along with their database instances incur zero downtime.

Use cases for this procedure include:

- You have an existing Oracle RAC cluster that uses only a subset of the database servers of an Oracle Exadata Rack, and now the nodes not being used by the cluster have become candidates for use.
- You have an existing Oracle RAC cluster on Exadata Database Machine that was recently extended with additional database servers.
- You have an existing Oracle RAC cluster that had a complete node failure and the node was removed and replaced with a newly re-imaged node.

Before performing the steps in this section, the new database servers should have been set up as detailed in [Adding a New Database Server to the Cluster](#), including the following:

- The new database server is installed and configured on the network with a KVM host.
- Download the latest Oracle Exadata Deployment Assistant (OEDA); ensure the version you download is the July 2019 release, or later.
- You have an OEDA configuration XML file that accurately reflects the existing cluster configuration. You can validate the XML file by generating an installation template from it and comparing it to the current configuration. See the OEDACLI command SAVE FILES.
- Review the OEDA Installation Template report for the current system configuration to obtain node names and IP addresses for existing nodes. You will need to have new host names and IP addresses for the new nodes being added. The new host names and IP addresses required are:
  - Administration host names and IP addresses (referred to as ADMINNET) for the KVM host and the guests.
  - Private host names and IP addresses (referred to as PRIVNET) for the KVM host and the guests.
  - Integrated Lights Out Manager (ILOM) host names and IP addresses for the KVM host.



- Client host names and IP addresses (referred to as CLIENTNET) for the guests.
- Virtual IP (VIP) host names and IP addresses (referred to as VIPNET) for the guests.
- Physical rack number and location of the new node in the rack (in terms of U number)
- Each KVM host has been imaged or patched to the same image in use on the existing database servers. The current system image must match the version of the /EXAVMIMAGES/System.first.boot\*.img file on the new KVM host node.

 **Note:**

The ~/dom0\_group file referenced below is a text file that contains the host names of the KVM hosts for all existing and new nodes being added.

Check that the image version across all KVM hosts are the same.

```
dcli -g ~/dom0_group -l root "imageinfo -ver"

exa01adm01: 19.2.0.0.0.190225
exa01adm02: 19.2.0.0.0.190225
exa01adm03: 19.2.0.0.0.190225
```

If any image versions differ, you must upgrade the nodes as needed so that they match.

Ensure that the System.first.boot version across all KVM hosts matches the image version retrieved in the previous step.

```
dcli -g ~/dom0_group -l root "ls -l /EXAVMIMAGES/System.first.boot*.img"
exa01adm01: /EXAVMIMAGES/System.first.boot.19.2.0.0.0.190225.img
exa01adm02: /EXAVMIMAGES/System.first.boot.19.2.0.0.0.190225.img
exa01adm03: /EXAVMIMAGES/System.first.boot.19.2.0.0.0.190225.img
```

If **any** nodes are missing the System.first.boot.img file that corresponds to the current image, then obtain the required file. See the “Supplemental README note” for your Exadata release in My Oracle Support Doc ID 888828.1 and look for the patch file corresponding to this description, “DomU System.img OS image for V.V.0.0.0 VM creation on upgraded dom0s”

- Place the kclone.zip files (gi-kclone\*.zip and db-kclone\*.zip) in the /EXAVMIMAGES location on the freshly imaged KVM host node you are adding to the cluster. These files can be found in the /EXAVMIMAGES directory on the KVM host node from where the system was initially deployed.

The steps here show how to add a new KVM host node called exa01adm03 that will have a new guest called exa01adm03vm01. The steps show how to extend an existing Oracle RAC cluster onto the guest using OEDACLI commands. The existing cluster has KVM host nodes named exa01adm01 and exa01adm02 and guest nodes named exa01adm01vm01 and exa01adm02vm01.

1. Add the KVM host information to the OEDA XML file using the CLONE COMPUTE command.

In the examples below, the OEDA XML file is assumed to be in:  
*unzipped\_OEDA\_location/ExadataConfigurations.*

```
OEDACLI> LOAD FILE NAME=exa01_original_deployment.xml

OEDACLI> CLONE COMPUTE SRCNAME=exa01adm01 TGTNAME=exa01adm03
OEDACLI> SET ADMINNET NAME=exa01adm03,IP=xx.xx.xx.xx
OEDACLI> SET PRIVNET NAME1=exa01adm03-priv1,IP1=xx.xx.xx.xx,NAME2=exa01adm03-priv2,IP2=xx.xx.xx.xx
OEDACLI> SET ILOMNET NAME=exa01adm03-c,IP=xx.xx.xx.xx
OEDACLI> SET RACK NUM=MN,ULOC=XX

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS FORCE
OEDACLI> SAVE FILE NAME=exa01_plus_adm03_node.xml
```

At this point we have a new XML file that has the new compute node KVM host in the configuration. This file will be used by the subsequent steps.

2. Add the new guest information to the OEDA XML file using the `CLONE GUEST` command and deploy the guest.

```
OEDACLI> LOAD FILE NAME=exa01_plus_adm03_node.xml

OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01
WHERE STEPNAME=CREATE_GUEST
OEDACLI> SET PARENT NAME=exa01adm03
OEDACLI> SET ADMINNET NAME=exa01adm03vm01,IP=xx.xx.xx.xx
OEDACLI> SET PRIVNET NAME1=exa01db03vm01-priv1,IP1=xx.xx.xx.xx,NAME2=exa01db03vm01-priv2,IP2=xx.xx.xx.xx
OEDACLI> SET CLIENTNET NAME=exa01client03vm01,IP=xx.xx.xx.xx
OEDACLI> SET VIPNET NAME=exa01client03vm01-vip,IP=xx.xx.xx.xx

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS
```

If you prefer that OEDACLI runs all steps automatically, omit the `WHERE STEPNAME=CREATE_GUEST` clause from the `CLONE GUEST` command and skip step 3 below.

3. Use OEDACLI to extend the cluster to the new guest.

 **Note:**

Continue using the same XML file, `exa01_plus_adm03_node.xml` in this example. You will continue to update this file as you proceed through these steps. At the very end of the procedure, this XML file will properly reflect the new state of the clusters.

```
OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01
WHERE STEPNAME=CREATE_USERS
```

```

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS

OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=CELL_CONNECTIVITY

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS

OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=ADD_NODE

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS

OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=EXTEND_DBHOME

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS

OEDACLI> CLONE GUEST SRCNAME=exa01adm01vm01 TGTNAME=exa01adm03vm01 WHERE
STEPNAME=ADD_INSTANCE

OEDACLI> SAVE ACTION
OEDACLI> MERGE ACTIONS
OEDACLI> DEPLOY ACTIONS

```

OEDACLI prints out messages similar to the following as each step completes:

```

Deploying Action ID : 39 CLONE GUEST SRCNAME=exa01adm01vm01
TGTNAME=exa01adm03vm01 where STEPNAME=ADD_INSTANCE
Deploying CLONE GUEST
Cloning Guest
Cloning Guest : exa01adm03vm01.example.com_id
Adding new instance for database [dbm] on exa01adm03vm01.example.com
Setting up Huge Pages for Database..[dbm]
Adding instance dbm3 on host exa01adm03vm01.example.com
Successfully completed adding database instance on the new node [elapsed
Time [Elapsed =
249561 mS [4.0 minutes] Fri Jun 28 13:35:52 PDT 2019]]
Done...
Done

```

4. Save the current state of the configuration and generate configuration information.

```
OEDACLI> SAVE FILES LOCATION=/tmp/exa01_plus_adm03_config
```

The above command writes all the configuration files to the directory `/tmp/exa01_plus_adm03_config`. Save a copy of these files in a safe place since they now reflect the changes made to your cluster.

5. Gather an Oracle EXAchk report and examine it to ensure the cluster is in good health.

#### Related Topics

- OEDA Command Line Interface
- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

## 6.18 Moving a Guest to a Different Database Server

Guests can move to different database servers.

The target Exadata Database Machine database server must meet the following requirements:

- The target database server must have the same Oracle Exadata System Software release installed with Oracle Linux KVM.
- The target database server must have the same network visibility.
- The target database server must have access to the same Exadata Database Machine storage servers.
- The target database server must have sufficient free resources (CPU, memory, and local disk storage) to operate the guest.
  - It is possible to over-commit virtual CPUs such that the total number of virtual CPUs assigned to all domains exceeds the number of physical CPUs on the system. Over-committing CPUs can be done only when the competing workloads for over-subscribed resources are well understood and the concurrent demand does not exceed physical capacity.
  - It is not possible to over-commit memory.
  - Copying disk images to the target database server may increase space allocation of the disk image files because the copied files are no longer able to benefit from the disk space savings gained by using reflinks.
- The guest name must not be already in use on the target database server.

The following procedure moves a guest to a new database server in the same Oracle Exadata System Software configuration.

1. In the KVM host, shut down the guest that is being moved.

```
vm_maker --stop-domain GuestName
```

2. Copy the guest disk image and configuration files to the target database server.

In the following examples, replace *GuestName* with the name of the guest, and replace *target* with the host name of the target KVM host.

```
scp -r /EXAVMIMAGES/GuestImages/GuestName/ target:/EXAVMIMAGES/GuestImages
```

3. Copy the guest XML definition to the target database server.

```
scp /etc/libvirt/qemu/GuestName.xml target:/EXAVMIMAGES/GuestImages
```

4. In the target KVM host, define the domain.

```
virsh define /EXAVMIMAGES/GuestImages/GuestName.xml
```

5. If you are using Oracle Exadata System Software release 20.1 or later, run the following `vm_maker` command in the target KVM host to complete the guest migration.

```
vm_maker --update-mac GuestName
```

 **Note:**

- The `vm_maker --update-mac` command is first introduced in Oracle Exadata System Software release 20.1.4 (November 2020). If you are using an earlier 20.1 release, you must perform an update to get this command.
- This step is not required on systems using a release of Oracle Exadata System Software prior to 20.1.

6. Start the migrated guest on the target KVM host.

```
vm_maker --start-domain GuestName
```

## 6.19 Recovering a KVM Deployment

A KVM host can be recovered from backups and guests can be recovered from snapshot backups.

A KVM host can be recovered from a snapshot-based backup when severe disaster conditions damage the Oracle KVM host, or when the server hardware is replaced to such an extent that it amounts to new hardware.

For example, replacing all hard disks leaves no trace of original software on the system. This is similar to replacing the complete system as far as the software is concerned.

The recovery procedures described in this section do not include backup or recovery of Exadata storage servers or the data in an Oracle Database. Oracle recommends testing the backup and recovery procedures on a regular basis.

- [Overview of Snapshot-Based Recovery of KVM Hosts](#)  
The recovery of a KVM host consists of a series of tasks.
- [KVM System Recovery Scenarios](#)  
How to recover a KVM system deployment.

### 6.19.1 Overview of Snapshot-Based Recovery of KVM Hosts

The recovery of a KVM host consists of a series of tasks.

The recovery procedures use the `diagnostics.iso` image as a virtual CD-ROM to restart the KVM host in rescue mode using the Integrated Lights Out Manager (ILOM). At a high-level, the steps are:

1. Re-create the following:
  - Boot partitions
  - Physical volumes
  - Volume groups
  - Logical volumes
  - File system
  - Swap partition
2. Activate the swap partition
3. Ensure the `/boot` partition is the active boot partition
4. Restore the data
5. Reconfigure GRUB
6. Restart the server

## 6.19.2 KVM System Recovery Scenarios

How to recover a KVM system deployment.

The following scenarios are applicable to a KVM system recovery:

- [Recovering a KVM Host and the Guests from Backup](#)  
This procedure recovers the KVM host and all its guest from a backup of the KVM host and a backup of the guests from the KVM host.
- [Re-imaging a KVM Host and Restoring the Guests from Backup](#)  
This procedure re-images the management domain and reconstructs all the user domains from a backup of the KVM guests.
- [Recover and Restart a KVM Guest](#)  
From the KVM host you can recover a specific guest.
- [Restoring and Recovering Guests from Snapshot Backups](#)  
This procedure can be used to restore lost or damaged files of a KVM guest using a snapshot-based backup of the guest taken from inside the guest.

### 6.19.2.1 Recovering a KVM Host and the Guests from Backup

This procedure recovers the KVM host and all its guest from a backup of the KVM host and a backup of the guests from the KVM host.

A KVM host can be recovered from a snapshot-based backup using the steps below when severe disaster conditions damage the management domain, or when the server hardware is replaced to such an extent that it amounts to new hardware.

Prepare an NFS server to host the backup archives created in [Backing up the KVM host Using Snapshot-Based Backup](#)

The NFS server must be accessible by IP address. For example, on an NFS server with the IP address `nfs_ip`, where the directory `/Backup` contains the backup archives.

- [Recover the KVM Host on Exadata X9M-2](#)  
This procedure describes how to recover the KVM host on an Exadata Database Machine X9M-2 database server.
- [Recover the KVM Host on Exadata X8M-2](#)  
This procedure describes how to recover the KVM host on an Exadata Database Machine X8M-2 database server.
- [Reboot the KVM Host](#)  
Start the KVM host, and continue administrator operations.
- [Recover and Restart the KVM Guests](#)  
From the KVM host recover all of the KVM guests.

### 6.19.2.1.1 Recover the KVM Host on Exadata X9M-2

This procedure describes how to recover the KVM host on an Exadata Database Machine X9M-2 database server.

1. Boot the server and use the system BIOS menus to check the disk controller status. If required, configure the disk controller and set up the disks.
2. Boot the server in diagnostic mode.  
See Booting a Server using the Diagnostic ISO File in *Oracle Exadata System Software User's Guide*.
3. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

4. If it is mounted, unmount `/mnt/cell`

```
umount /mnt/cell
```

5. Confirm the `md` devices on the server.

Confirm that the server contains the devices listed in the following example. Do not proceed and contact Oracle Support if your server differs substantially.

```
ls -al /dev/md*
brw-rw---- 1 root disk 9, 126 Jul 15 06:59 /dev/md126
brw-rw---- 1 root disk 259, 4 Jul 15 06:59 /dev/md126p1
brw-rw---- 1 root disk 259, 5 Jul 15 06:59 /dev/md126p2
brw-rw---- 1 root disk 9, 127 Jul 15 06:28 /dev/md127
brw-rw---- 1 root disk 9, 25 Jul 15 06:28 /dev/md25
```

```

/dev/md:
total 0
drwxr-xr-x 2 root root 140 Jul 15 06:59 .
drwxr-xr-x 18 root root 3400 Jul 15 06:59 ..
lrwxrwxrwx 1 root root 8 Jul 15 06:59 24_0 -> ../md126
lrwxrwxrwx 1 root root 10 Jul 15 06:59 24_0p1 -> ../md126p1
lrwxrwxrwx 1 root root 10 Jul 15 06:59 24_0p2 -> ../md126p2
lrwxrwxrwx 1 root root 7 Jul 15 06:28 25 -> ../md25
lrwxrwxrwx 1 root root 8 Jul 15 06:28 imsm0 -> ../md127

```

6. Remove the logical volumes, the volume group, and the physical volume, in case they still exist after the disaster.

```

lvm vgremove VGExaDb --force
lvm pvremove /dev/md25 --force

```

7. Remove the existing partitions, then verify all partitions were removed.

- a. Use the following command to remove the existing partitions:

```

for v_partition in $(parted -s /dev/md126 print|awk '/^ /
{print $1}')
do
 parted -s /dev/md126 rm ${v_partition}
done

```

- b. Verify by running the following command:

```

parted -s /dev/md126 unit s print

```

The command output should not display any partitions.

8. Create the boot partition.

- a. Start an interactive session using the `parted` command.

```

parted /dev/md126

```

- b. Assign a disk label.

```

(parted) mklabel gpt

```

- c. Set the unit size as sector.

```

(parted) unit s

```

- d. Check the partition table by displaying the existing partitions.

```

(parted) print

```

- e. Remove the partitions listed in the previous step.

```

(parted) rm part#

```



- f. Create a new first partition.

```
(parted) mkpart primary 64s 15114206s
```

- g. Make the new partition bootable.

```
(parted) set 1 boot on
```

9. Create the second primary (boot) partition.

- a. Create a second primary partition as a UEFI boot partition with `fat32`.

```
(parted) mkpart primary fat32 15114207s 15638494s
```

```
(parted) set 2 boot on
```

- b. Write the information to disk, then quit.

```
(parted) quit
```

10. Create the physical volume and volume group.

```
lvm pvcreate /dev/md25
lvm vgcreate VGExaDb /dev/md25
```

If the physical volume or volume group already exists, then remove and then re-create them as follows:

```
lvm vgremove VGExaDb
lvm pvremove /dev/md25
lvm pvcreate /dev/md25
lvm vgcreate VGExaDb /dev/md25
```

11. Create the LVM partitions, then create and mount the file systems.

- a. Create the logical volumes.

```
lvm lvcreate -n LVDbSys1 -L15G VGExaDb -y
lvm lvcreate -n LVDbSwap1 -L16G VGExaDb -y
lvm lvcreate -n LVDbSys2 -L15G VGExaDb -y
lvm lvcreate -n LVDbHome -L4G VGExaDb -y
lvm lvcreate -n LVDbVar1 -L2G VGExaDb -y
lvm lvcreate -n LVDbVar2 -L2G VGExaDb -y
lvm lvcreate -n LVDbVarLog -L18G VGExaDb -y
lvm lvcreate -n LVDbVarLogAudit -L1G VGExaDb -y
lvm lvcreate -n LVDbTmp -L3G VGExaDb -y
lvm lvcreate -n LVDoNotRemoveOrUse -L2G VGExaDb -y
lvm lvcreate -n LVDbExaVMImages -L1500G VGExaDb -y
lvextend -l +98%FREE /dev/VGExaDb/LVDbExaVMImages
```

- b. Create the file systems.

```
mkfs.xfs -f /dev/VGExaDb/LVDbSys1
mkfs.xfs -f /dev/VGExaDb/LVDbSys2
mkfs.xfs -f /dev/VGExaDb/LVDbHome
mkfs.xfs -f /dev/VGExaDb/LVDbVar1
```

```
mkfs.xfs -f /dev/VGExaDb/LVDbVar2
mkfs.xfs -f /dev/VGExaDb/LVDbVarLog
mkfs.xfs -f /dev/VGExaDb/LVDbVarLogAudit
mkfs.xfs -f /dev/VGExaDb/LVDbTmp
mkfs.xfs -m crc=1 -m reflink=1 -f /dev/VGExaDb/LVDbExaVMImages
mkfs.xfs -f /dev/md126p1
mkfs.vfat -v -c -F 32 -s 2 /dev/md126p2
```

- c. Label the file systems.

```
xfs_admin -L DBSYS /dev/VGExaDb/LVDbSys1
xfs_admin -L HOME /dev/VGExaDb/LVDbHome
xfs_admin -L VAR /dev/VGExaDb/LVDbVar1
xfs_admin -L DIAG /dev/VGExaDb/LVDbVarLog
xfs_admin -L AUDIT /dev/VGExaDb/LVDbVarLogAudit
xfs_admin -L TMP /dev/VGExaDb/LVDbTmp
xfs_admin -L EXAVMIMAGES /dev/VGExaDb/LVDbExaVMImages
xfs_admin -L BOOT /dev/md126p1
dosfslabel /dev/md126p2 ESP
```

- d. Create mount points for all the partitions, and mount the respective partitions.

For example, assuming that `/mnt` is used as the top level directory for the recovery operation, you could use the following commands to create the directories and mount the partitions:

```
mount -t xfs /dev/VGExaDb/LVDbSys1 /mnt
mkdir -p /mnt/home
mount -t xfs /dev/VGExaDb/LVDbHome /mnt/home
mkdir -p /mnt/var
mount -t xfs /dev/VGExaDb/LVDbVar1 /mnt/var
mkdir -p /mnt/var/log
mount -t xfs /dev/VGExaDb/LVDbVarLog /mnt/var/log
mkdir -p /mnt/var/log/audit
mount -t xfs /dev/VGExaDb/LVDbVarLogAudit /mnt/var/log/audit
mkdir -p /mnt/tmp
mount -t xfs /dev/VGExaDb/LVDbTmp /mnt/tmp
mkdir -p /mnt/EXAVMIMAGES
mount -t xfs /dev/VGExaDb/LVDbExaVMImages /mnt/EXAVMIMAGES
mkdir -p /mnt/boot
mount -t xfs /dev/md126p1 /mnt/boot
mkdir -p /mnt/boot/efi
mount -t vfat /dev/md126p2 /mnt/boot/efi
```

12. Create the system swap space.

For example:

```
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

13. Bring up the network.

```
ip address add ip_address_for_eth0/netmask_for_eth0 dev eth0
ip link set up eth0
ip route add default via gateway_address dev eth0
```

**14.** Mount the NFS server containing the backup.

The following example assumes that the backup is located in the `/export` directory of the NFS server with IP address `nfs_ip`.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/export /root/mnt
```

**15.** Restore the files from the backup.

Assuming that the backup was created using the procedure in [Backing up the KVM host Using Snapshot-Based Backup](#), you can restore the files by using the following command:

```
tar --acls --xattrs --xattrs-include=* --format=pax -pjxvf /root/mnt/
myKVMbackup.tar.bz2 -C /mnt
```

**16.** Create the directory for kdump service.

```
mkdir /mnt/EXAVMIMAGES/crashfiles
```

**17.** Check the restored `fstab` file (at `/mnt/etc/fstab`), and comment out any line that references `/EXAVMIMAGES`.**18.** Unmount the restored file systems.

For example:

```
umount /mnt/tmp
umount /mnt/var/log/audit
umount /mnt/var/log
umount /mnt/var
umount /mnt/home
umount /mnt/EXAVMIMAGES
umount /mnt/boot/efi
umount /mnt/boot
umount /mnt
```

**19.** Check the boot devices and set the boot order.**a.** Check the available boot devices, and identify the boot device that is associated with Redhat Boot Manager (`\EFI\REDHAT\SHIMX64.EFI`).

For example:

```
efibootmgr -v
BootCurrent: 0019
Timeout: 1 seconds
BootOrder:
0019,0000,0002,0010,0009,0017,000A,000B,0018,0005,0006,0007,0008,0013,
0014,0015,0016,0003,0011,0004,0012,001A
Boot0000* RedHat Boot Manager
HD(2,GPT,eec54dfd-8928-4874-833d-5b0b9e914b99,0xe69fdf,0x80000)/
File(\EFI\REDHAT\SHIMX64.EFI)
Boot0002* NET0:PXE IPv4 Intel(R) I210 Gigabit Network Connection /
Pci(0x1c,0x4)/Pci(0x0,0x0)/MAC(0010e0fc6e94,1)/
IPv4(0.0.0.00.0.0.0,0,0)..BO
```

```
Boot0003* PCIE5:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:22:38:0A /Pci (0x2,0x0)/Pci (0x0,0x0)/MAC (b8cef622380a,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0004* PCIE5:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:22:38:0B /Pci (0x2,0x0)/Pci (0x0,0x1)/MAC (b8cef622380b,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0005* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci (0x2,0x0)/Pci (0x0,0x0)/MAC (3cfdfe915070,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0006* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci (0x2,0x0)/Pci (0x0,0x1)/MAC (3cfdfe915071,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0007* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci (0x2,0x0)/Pci (0x0,0x2)/MAC (3cfdfe915072,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0008* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci (0x2,0x0)/Pci (0x0,0x3)/MAC (3cfdfe915073,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0009* PCIE1:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:44:51:9C /Pci (0x2,0x0)/Pci (0x0,0x0)/MAC (b8cef644519c,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot000A* PCIE1:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:44:51:9D /Pci (0x2,0x0)/Pci (0x0,0x1)/MAC (b8cef644519d,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot000B* PCIE1:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:44:51:9D /Pci (0x2,0x0)/Pci (0x0,0x1)/MAC (b8cef644519d,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0010* NET0:PXE IPv4 Intel(R) I210 Gigabit Network
Connection /Pci (0x1c,0x4)/Pci (0x0,0x0)/MAC (0010e0fc6e94,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0011* PCIE5:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:22:38:0A /Pci (0x2,0x0)/Pci (0x0,0x0)/MAC (b8cef622380a,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0012* PCIE5:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:22:38:0B /Pci (0x2,0x0)/Pci (0x0,0x1)/MAC (b8cef622380b,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0013* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci (0x2,0x0)/Pci (0x0,0x0)/MAC (3cfdfe915070,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0014* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci (0x2,0x0)/Pci (0x0,0x1)/MAC (3cfdfe915071,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0015* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci (0x2,0x0)/Pci (0x0,0x2)/MAC (3cfdfe915072,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0016* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter /
Pci (0x2,0x0)/Pci (0x0,0x3)/MAC (3cfdfe915073,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0017* PCIE1:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:44:51:9C /Pci (0x2,0x0)/Pci (0x0,0x0)/MAC (b8cef644519c,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0018* PCIE1:PXE IPv4 Mellanox Network Adapter -
B8:CE:F6:44:51:9D /Pci (0x2,0x0)/Pci (0x0,0x1)/MAC (b8cef644519d,1)/
IPv4 (0.0.0.00.0.0.0,0,0)..BO
Boot0019* USB:SP:SUN Remote ISO CDROM1.01 /Pci (0x14,0x0)/
```

```

USB(7,0)/USB(3,0)/CDROM(1,0x28,0x3100)..BO
Boot001A* Oracle Linux (grubx64.efi)
HD(2,GPT,eec54dfd-8928-4874-833d-5b0b9e914b99,0xe69fdf,0x80000)/
File(\EFI\REDHAT\GRUBX64.EFI)..BO
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false

```

- b. Configure the device that is associated with Redhat Boot Manager (\EFI\REDHAT\SHIMX64.EFI) to be first in the boot order.

In this example, Redhat Boot Manager is associated with boot device 0000:

```

efibootmgr -o 0000
BootCurrent: 0019
Timeout: 1 seconds
BootOrder: 0000
Boot0000* RedHat Boot Manager
Boot0002* NET0:PXE IPv4 Intel(R) I210 Gigabit Network Connection
Boot0003* PCIE5:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:22:38:0A
Boot0004* PCIE5:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:22:38:0B
Boot0005* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0006* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0007* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0008* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0009* PCIE1:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:44:51:9C
Boot000A* PCIE1:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:44:51:9D
Boot000B* PCIE1:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:44:51:9D
Boot0010* NET0:PXE IPv4 Intel(R) I210 Gigabit Network Connection
Boot0011* PCIE5:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:22:38:0A
Boot0012* PCIE5:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:22:38:0B
Boot0013* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0014* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0015* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0016* PCIE3:PXE IPv4 Oracle Quad Port 10GBase-T Adapter
Boot0017* PCIE1:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:44:51:9C
Boot0018* PCIE1:PXE IPv4 Mellanox Network Adapter - B8:CE:F6:44:51:9D
Boot0019* USB:SP:SUN Remote ISO CDROM1.01
Boot001A* Oracle Linux (grubx64.efi)
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false

```

20. Restart the system.

```
reboot
```

21. Disconnect the `diagnostics.iso` file.

See *Booting a Server using the Diagnostic ISO File* in *Oracle Exadata System Software User's Guide*.

22. Log back into the server as the `root` user.

23. Run the `imageinfo` command and verify that the image status is `success`.

For example:

```
imageinfo

Kernel version: 4.14.35-2047.502.5.el7uek.x86_64 #2 SMP Wed Apr 14
15:08:41
PDT 2021 x86_64
Uptrack kernel version: 4.14.35-2047.503.1.el7uek.x86_64 #2 SMP Fri
Apr 23
15:20:41 PDT 2021 x86_64
Image kernel version: 4.14.35-2047.502.5.el7uek
Image version: 21.2.1.0.0.210608
Image activated: 2021-07-12 14:58:03 +0900
Image status: success
Node type: COMPUTE
System partition on device: /dev/mapper/VGExaDb-LVDbSys1
```

The KVM host has been recovered.

### 6.19.2.1.2 Recover the KVM Host on Exadata X8M-2

This procedure describes how to recover the KVM host on an Exadata Database Machine X8M-2 database server.

1. Boot the server in diagnostic mode.

See Booting a Server using the Diagnostic ISO File in *Oracle Exadata System Software User's Guide*.

2. Log in to the diagnostics shell as the `root` user.

When prompted, enter the diagnostics shell.

For example:

```
Choose from following by typing letter in '()':
(e)nter interactive diagnostics shell. Must use credentials
from Oracle support to login (reboot or power cycle to exit
the shell),
(r)estore system from NFS backup archive,
Type e to enter the diagnostics shell and log in as the root user.
```

If prompted, log in to the system as the `root` user. If you are prompted for the `root` user password and do not have it, then contact Oracle Support Services.

3. If required, use `/opt/MegaRAID/storcli/storcli64` to configure the disk controller to set up the disks.
4. Remove the logical volumes, the volume group, and the physical volume, in case they still exist after the disaster.

```
lvm vgremove VGExaDb --force
lvm pvremove /dev/sda3 --force
```

5. Remove the existing partitions, then verify all partitions were removed. The below script can be used.

```
for v_partition in $(parted -s /dev/sda print|awk '/^ / {print $1}')
do
 parted -s /dev/sda rm ${v_partition}
done

parted -s /dev/sda unit s print
Model: AVAGO MR9[2783.921605] sda:361-16i (scsi)
Disk /dev/sda: 3509760000s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number Start End Size File system Name Flags
```

6. Create three partitions on /dev/sda
- a. Get the end sector for the disk /dev/sda from a surviving KVM host and store it in a variable:

```
end_sector_logical=$(parted -s /dev/sda unit s print|perl -ne '/
^Disk\s+\S+:\s+(\d+)\s/
 and print $1')
end_sector=$((expr $end_sector_logical - 34)
echo $end_sector
```

The values for the start and end sectors in the commands below were taken from a surviving KVM host. Because these values can change over time, it is recommended that these values are checked from a KVM host at the time of performing this procedure. For example, for an Oracle Exadata Database Machine X8M-2 database server the following might be seen:

```
parted -s /dev/sda unit s print
Model: AVAGO MR9361-16i (scsi)
Disk /dev/sda: 7025387520s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:
Number Start End Size File system Name
Flags
1 64s 1048639s 1048576s xfs primary
boot
2 1048640s 1572927s 524288s fat32 primary
boot
3 1572928s 7025387486s 7023814559s primary lvm
```

- b. Create the boot partition, /dev/sda1.

```
parted -s /dev/sda mklabel gpt mkpart primary 64s 1048639s set 1
boot on
```

- c. Create the efi boot partition , /dev/sda2.

```
parted -s /dev/sda mkpart primary fat32 1048640s 1572927s set
2 boot on
```

- d. Create the partition that will hold the logical volumes, /dev/sda3.

```
parted -s /dev/sda mkpart primary 1572928s ${end_sector}s set
3 lvm on
```

- e. Verify all the partitions have been created.

```
parted -s /dev/sda unit s print
Model: AVAGO MR9[2991.834796] sda: sda1 sda2 sda3
361-16i(scsi)
Disk /dev/sda:3509760000s
Sector size(logical/physical): 512B/512B
Partition Table:gpt
Disk Flags:
Number Start End Size File system
Name Flags
1 64s 1048639s 1048576s xfs
primary boot
2 1048640s 1572927s 524288s fat32
primary boot
3 1572928s 3509759966s 3508187039s
primary lvm
```

7. Create logical volumes and file systems.

- a. Create the physical volume and the volume group.

```
lvm pvcreate /dev/sda3
lvm vgcreate VGExaDb /dev/sda3
```

- b. Create and label the logical volume for the file system that will contain the first system partition and build a xfs file system on it.

```
lvm lvcreate -n LVDbSys1 -L15G VGExaDb -y
mkfs.xfs -L DBSYS /dev/VGExaDb/LVDbSys1 -f
```

- c. Create and label the logical volume for the swap directory.

```
lvm lvcreate -n LVDbSwap1 -L16G VGExaDb -y
mkswap -L SWAP /dev/VGExaDb/LVDbSwap1
```

- d. Create the logical volume for the second system partition and build a xfs file system on it.

```
lvm lvcreate -n LVDbSys2 -L15G VGExaDb -y
mkfs.xfs /dev/VGExaDb/LVDbSys2
```



- e. Create and label the logical volume for the HOME partition and build a xfs file system on it.

```
lvm lvcreate -n LVDbHome -L4G VGEaDb -y
mkfs.xfs -L HOME /dev/VGEaDb/LVDbHome
```

- f. Create the logical volume for the tmp partition and build a xfs file system on it.

```
lvm lvcreate -n LVDbTmp -L3G VGEaDb -y
mkfs.xfs -L TMP /dev/VGEaDb/LVDbTmp -f
```

- g. Create the logical volume for the first var partition and build a xfs file system on it.

```
lvm lvcreate -n LVDbVar1 -L2G VGEaDb -y
mkfs.xfs -L VAR /dev/VGEaDb/LVDbVar1 -f
```

- h. Create the logical volume for the second var partition and build a xfs file system on it.

```
lvm lvcreate -n LVDbVar2 -L2G VGEaDb -y
mkfs.xfs /dev/VGEaDb/LVDbVar2 -f
```

- i. Create and label the logical volume for the LVDbVarLog partition and build a xfs file system on it.

```
lvm lvcreate -n LVDbVarLog -L18G VGEaDb -y
mkfs.xfs -L DIAG /dev/VGEaDb/LVDbVarLog -f
```

- j. Create and label the logical volume for the LVDbVarLogAudit partition and build a xfs file system on it.

```
lvm lvcreate -n LVDbVarLogAudit -L1G VGEaDb -y
mkfs.xfs -L AUDIT /dev/VGEaDb/LVDbVarLogAudit -f
```

- k. Create the LVDoNotRemoveOrUse logical volume.

```
lvm lvcreate -n LVDoNotRemoveOrUse -L2G VGEaDb -y
```

- l. Create the logical volume for the guest storage repository and build a xfs file system on it.

```
lvm lvcreate -n LVDbExaVMImages -L1500G VGEaDb -y
mkfs.xfs -m crc=1 -m reflink=1 -L EXAVMIMAGES /dev/VGEaDb/
LVDbExaVMImages -f
```

- m. Create a file system on the /dev/sda1 partition, and label it.

```
mkfs.xfs -L BOOT /dev/sda1 -f
```

- n. Create a file system on the /dev/sda2 partition, and label it.

```
mkfs.vfat -v -c -F 32 -s 2 /dev/sda2
dosfslabel /dev/sda2 ESP
```

8. Create mount points for all the partitions and mount the respective partitions.

For example, if `/mnt` is used as the top-level directory, the mounted list of partitions might look like:

```
/dev/VGExaDb/LVDbSys1 on /mnt
/dev/sda1 on /mnt/boot
/dev/sda2 on /mnt/boot/efi
/dev/VGExaDb/LVDbHome on /mnt/home
/dev/VGExaDb/LVDbTmp on /mnt/tmp
/dev/VGExaDb/LVDbVar1 on /mnt/var
/dev/VGExaDb/LVDbVarLog on /mnt/var/log
/dev/VGExaDb/LVDbVarLogAudit on /mnt/var/log/audit
/dev/VGExaDb/LVDbExaVMImages on /mnt/EXAVMIMAGES
```

The following example mounts the system partition and creates 2 mount points for the boot partitions.

```
mount /dev/VGExaDb/LVDbSys1 /mnt -t xfs
mkdir /mnt/boot
mount /dev/sda1 /mnt/boot -t xfs
mkdir /mnt/boot/efi
mount /dev/sda2 /mnt/boot/efi -t vfat
mkdir /mnt/home
mount /dev/VGExaDb/LVDbHome /mnt/home -t xfs
mkdir /mnt/tmp
mount /dev/VGExaDb/LVDbTmp /mnt/tmp -t xfs
mkdir /mnt/var
mount /dev/VGExaDb/LVDbVar1 /mnt/var -t xfs
mkdir /mnt/var/log
mount /dev/VGExaDb/LVDbVarLog /mnt/var/log -t xfs
mkdir /mnt/var/log/audit
mount /dev/VGExaDb/LVDbVarLogAudit /mnt/var/log/audit -t xfs
mkdir /mnt/EXAVMIMAGES
mount /dev/VGExaDb/LVDbExaVMImages /mnt/EXAVMIMAGES -t xfs
```

9. Bring up the network on `eth0` and (if not using DHCP) assign the host's IP address and netmask to it. If using DHCP then manually configure the IP address for the host.

```
ip link set eth0 up
ip address add ip_address_for_eth0/netmask_for_eth0 dev eth0
ip route add default via gateway_ip_address dev eth0
```

10. Mount the NFS server holding the backup.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

11. Restore the files from the backup.

Assuming that the backup was created using the procedure in [Backing up the KVM host Using Snapshot-Based Backup](#), you can restore the files by using the following command:

```
tar --acls --xattrs --xattrs-include=* --format=pax -pjxvf /root/mnt/
myKVMbackup.tar.bz2 -C /mnt
```

12. Create the directory for kdump service.

```
mkdir /mnt/EXAVMIMAGES/crashfiles
```

13. Set the boot device using `efibootmgr`.

- a. Disable and delete the Oracle Linux boot device.

If the entry `ExadataLinux_1` is seen, then remove this entry and recreate it. Example:

```
efibootmgr
BootCurrent: 0000
Timeout: 1 seconds
BootOrder: 0000,0001,000A,000B,0007,0008,0004,0005
Boot0000* ExadataLinux_1
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0004* PCIE1:PXE IP4 Oracle dual 25Gb Ethernet Adapter or dual
10Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle dual 25Gb Ethernet Adapter or dual
10Gb Ethernet Adapter
Boot0007* NET1:PXE IP4 Oracle Dual Port 10Gb/25Gb SFP28 Ethernet
Controller
Boot0008* NET2:PXE IP4 Oracle Dual Port 10Gb/25Gb SFP28 Ethernet
Controller
Boot000A PCIE2:PXE IP4 Mellanox Network Adapter - 50:6B:4B:CB:EF:F2
Boot000B PCIE2:PXE IP4 Mellanox Network Adapter - 50:6B:4B:CB:EF:F3
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false
```

In this example, `ExadataLinux_1` (Boot000) would be disabled and removed. Use the commands below to disable and delete the boot device.

Disable old `ExadataLinux_1`:

```
efibootmgr -b 0000 -A
```

Delete old `ExadataLinux_1`:

```
efibootmgr -b 0000 -B
```

- b. Recreate the boot entry for `ExadataLinux_1` and then view the boot order entries.

```
efibootmgr -c -d /dev/sda -p 2 -l '\EFI\REDHAT\SHIM.EFI' -L
'ExadataLinux_1'
efibootmgr
BootCurrent: 0000
Timeout: 1 seconds
BootOrder: 0000,0001,0007,0008,0004,0005,000B,000C
```

```

Boot0000* ExadataLinux_1
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0004* PCIE1:PXE IP4 Oracle dual 25Gb Ethernet Adapter or
dual 10Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle dual 25Gb Ethernet Adapter or
dual 10Gb Ethernet Adapter
Boot0007* NET1:PXE IP4 Oracle Dual Port 10Gb/25Gb SFP28
Ethernet Controller
Boot0008* NET2:PXE IP4 Oracle Dual Port 10Gb/25Gb SFP28
Ethernet Controller
Boot000B* PCIE2:PXE IP4 Mellanox Network Adapter -
EC:0D:9A:CC:1E:46
Boot000C* PCIE2:PXE IP4 Mellanox Network Adapter -
EC:0D:9A:CC:1E:47
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false

```

- c. In the output from the `efibootmgr` command, make note of the boot order number for `ExadataLinux_1` and use that value in the following commands.

```

efibootmgr -b entry number -A
efibootmgr -b entry number -a

```

For example, in the previous output shown in step 13.b, `ExadataLinux_1` was listed as (`Boot0000`). So you would use the following commands:

```

efibootmgr -b 0000 -A
efibootmgr -b 0000 -a

```

- d. Set the correct boot order. Set `ExadataLinux_1` as the first boot device. The remaining devices should stay in the same boot order.

```

efibootmgr -o 0000,0001,0004,0005,0007,0008,000B,000C

```

- e. Check the boot order.

The boot order should now look like the following:

```

efibootmgr
BootCurrent: 0000
Timeout: 1 seconds
BootOrder: 0000,0001,0004,0005,0007,0008,000B,000C
Boot0000* ExadataLinux_1
Boot0001* NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection
Boot0004* PCIE1:PXE IP4 Oracle dual 25Gb Ethernet Adapter or
dual 10Gb Ethernet Adapter
Boot0005* PCIE1:PXE IP4 Oracle dual 25Gb Ethernet Adapter or
dual 10Gb Ethernet Adapter
Boot0007* NET1:PXE IP4 Oracle Dual Port 10Gb/25Gb SFP28 Ethernet
Controller
Boot0008* NET2:PXE IP4 Oracle Dual Port 10Gb/25Gb SFP28 Ethernet
Controller
Boot000B* PCIE2:PXE IP4 Mellanox Network Adapter -
EC:0D:9A:CC:1E:46

```

```
Boot000C* PCIE2:PXE IP4 Mellanox Network Adapter - EC:0D:9A:CC:1E:47
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false
```

- f. Check the boot order using the `ubiosconfig` command.

```
ubiosconfig export all -x /tmp/ubiosconfig.xml
```

Make sure the `ExadataLinux_1` entry is the first child element of `boot_order`.

```
cat /tmp/ubiosconfig.xml
<boot_order>
 <boot_device>
 <description>ExadataLinux_1</description>
 <instance>1</instance> </boot_device>
 <boot_device>
 <description>NET0:PXE IP4 Intel(R) I210 Gigabit Network Connection</
description>
 <instance>1</instance>
 </boot_device>
 <boot_device>
 <description>PCIE1:PXE IP4 Oracle dual 25Gb Ethernet Adapter or
dual 10Gb Ethernet Adapter</description>
 <instance>1</instance>
 </boot_device>
 <boot_device>
 <description>PCIE1:PXE IP4 Oracle dual 25Gb Ethernet Adapter or
dual 10Gb Ethernet Adapter</description>
 <instance>2</instance>
 </boot_device>
 <boot_device>
 <description>NET1:PXE IP4 Oracle Dual Port 10Gb/25Gb SFP28 Ethernet
Controller</description>
 <instance>1</instance>
 </boot_device>
 <boot_device>
 <description>NET2:PXE IP4 Oracle Dual Port 10Gb/25Gb SFP28 Ethernet
Controller</description>
 <instance>1</instance>
 </boot_device>
 <boot_device>
 <description>PCIE2:PXE IP4 Mellanox Network Adapter -
EC:0D:9A:CC:1E:46</description>
 <instance>1</instance>
 </boot_device>
 <boot_device> <description>PCIE2:PXE IP4 Mellanox Network Adapter -
EC:0D:9A:CC:1E:47</description>
 <instance>1</instance> </boot_device> </boot_order>
```

- g. Check the restored `fstab` file and comment out any reference to `/EXAVMIMAGES`.

Navigate to `/mnt/etc`.

```
cd /mnt/etc
```

In the `/mount/etc/fstab` file, comment out any line that references `/EXAVMIMAGES`.

The KVM host has been recovered.

### 6.19.2.1.3 Reboot the KVM Host

Start the KVM host, and continue administrator operations.

The KVM host partitions have been restored to the server and the boot order now has `ExadataLinux_1` as the first device.

1. Detach the `diagnostics.iso` file.

Assuming that the ILOM web interface console was used to connect the CDROM image, click the **Disconnect** button.

2. Unmount the restored `/dev/sda1` partitions so `/dev/sda1` can be remounted on `/boot`.

```
cd /
umount /mnt/boot/efi
umount /mnt/boot
umount /mnt/home
umount /mnt/var/log/audit
umount /mnt/var/log/
umount /mnt/var
umount /mnt/EXAVMIMAGES/
umount /mnt
umount /root/mnt
```

3. Restart the system.

```
reboot
```

The current `/boot` partition will be used to start the server.

### 6.19.2.1.4 Recover and Restart the KVM Guests

From the KVM host recover all of the KVM guests.

This procedure complements the recommended backup procedure described in [Method 1: Back Up All of the KVM Guests](#). It assumes that the KVM host is operational after recovery and all the guests are being recovered.

1. Mount the backup NFS server that holds the guest storage repository (`/EXAVMIMAGES`) backup to restore the `/EXAVMIMAGES` file system.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

2. Mount the `/EXAVMIMAGES` file system.

```
mount -t xfs /dev/VGExaDb/LVDbExaVMImages /EXAVMIMAGES
```

3. Restore the `/EXAVMIMAGES` file system.

For example, use the command below to restore all guests:

```
tar --acls --xattrs --xattrs-include=* --format=pax -Spvxf /root/mnt/
exavmimages.tar.bz2 -C /
```

4. Check each guest XML configuration file.

During recovery of the KVM host, each guest XML configuration file is restored to `/etc/libvirt/qemu/guestname.xml`. Additionally, the previous step restores a copy of each guest XML file to `/XML/guestname.xml`.

For each guest, compare the XML configuration files. For example:

```
diff /etc/libvirt/qemu/guestname.xml /XML/guestname.xml
```

In most cases the XML configuration files will be identical. However, if there are differences, you should examine both files and copy the most up-to-date configuration to `/etc/libvirt/qemu/guestname.xml`.

5. Define each restored guest in the KVM hypervisor.

```
virsh define /etc/libvirt/qemu/guestname.xml
```

6. Enable autostart for each restored guest.

```
/opt/exadata_ovm/vm_maker --autostart guestname --disable
/opt/exadata_ovm/vm_maker --autostart guestname --enable
```

7. Unmount the NFS backup.

```
umount /root/mnt
```

8. Restore any reference to `/EXAVMIMAGES` in the `fstab` file.

In the `/etc/fstab` file, remove comments for any line that references `/EXAVMIMAGES`.

9. Bring up each guest.

```
/opt/exadata_ovm/vm_maker --start-domain guestname
```

At this point all of the guests should come up along with Oracle Grid Infrastructure and the Oracle Database instances.

### 6.19.2.2 Re-imaging a KVM Host and Restoring the Guests from Backup

This procedure re-images the management domain and reconstructs all the user domains from a backup of the KVM guests.

The following procedure can be used when the KVM host is damaged beyond repair and no backup exists for the KVM host, but there is a backup available of the storage repository (`/EXAVMIMAGES` file system) containing all the guests.

1. Re-image the KVM host with the image used in the other hosts in the rack using the procedure described in [Re-Imaging the Oracle Exadata Database Server](#).

2. If the recovery is on Oracle Exadata Database Machine eighth rack, then perform the procedure described in [Configuring Oracle Exadata Database Machine Eighth Rack Oracle Linux Database Server After Recovery](#).

3. Create the bonded interface and bonded network bridge.

The following command uses `vmbondeth0` as the name of the network bridge.

```
/opt/exadata_ovm/vm_maker --add-bonded-bridge vmbondeth0 --first-slave eth1 --second-slave eth2
```

4. Mount the backup NFS server containing the guest backups.

In this procedure, the backup NFS server is mounted on `/remote_FS`.

```
mkdir -p /remote_FS
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/location_of_backup /remote_FS
```

5. Restore the `/EXAVMIMAGES` file system.

```
tar --acls --xattrs --xattrs-include=* --format=pax -Spxvf /remote_FS/backup-of-exavmimages.tar -C /
```

#### Note:

Following restoration, the guest image files under `/EXAVMINAGES` are separate regular files. Consequently, the space usage in `/EXAVMINAGES` may go up after the restoration process compared to the original space usage, which may have benefited from savings associated with reflinked files.

6. Restore the guest specific files for the KVM hypervisor.

```
/usr/bin/cp /XML/*.xml /etc/libvirt/qemu/
```

#### Note:

This step assumes that the files restored to `/XML` were copied into the backup location during the backup procedure.

7. Define each guest in the KVM hypervisor.

```
virsh define /etc/libvirt/qemu/guestname.xml
```

8. Enable autostart for each restored guest.

```
/opt/exadata_ovm/vm_maker --autostart guestname --disable
/opt/exadata_ovm/vm_maker --autostart guestname --enable
```



9. Bring up each guest.

```
/opt/exadata_ovm/vm_maker --start-domain guestname
```

At this point all the guests should start along with the Oracle Grid Infrastructure and the database instances.

### 6.19.2.3 Recover and Restart a KVM Guest

From the KVM host you can recover a specific guest.

This procedure complements the backup procedure described in [Method 2: Back Up an Individual Guest](#). It assumes that the KVM host is operational and that the guest being recovered does not exist on the KVM host.

1. Mount the backup NFS server that holds the guest backup.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

2. Restore the guest image files.

For example, use the following command to extract the contents of the guest backup file named `exavmimage.tar.bz2`:

```
tar --acls --xattrs --xattrs-include=* --format=pax -Spxvf /root/mnt/
exavmimage.tar.bz2 -C /
```

3. Restore the guest XML configuration file.

```
cp /XML/guestname.xml /etc/libvirt/qemu
```

4. Define the guest in the KVM hypervisor.

```
virsh define /etc/libvirt/qemu/guestname.xml
```

5. Enable autostart for the restored guest.

```
/opt/exadata_ovm/vm_maker --autostart guestname --disable
/opt/exadata_ovm/vm_maker --autostart guestname --enable
```

6. Unmount the NFS backup.

```
umount /root/mnt
```

7. Bring up the guest.

```
/opt/exadata_ovm/vm_maker --start-domain guestname
```

## 6.19.2.4 Restoring and Recovering Guests from Snapshot Backups

This procedure can be used to restore lost or damaged files of a KVM guest using a snapshot-based backup of the guest taken from inside the guest.

Use this procedure to restore lost/damaged files of a KVM guest by using a snapshot-based backup created from within the guest as described in section [Method 3: Back Up a Guest Internally](#).

Log in to the KVM guest as the root user.

1. Mount the backup NFS server to restore the damaged or lost files.

```
mkdir -p /root/mnt
mount -t nfs -o ro,intr,soft,proto=tcp,nolock nfs_ip:/
location_of_backup /root/mnt
```

2. Extract the damaged or lost files from the backup to a staging area.

- a. Prepare a staging area to hold the extracted files. The backup logical volume `LVDbSys2` may be used for this:

```
mkdir /backup-LVM
mount /dev/mapper/VGExaDb-LVDbSys2 /backup-LVM
mkdir /backup-LVM/tmp_restore
```

- b. Extract the needed files.

```
tar --acls --xattrs --xattrs-include=* --format=pax -pjxvf /
root/mnt/tar_file_name -C /backup-LVM/tmp_restore
absolute_path_of_file_to_be_restored
```

3. Restore the damaged or lost files from the temporary staging area as needed.
4. Restart the KVM guest if needed.

## 6.20 Removing a Guest

You can remove a guest in Oracle Linux KVM using either `OEDACLI` or the `vm_maker` utility.

- [Removing a Guest from an Oracle RAC Cluster Using OEDACLI](#)  
You can use `OEDACLI` to remove a guest from a cluster.
- [Removing a Guest Using vm\\_maker](#)  
You can use the `vm_maker` utility to remove a guest.

### 6.20.1 Removing a Guest from an Oracle RAC Cluster Using OEDACLI

You can use `OEDACLI` to remove a guest from a cluster.

The following procedure removes a guest from a cluster. If the guest is not part of a cluster, then you can skip the cluster-related commands.

1. Load the XML configuration file (`es.xml` for example) into OEDACLI.

```
./oedacli -c full_path_to_XML_file
```

2. Use the following OEDACLI commands to delete the database instance from the cluster node.

```
DELETE GUEST WHERE srcname=guest_FQDN stepname=ADD_INSTANCE
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

3. Use the following OEDACLI commands to delete the Oracle Database home from the cluster.

```
DELETE GUEST WHERE srcname=guest_FQDN stepname=EXTEND_DBHOME
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

4. Use the following OEDACLI commands to delete the guest node from the cluster.

```
DELETE GUEST WHERE srcname=guest_FQDN stepname=ADD_NODE
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

5. Use the following OEDACLI commands to remove connectivity to the storage servers and delete the users on the guest.

```
DELETE GUEST WHERE srcname=guest_FQDN stepname=CELL_CONNECTIVITY
SAVE ACTION
DELETE GUEST WHERE srcname=guest_FQDN stepname=CREATE_USERS
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

6. Use the following OEDACLI commands to remove the guest.

```
DELETE GUEST WHERE srcname=guest_FQDN stepname=CREATE_GUEST
SAVE ACTION
MERGE ACTIONS
DEPLOY ACTIONS
```

7. Save the updated configuration information.

Specify the full path to a directory where the updated configuration information will be saved.

```
SAVE FILES LOCATION=output_directory
```

8. View the list of existing guests on the KVM host.

```
/opt/exadata_ovm/vm_maker --list-domains
```

The guest you just removed should not be listed.

## 6.20.2 Removing a Guest Using `vm_maker`

You can use the `vm_maker` utility to remove a guest.

The following procedure removes a guest from a cluster. If the guest is not part of a cluster, then you can skip the commands related to Oracle Clusterware.

1. If your cluster uses Quorum failure groups, these need to be deleted first.
  - a. List the existing quorum devices.

```
~]# /opt/oracle.SupportTools/quorumdiskmgr --list --target
```

- b. Delete the quorum devices.

```
~]# /opt/oracle.SupportTools/quorumdiskmgr --delete --target
```

2. Remove the database instance on the guest from the cluster node.  
Refer to [Deleting Instances from Oracle RAC Databases](#) for the instructions.
3. Remove the guest node from the cluster.  
Refer to [Deleting a Cluster Node on Linux and UNIX Systems](#) for the instructions.
4. Delete the guest using `vm_maker`.

```
/opt/exadata_ovm/vm_maker --remove-domain guest_name
```

5. View the list of existing guests on the KVM host.

```
/opt/exadata_ovm/vm_maker --list-domains
```

The guest you just removed should not be listed.

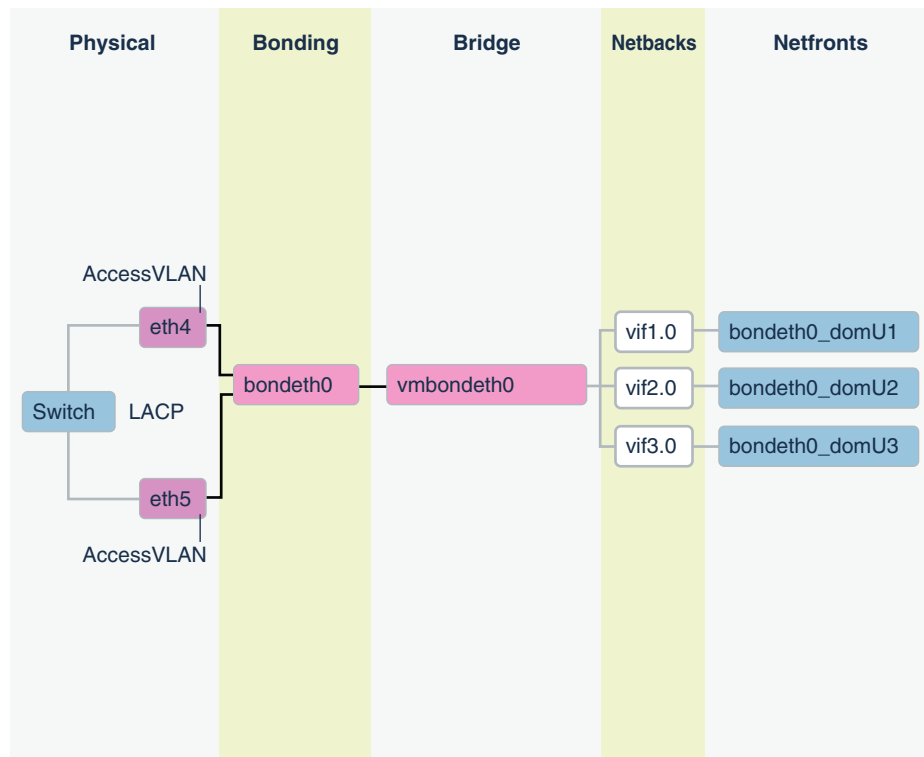
## 6.21 Using Client Network VLAN Tagging with Oracle Linux KVM

This topic describes the implementation of tagged VLAN interfaces for the client network in conjunction with Oracle Linux KVM.

Oracle databases running in Oracle Linux KVM guests on Exadata Database Machine are accessed through the client Ethernet network defined in the Oracle Exadata Deployment Assistant (OEDA) configuration tool. Client network configuration in both the KVM host and guests is done automatically when the OEDA installation tool creates the first guest during initial deployment.

The following figure shows a default bonded client network configuration:

**Figure 6-1 NIC Layout in an Oracle Virtual Environment**



The network has the following configuration:

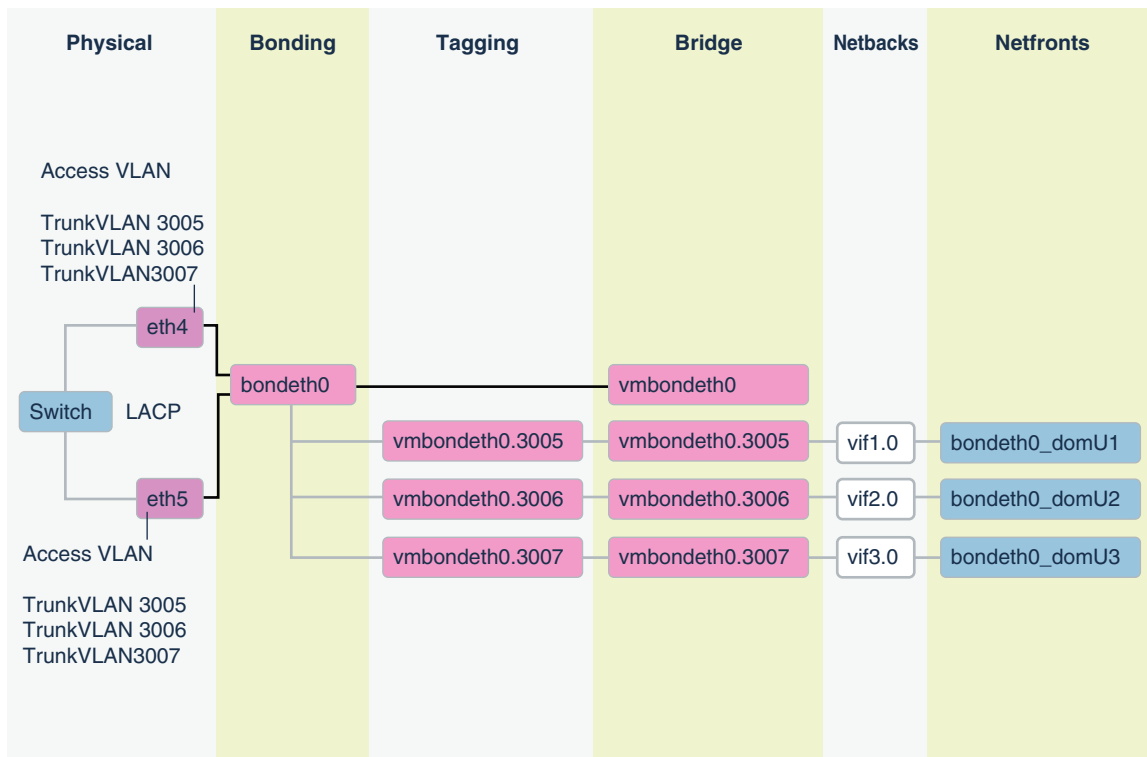
1. In the KVM host, eth slave interfaces (for example, eth1 and eth2, or eth4 and eth5) that allow access to the guest client network defined in OEDA are discovered, configured, and brought up, but no IP is assigned.
2. In the KVM host, bondeth0 master interface is configured and brought up, but no IP is assigned.
3. In the KVM host, bridge interface vmbondeth0 is configured, but no IP is assigned.
4. In the KVM host, one virtual backend interface (VIF) per guest that maps to that particular guest's bondeth0 interface is configured and brought up, but no IP is assigned. These VIFs are configured on top of the bridge interface vmbondeth0, and the mapping between the KVM host VIF interface and its corresponding guest interface bondeth0 is defined in the guest configuration file called `vm.cfg`, located in `/EXAVMIMAGES/GuestImages/guest name`.

For default installations, a single bondeth0 and a corresponding vmbondeth0 bridge interface is configured in the KVM host as described above. This bondeth0 interface is based on the default Access VLAN. The ports on the switch used by the slave interfaces making up bondeth0 are configured for Access VLAN.

### Using VLAN Tagging

If there is a need for virtual deployments on Exadata to access additional VLANs on the client network, such as enabling network isolation across guests, then 802.1Q-based VLAN tagging is a solution. The following figure shows a client network configuration with VLAN tagging.

**Figure 6-2 NIC Layout for Oracle Virtual Environments with VLAN Tagging**



**Note:**

Commencing with the March 2020 OEDA release, the bridge names now have the form `vmbethXY.VLANID`, where *X* and *Y* are the numeric identifiers associated with the slave interface, and *VLANID* is the VLAN ID.

This avoids a potential naming conflict that could previously occur in some cases.

For example, under the new naming scheme the bridges in the previous diagram would be named `vmbeth45.3005`, `vmbeth45.3006`, and `vmbeth45.3007`.

For instructions on how to manually configure tagged VLAN interfaces in conjunction with Oracle Linux KVM, see My Oracle Support note 2710712.1.

**Related Topics**

- [Implementing Tagged VLAN Interfaces in KVM based Oracle VM Environments on Exadata \(Doc ID 2710712.1\)](#)

## 6.22 Using Exadata Secure RDMA Fabric Isolation with Oracle Linux KVM

This topic describes the implementation of Exadata Secure RDMA Fabric Isolation in conjunction with Oracle Linux KVM.

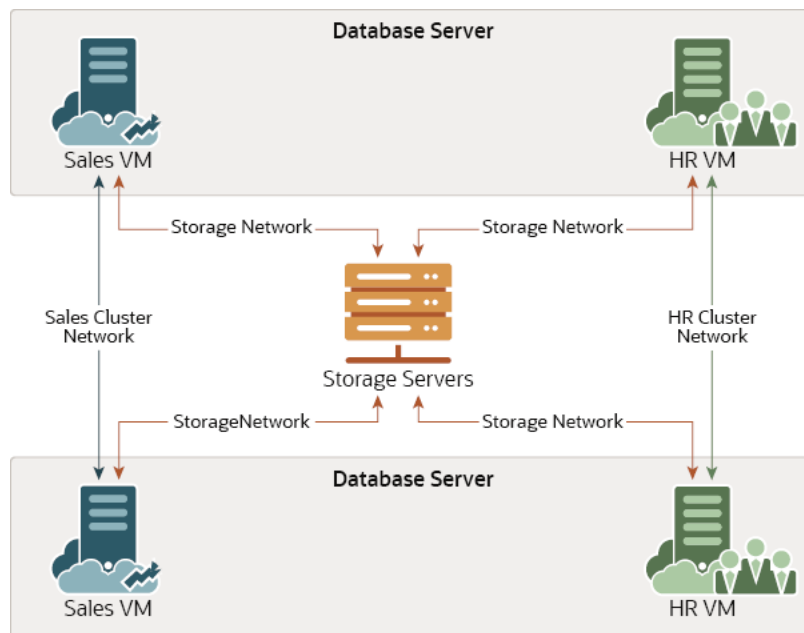
Secure Fabric enables secure consolidation and strict isolation between multiple tenants on Exadata Database Machine. Each tenant resides in their own dedicated virtual machine (VM) cluster, using database server VMs running on Oracle Linux KVM.

With Secure Fabric, each database cluster uses a dedicated network partition and VLAN ID for cluster networking between the database servers, which supports Oracle RAC inter-node messaging. In this partition, all of the database servers are full members. They can communicate freely within the partition but cannot communicate with database servers in other partitions.

Another partition, with a separate VLAN ID, supports the storage network partition. The storage servers are full members in the storage network partition, and every database server VM is also a limited member. By using the storage network partition:

- Each database server can communicate with all of the storage servers.
- Each storage server can communicate with all of the database servers that they support.
- Storage servers can communicate directly with each other to perform cell-to-cell operations.

**Figure 6-3 Secure Fabric Network Partitions**



To support the cluster network partition and the storage network partition, each database server VM is plumbed with 4 virtual interfaces:

- `clre0` and `clre1` support the cluster network partition.
- `stre0` and `stre1` support the storage network partition.

Corresponding `stre0` and `stre1` interfaces are also plumbed on each storage server.

On each server, the RoCE network interface card acts like a switch on the hypervisor, which performs VLAN tag enforcement. Since this is done at the KVM host level, cluster isolation cannot be bypassed by any software exploits or misconfiguration on the database server VMs.

In this release, you can only enable Secure Fabric as part of the initial system deployment using Oracle Exadata Deployment Assistant (OEDA). You cannot enable Secure Fabric on an existing system without wiping the system and re-deploying it using OEDA. When enabled, Secure Fabric applies to all servers and clusters that share the same RoCE Network Fabric.

#### Related Topics

- Using Exadata Secure RDMA Fabric Isolation

## 6.23 Adding a Bonded Network Interface to an Oracle Linux KVM Guest

Use this procedure to add a bonded network interface to an existing Oracle Linux KVM guest.

For example, you can use this procedure to add a backup network to an Oracle Linux KVM guest.

To add a bonded network interface to an Oracle Linux KVM guest:

1. On the KVM host, create a bonded network bridge.

Use the following `vm_maker --add-bonded-bridge` command:

```
vm_maker --add-bonded-bridge bond-name --first-slave ethX --
second-slave ethY
```

In the command:

- `bond-name` specifies the name of the bonded network interface. The recommended naming convention uses the string `vmbeth` concatenated with the numbers that identify the network interfaces being bonded. For example, when bonding `eth5` and `eth6`, the recommended bond name is `vmbeth56`.
  - `ethX` and `ethY` specify the KVM host network interfaces that are being bonded.
2. On the KVM host, associate the bonded network bridge with the guest.

Use the following `vm_maker --allocate-bridge` command:

```
vm_maker --allocate-bridge bond-name --domain guest-name
```

In the command:



- *bond-name* specifies the name of the bonded network bridge created in the previous step.
- *guest-name* specifies the name of the guest to which you want to add the bonded network interface.

The output from the `vm_maker --allocate-bridge` command includes a series of additional manual steps that must be performed to complete the procedure.

3. Perform the additional manual steps outlined in the output from the previous `vm_maker --allocate-bridge` command.

In the first manual step, you must select a unique network interface name to use inside the guest. This guest interface maps to the bridged network interface on the KVM host. You can use your naming convention for this interface name, or you can choose to carry through the interface name from the KVM host. Whatever your choice, ensure that the `NAME="guest-interface"` entry in manual step 2 is adjusted accordingly.

The following is an example of the manual steps.

#### Note:

Your procedure will be specific to your environment. So, ensure that you perform the additional manual steps outlined in your output from the `vm_maker --allocate-bridge` command. Do not copy the commands from the following example.

```
[INFO] Please perform the following manual steps:
[INFO] 1. Determine a unique network interface name within the domain to
[INFO] which
[INFO] you are attaching this interface. Typically bonded
[INFO] interfaces are named bondeth<number>, for example 'bondeth1',
[INFO] and
[INFO] non-bonded interfaces are named eth<number>, for example 'eth2'.
[INFO] The name must be unique within the domain. In the example below
[INFO] the name 'bondeth0' has been chosen.
[INFO] 2. Add the following line to the file
[INFO] '/etc/udev/rules.d/70-persistent-net.rules' within the domain:
[INFO] SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*",
[INFO] KERNELS=="0000:00:PCI-slot-number.0", ATTR{type}=="1", NAME="guest-
[INFO] interface"
[INFO] You MUST execute steps 4 and 5! You CANNOT simply reboot from
[INFO] within
[INFO] the domain.
[INFO] 4. vm_maker --stop-domain guest-name
[INFO] 5. vm_maker --start-domain guest-name
[INFO] 6. Login to the domain and run the following command:
[INFO] /opt/oracle.cellos/ipconf.pl -nocodes
[INFO] This command will ask you for configuration information for this
[INFO] new interface.
[INFO] NOTE: if you have more than one interface to add to this domain,
[INFO] please execute the instructions above, and then call this command
[INFO] again.
```

```
[INFO] The domain must be stopped and started between invocations
of this
[INFO] command.
```

## 6.24 Using Oracle EXAchk in Oracle Linux KVM Environments

Oracle EXAchk version 12.1.0.2.2 and higher supports virtualization on Exadata Database Machine.

- [Running Oracle EXAchk in Oracle Linux KVM Environments](#)  
To perform the complete set of Oracle EXAchk audit checks in an Exadata Database Machine Oracle Linux KVM environment, Oracle EXAchk must be installed in and run from multiple locations.
- [Audit Checks Performed by Oracle EXAchk](#)  
Oracle EXAchk runs different audit checks on the KVM host and the guests.
- [Oracle EXAchk Command Line Options for Exadata Database Machine](#)

### 6.24.1 Running Oracle EXAchk in Oracle Linux KVM Environments

To perform the complete set of Oracle EXAchk audit checks in an Exadata Database Machine Oracle Linux KVM environment, Oracle EXAchk must be installed in and run from multiple locations.

1. Run Oracle EXAchk from one KVM host.
2. Run Oracle EXAchk from one guest in each Oracle Real Application Clusters (Oracle RAC) cluster running in Oracle Linux KVM.

For example, an Exadata Database Machine Quarter Rack with two database servers containing 4 Oracle RAC clusters (2 nodes per cluster for a total of 8 guests across both database servers) requires running Oracle EXAchk five separate times, as follows:

1. Run Oracle EXAchk in the first guest for the first cluster.
2. Run Oracle EXAchk in the first guest for the second cluster.
3. Run Oracle EXAchk in the first guest for the third cluster.
4. Run Oracle EXAchk in the first guest for the fourth cluster.
5. Run Oracle EXAchk in the first KVM host.

### 6.24.2 Audit Checks Performed by Oracle EXAchk

Oracle EXAchk runs different audit checks on the KVM host and the guests.

When you install and run Oracle EXAchk on the KVM host, it performs the following hardware and operating system level checks:

- Database servers (KVM hosts)
- Storage servers
- RDMA Network Fabric

- RDMA Network Fabric switches

When you install and run Oracle EXAchk on the guest, it performs operating system checks for guests, and checks for Oracle Grid Infrastructure and Oracle Database.

### 6.24.3 Oracle EXAchk Command Line Options for Exadata Database Machine

Oracle EXAchk requires no special command line options. It automatically detects that it is running in an Exadata Database Machine Oracle Linux KVM environment. However, you can use command line options to run Oracle EXAchk on a subset of servers or switches. Oracle EXAchk automatically detects whether it is running in a KVM host or guest and performs the applicable audit checks. For example, in the simplest case, you can run Oracle EXAchk with no command line options:

```
./exachk
```

When Oracle EXAchk is run in the KVM host, it performs audit checks on all database servers, storage servers, and RDMA Network Fabric switches accessible through the RDMA Network Fabric network.

To run Oracle EXAchk on a subset of servers or switches, use the following command line options:

#### Options

- `-clusternodes`: Specifies a comma-separated list of database servers.
- `-cells`: Specifies a comma-separated list of storage servers.

#### Example 6-1 Running Oracle EXAchk on a Subset of Nodes and Switches

For example, for an Exadata Database Machine Full Rack where only the first Quarter Rack is configured for virtualization, but all components are accessible through the RDMA Network Fabric network, you can run a command similar to the following from the database server `dm01adm01`:

```
./exachk -clusternodes dm01adm01,dm01adm02
 -cells dm01celadm01,dm01celadm02,dm01celadm03
```

#### Related Topics

- Compliance Framework (Oracle ORAchk and Oracle EXAchk) Command-Line Options

# 7

## VM\_MAKER Command Reference

The `vm_maker` utility provides commands to create, destroy, and otherwise manipulate Oracle Linux KVM guests.

- [About the `vm\_maker` Utility](#)  
You can use the `vm_maker` command-line utility to administer an Oracle Linux KVM guest.
- [attach Command](#)  
Use the `--attach` command to attach a disk image to a guest.
- [add-bonded-bridge Command](#)  
Use the `--add-bonded-bridge` command to add a bridge over a bonded Ethernet interface.
- [add-single-bridge Command](#)  
Use the `--add-single-bridge` command to add a bridge over a single Ethernet interface.
- [allocate-bridge Command](#)  
Use the `--allocate-bridge` command to associate a network bridge with a guest.
- [autostart Command](#)  
Use the `--autostart` command to manage automatic starting of guests.
- [boot-from-hd Command](#)  
Use the `--boot-from-hd` command to configure a guest to boot using its primary boot device after a previous `--boot-from-iso` command.
- [boot-from-iso Command](#)  
Use the `--boot-from-iso` command to configure a guest to boot from a specified diagnostic ISO file.
- [check Command](#)  
Use the `--check` command to check and display the system configuration status.
- [cleanup-ip Command](#)  
Use the `--cleanup-ip` command to remove and clean up a network interface plumbed by the `--setup-ip` command.
- [configure-system Command](#)  
Use the `--configure-system` command to configure the system.
- [create Command](#)  
Use the `--create` command to create a disk image.
- [detach Command](#)  
Use the `--detach` command to detach a disk image from a guest.
- [dumpxml Command](#)  
Use the `--dumpxml` command to send the XML definition of the specified guest to standard output.
- [list Command](#)  
Use the `--list` command to display information about guests.

- **list-domains Command**  
Use the `--list-domains` command to list all guests.
- **network-discovery Command**  
Use the `--network-discovery` command to identify the network for the Oracle RAC single client access name (SCAN) and other interfaces.
- **reboot Command**  
Use the `--reboot` command to restart a specific guest.
- **remove-bridge Command**  
Use the `--remove-bridge` command to remove a network bridge.
- **remove-domain Command**  
Use the `--remove-domain` command to stop and remove a specific guest or all guests.
- **set Command**  
Use the `--set` command to configure vCPUs or memory for a guest.
- **setup-ip Command**  
Use the `--setup-ip` command to plumb a network interface on a KVM host.
- **start-domain Command**  
Use the `--start-domain` command to create and start an Oracle Linux KVM guest.
- **stop-domain Command**  
Use the `--stop-domain` command to stop the specified Oracle Linux KVM guest.
- **validate-ip Command**  
Use the `--validate-ip` command to validate specified network details for an Oracle Linux KVM guest.
- **vhostmd Command**  
Use the `--vhostmd` command to enable, disable, or show the status of the `vhostmd` service.

## 7.1 About the `vm_maker` Utility

You can use the `vm_maker` command-line utility to administer an Oracle Linux KVM guest.

The `vm_maker` command-line utility is located on each Oracle Linux KVM host at `/opt/exadata_ovm/vm_maker`.

### Syntax

The general syntax for the `vm_maker` command is shown here. The following topics describe the individual commands.

```
vm_maker --command --option [--option ...]
```

Some commands or options take arguments, such as the guest name. Only enough characters to uniquely identify the guest name are required.

 **Note:**

Oracle Exadata Deployment Assistant (OEDA) contains the OEDA Command Line Interface (OEDACLI), which provides an alternative command-line utility to administer Oracle Linux KVM guests. A primary difference is that OEDACLI can orchestrate operations across multiple guests, while `vm_maker` is limited to operating on the Oracle Linux KVM host where it resides. Consider your requirements before deciding which utility to use.

**Related Topics**

- OEDA Command Line Interface

## 7.2 attach Command

Use the `--attach` command to attach a disk image to a guest.

**Syntax**

```
vm_maker --attach --disk-image image_name --domain guest_name [--no-reflink]
```

**Options**

- `--disk-image`: The name of the disk image you want to attach.
  - If you specify a disk image name that is not a fully-qualified file name (for example, `shared.img`), then the shared disk image must already exist in `/EXAVMIMAGES` on the KVM host. In this case, the shared disk image file is copied into the guest-specific directory at `/EXAVMIMAGES/GuestImages/guest_name`. By default, a reflink copy is created in the guest-specific directory, which references the shared image file in `/EXAVMIMAGES`. To attach a shared disk image file without using a reflink, add the `--no-reflink` option. In this case, the image file in the guest-specific directory is a discrete copy of the image file in `/EXAVMIMAGES`.
  - If you specify a fully-qualified disk image name in a guest-specific directory (for example, `/EXAVMIMAGES/GuestImages/guest01.example.com/pv2_vgexadb.img`), then the disk image file must already exist in the specified location. In this case, the disk image file can only be attached to the guest identified in the file path (for example, `guest01.example.com`), and no additional copy occurs.
- `--domain`: The name of the guest.
- `--no-reflink`: Use this option to create a discrete copy in the guest-specific image directory of the shared image file (not a reflink). This option is ignored for non-shared image files.

**Usage Notes**

- A KVM guest can have a maximum of 112 disks attached.

## 7.3 add-bonded-bridge Command

Use the `--add-bonded-bridge` command to add a bridge over a bonded Ethernet interface.

### Syntax

```
vm_maker --add-bonded-bridge bridge_name --first-slave interface1 --
second-slave interface2
[--vlan vlanid] [--bond-mode { active-backup | lacp }]
```

### Options

- *bridge\_name*: The name to use for the new bridge. The bridge name must begin with `vm` and be followed by an interface name. For example, `vmeth1`.
- `--first-slave`: The first part of the interface pair that comprises the bonded bridge.
- `--second-slave`: The second part of the interface pair that comprises the bonded bridge.
- `--vlan`: Optionally, specifies the VLAN ID to use for the bonded bridge.
- `--bond-mode`: Optionally, specifies the bonding mode to use for the bonded bridge:
  - `active-backup` configures a bonded bridge where one network interface handles all of the network traffic, and the other network interface only becomes active if the first one fails. This mode provides fault tolerance and is the default setting.
  - `lacp` configures a bonded bridge where both network interfaces share the workload using Link Aggregation Control Protocol (LACP). This mode provides fault tolerance and increases the network bandwidth. Also known as Dynamic Link Aggregation, this bonding mode requires a switch that supports IEEE 802.3ad dynamic links.

## 7.4 add-single-bridge Command

Use the `--add-single-bridge` command to add a bridge over a single Ethernet interface.

### Syntax

```
vm_maker --add-single-bridge bridge_name [--vlan vlanid]
```

### Options

- *bridge\_name*: The name to use for the new bridge. The bridge name must begin with `'vm'` followed by an interface name, for example, `vmeth1`.
- `--vlan`: Optionally, specifies the VLAN ID to use for the bridge.

## 7.5 allocate-bridge Command

Use the `--allocate-bridge` command to associate a network bridge with a guest.

### Syntax

```
vm_maker --allocate-bridge bridge-name --domain guest-name
```

### Options

- *bridge-name*: The name of the previously created network bridge that you want to associate with the specified guest.
- *guest-name*: The name of the guest to which you want to add the network bridge.

### Usage Notes

The output from the `--allocate-bridge` command includes a series of additional manual steps that must be performed to complete the procedure.

In the first manual step, you must select a unique network interface name to use inside the KVM guest. This guest interface maps to the bridged network interface on the KVM host. You can use your naming convention for this interface name, or you can choose to carry through the interface name from the KVM host. Whatever your choice, ensure that the `NAME="guest-interface"` entry is adjusted accordingly in step 2.

The following is an example of the manual steps.

 **Note:**

Your procedure will be specific to your environment. So, ensure that you perform the additional manual steps outlined in your output from the `--allocate-bridge` command. Do not copy the commands from the following example.

```
[INFO] Please perform the following manual steps:
[INFO] 1. Determine a unique network interface name within the domain to
which
[INFO] you are attaching this interface. Typically bonded
[INFO] interfaces are named bondeth<number>, for example 'bondeth1', and
[INFO] non-bonded interfaces are named eth<number>, for example 'eth2'.
[INFO] The name must be unique within the domain. In the example below
[INFO] the name 'bondeth0' has been chosen.
[INFO] 2. Add the following line to the file
[INFO] '/etc/udev/rules.d/70-persistent-net.rules' within the domain:
[INFO] SUBSYSTEM=="net", ACTION=="add", DRIVERS=="*",
KERNELS=="0000:00:PCI-slot-number.0", ATTR{type}=="1", NAME="guest-interface"
[INFO] You MUST execute steps 4 and 5! You CANNOT simply reboot from within
[INFO] the domain.
[INFO] 4. vm_maker --stop-domain guest-name
[INFO] 5. vm_maker --start-domain guest-name
[INFO] 6. Login to the domain and run the following command:
```



```
[INFO] /opt/oracle.cellos/ipconf.pl -nocodes
[INFO] This command will ask you for configuration information for
this
[INFO] new interface.
[INFO] NOTE: if you have more than one interface to add to this domain,
[INFO] please execute the instructions above, and then call this
command again.
[INFO] The domain must be stopped and started between invocations of
this
[INFO] command.
```

## 7.6 autostart Command

Use the `--autostart` command to manage automatic starting of guests.

### Syntax

```
vm_maker --autostart guest_name { --enable | --disable }
```

### Options

- *guest\_name*: The name of the guest to configure.
- `--enable`: Configure the guest to start automatically when the KVM host is started.
- `--disable`: Configure the guest so that it does not start automatically when the KVM host is started.

## 7.7 boot-from-hd Command

Use the `--boot-from-hd` command to configure a guest to boot using its primary boot device after a previous `--boot-from-iso` command.

### Syntax

```
vm_maker --boot-from-hd --domain guest-name
```

### Options

- *guest-name*: The name of the guest that you want to configure.

### Usage Notes

The `--boot-from-hd` command configures the guest to boot using its primary boot device. You must start the domain separately by using the `--start-domain` command.

## 7.8 boot-from-iso Command

Use the `--boot-from-iso` command to configure a guest to boot from a specified diagnostic ISO file.

## Syntax

```
vm_maker --boot-from-iso ISO-file --domain guest-name
```

## Options

- *ISO-file*: The name of the diagnostic ISO that you want to use to boot the specified guest.
- *guest-name*: The name of the guest that you want to boot using the specified ISO file.

## Usage Notes

The `--boot-from-iso` command configures the guest to boot using the specified ISO file.

The output from the `--boot-from-iso` command includes additional commands to boot the guest and configure the network while in diagnostic mode.

The output also contains commands to stop the domain and reconfigure it to boot using its primary boot device. Use these commands when you are finished using the guest in diagnostic mode.

For example:

```
vm_maker --boot-from-iso /root/home/diagnostics.iso --domain dm01vm01
[INFO] Running 'virsh undefine dm01vm01.example.com'...
[INFO] Running 'virsh define /var/log/exadatatmp/
dm01vm01.example.com.xml.new.357b'...
[INFO] The domain 'dm01vm01.example.com' is ready for booting.
[INFO] Run the following command to boot from the diagnostic iso:
[INFO]
[INFO] virsh start dm01vm01.example.com --console
[INFO]
[INFO] If network is needed to be setup on the VM, run
[INFO] setup_management.sh from the console after the guest has booted.
[INFO]
[INFO] When finished, run the following commands to restore
[INFO] the domain to boot from its hard disk:
[INFO]
[INFO] vm_maker --stop-domain dm01vm01.example.com --force
[INFO] vm_maker --boot-from-hd --domain dm01vm01.example.com
```

## 7.9 check Command

Use the `--check` command to check and display the system configuration status

## Syntax

```
vm_maker --check
```

## 7.10 cleanup-ip Command

Use the `--cleanup-ip` command to remove and clean up a network interface plumbed by the `--setup-ip` command.

### Syntax

```
vm_maker --cleanup-ip network_interface [--vlan vlanid]
```

### Options

- `network_interface`: Specifies the name of the network interface to clean up.
- `--vlan`: Optionally, specifies the VLAN ID used with the specified network interface.

### Example 7-1 Cleaning Up a Network Interface

The following shows an example of using the `--cleanup-ip` command to remove and clean up a network interface plumbed by the `--setup-ip` command.

```
ping 192.0.2.1
PING 192.0.2.1 (192.0.2.1) 56(84) bytes of data.
From 10.32.106.12 icmp_seq=1 Destination Host Unreachable
^C
--- 192.0.2.1 ping statistics ---
2 packets transmitted, 0 received, +1 errors, 100% packet loss, time
999ms
/opt/exadata_ovm/vm_maker --setup-ip 192.0.2.6 --interface eth3 --
netmask 255.255.255.128 --gateway 192.0.2.1 --vlan 144
[INFO] Interface eth3 currently down
[INFO] Bringing up interface eth3
[INFO] Interface eth3.144 is plumbed.
ping 192.0.2.1
PING 192.0.2.1 (192.0.2.1) 56(84) bytes of data.
64 bytes from 192.0.2.1: icmp_seq=2 ttl=255 time=0.317 ms
64 bytes from 192.0.2.1: icmp_seq=3 ttl=255 time=0.302 ms
64 bytes from 192.0.2.1: icmp_seq=4 ttl=255 time=0.307 ms
^C
--- 192.0.2.1 ping statistics ---
4 packets transmitted, 3 received, 25% packet loss, time 3077ms
rtt min/avg/max/mdev = 0.302/0.308/0.317/0.021 ms
ip rule
0: from all lookup local
32757: from 192.0.2.6 to 192.0.2.1 iif eth3.144 lookup main
32758: from all to 10.32.96.5 lookup 220
32759: from 10.32.96.5 lookup 220
32760: from all to 10.32.106.0/23 lookup 210
32761: from 10.32.106.0/23 lookup 210
32762: from all to 192.168.0.4 lookup 181
32763: from 192.168.0.4 lookup 181
32764: from all to 192.168.0.3 lookup 180
32765: from 192.168.0.3 lookup 180
32766: from all lookup main
```

```

32767: from all lookup default
ip a s | grep eth3
5: eth3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group
default qlen 1000
67: eth3.144@eth3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue
state UP group default qlen 1000
inet 192.0.2.6/25 scope global eth3.144
/opt/exadata_ovm/vm_maker --cleanup-ip eth3 --vlan 144
[INFO] Deleting routing rules.
[INFO] Flushing ip-address from interface.
ip a s | grep eth3
5: eth3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group
default qlen 1000

```

## 7.11 configure-system Command

Use the `--configure-system` command to configure the system.

### Syntax

```
vm_maker --configure-system [--no-build-base-image]
```

### Usage Notes

This command creates `/EXAVMIMAGES` and creates the virtual functions. If you include the `--no-build-base-image` option, then `vm_maker` does not check for the existence of the base image.

## 7.12 create Command

Use the `--create` command to create a disk image.

### Syntax

```
vm_maker --create --disk-image image_name [--size size] [--filesystem
filesystem_type]
[--from-zip zipfile] [--attach --domain guest_name [--no-reflink]]
```

### Options

- `--disk-image`: Create a disk image using the specified image name:
  - If you specify a disk image name that is not a fully-qualified file name (for example, `shared.img`), a shared disk image file is created in `/EXAVMIMAGES` on the KVM host. In this case, the disk image file can be attached to any guest, either by adding the `--attach` option or by using a separate `vm_maker --attach` command.
  - If you specify a fully-qualified disk image name in a guest-specific directory (for example, `/EXAVMIMAGES/GuestImages/guest01.example.com/pv2_vgexadb.img`), a guest-specific disk image file is created in the specified location. In this case, the disk image file can only be attached to the guest identified in the file path (for example, `guest01.example.com`).

- `--size`: Optionally specifies the size of the image. If not specified, the default size is 50 GB.
- `--filesystem`: Specifies the filesystem type to use for the disk image. Allowed filesystem types are: `ext3`, `ext4`, and `xf`s. The default is `xf`s.
- `--from-zip`: Indicates the disk image should be created from the specified ZIP file.
- `--attach`: Use this option to directly attach the disk image to a guest after creation. If you specify this option, then you must also specify the `--domain` option.
  - If you are attaching a shared disk image (created in `/EXAVMIMAGES`), then a copy of the shared disk image file is created in the guest-specific directory at `/EXAVMIMAGES/GuestImages/guest_name`. By default, a reflink copy is created in the guest-specific directory, which references the shared image file in `/EXAVMIMAGES`. To attach a shared disk image file without using a reflink, add the `--no-reflink` option. In this case, the new disk image in the guest-specific directory is a discrete copy of the image file in `/EXAVMIMAGES`.
  - No additional copy occurs when you attach a non-shared image file that resides in a guest-specific directory.
- `--domain`: Use this option to specify which guest to attach the disk image to after creation.
- `--no-reflink`: Use this option to create a discrete copy in the guest-specific image directory of a shared image file (not a reflink). This option is ignored for non-shared image files.

### Usage Notes

- A KVM guest can have a maximum of 112 disks attached.

## 7.13 detach Command

Use the `--detach` command to detach a disk image from a guest.

### Syntax

```
vm_maker --detach --disk-image image_name --domain guest_name [--delete]
```

### Options

- `--disk image`: The name of the disk image that you want to detach.
- `--domain`: The name of the guest from which you want to detach the disk image.
- `--delete`: Optionally specifies to delete the disk image from `/EXAVMIMAGES`. This option succeeds only if the disk image is not in use by any other guest. Otherwise, the command fails and reports an error.

## 7.14 dumpxml Command

Use the `--dumpxml` command to send the XML definition of the specified guest to standard output.

### Syntax

```
vm_maker --dumpxml guest_name
```

### Usage Notes

You can save the data by piping the output to a file, for example:

```
vm_maker --dumpxml guest1.example.com > guest.xml
```

## 7.15 list Command

Use the `--list` command to display information about guests.

### Syntax

To view information about disk images:

```
vm_maker --list --disk-image [--domain guest_name]
```

To view information about guests:

```
vm_maker --list --domain [guest_name [--detail]]
```

To view information about virtual CPU configuration for guests:

```
vm_maker --list --vcpu [--domain guest_name]
```

To view information about memory configuration for guests:

```
vm_maker --list --memory [--domain guest_name]
```

### Options

- `--disk-image`: List all shared disk images in `/EXAVMIMAGES`.
- `--domain`: List all guests and their status. If you include the guest name, then the information displayed is limited to that guest.
- `--vcpu`: List the virtual CPU usage for the entire system.
- `--memory`: List the memory usage for the entire system.

### Optional Arguments

- `--domain guest_name`: Limits the information displayed to a single guest.
- `--detail`: Enables detailed output.

## 7.16 list-domains Command

Use the `--list-domains` command to list all guests.

### Syntax

```
vm_maker --list-domains
```

## 7.17 network-discovery Command

Use the `--network-discovery` command to identify the network for the Oracle RAC single client access name (SCAN) and other interfaces.

### Syntax

```
vm_maker --network-discovery XML-config-file
```

### Options

*XML-config-file*: The full file name of the guest XML configuration file. For example, `/EXAVMIMAGES/conf/domain.xml`.

## 7.18 reboot Command

Use the `--reboot` command to restart a specific guest.

### Syntax

```
vm_maker --reboot guest_name
```

## 7.19 remove-bridge Command

Use the `--remove-bridge` command to remove a network bridge.

### Syntax

```
vm_maker --remove-bridge bridge_name [--vlan vlanid] [--force]
```

### Options

- *bridge\_name*: The name of the bridge to remove. The bridge name starts with 'vm' and is followed by an interface name, for example, `vmeth1`.
- `--vlan`: Optionally, specifies the VLAN ID for the bridge.
- `--force`: Use this option to forcefully remove the bridge.

### Usage Notes

A network bridge can be removed only if the bridge is not being used by the KVM host or other guests.

## 7.20 remove-domain Command

Use the `--remove-domain` command to stop and remove a specific guest or all guests.

### Syntax

```
vm_maker --remove-domain { guest_name | --all [--reset-kvmhost] }
```

### Options

- *guest\_name*: The name of a specific guest to shutdown and then delete. You can use `--all` instead of a single guest name if you want to remove all guests.
- `--reset-kvmhost`: Resets the system to the state when the KVM host had no guests. This option can only be used when you specify `--all` for the guest name.

### Usage Notes

When you use the `--remove-domain` command `vm_maker` forces a shutdown of the guest and deletes it. If `--all` is specified for the guest name, then the command shuts down and deletes all guests and intermediate image files. Underlying network bridges are removed if they are not being used anymore.

If you include the option `--reset-kvmhost`, then the system is reset to the state it was in before the guests were created.

## 7.21 set Command

Use the `--set` command to configure vCPUs or memory for a guest.

### Syntax

To configure the number of virtual CPUs for a guest:

```
vm_maker --set --vcpu vcpu [--domain guest_name [--config] [--force]]
```

To configure the memory for a guest:

```
vm_maker --set --memory size --domain guest_name [--restart-domain] [--force]
```

To configure the MAC for specified Ethernet interface:

```
vm_maker --set --mac mac_address --interface interface_name [--restart-domain]
```



## Options

- `--vcpu`: Specifies the number of virtual CPUs for a guest.
- `--memory`: Sets the memory size for the specified guest. Memory can be specified using K, M, or G, for example, 40G or 25000M.
- `--mac`: Sets the media access control (MAC) address for the specified network interface.
- `--domain`: When used with `--vcpu` or `--memory`, this argument indicates that the setting applies to a single guest. Otherwise the setting applies to all guests. You must use `--domain` when changing the memory size.
- `--config`: Indicates that the number of virtual CPUs should not be changed until you restart the guest. This option can only be used when modifying a single guest.
- `--interface`: The Ethernet network interface to configure.
- `--restart-domain`: Use this option to automatically restart the guest after `vm_maker` changes the configuration.
- `--force`: Forces the new setting to be implemented. This option can only be used when modifying a single guest.

## Usage Notes

- Changing the number of virtual CPUs for a guest happens immediately by default, without restarting the guest.
- When changing the memory, you must restart the guest before the settings take effect.

## 7.22 setup-ip Command

Use the `--setup-ip` command to plumb a network interface on a KVM host.

### Syntax

```
vm_maker --setup-ip ip_address --interface network_interface
--netmask netmask --gateway gateway_address [--vlan vlanid]
```

```
vm_maker --setup-ip ip_address --netmask netmask
--discover --gateway gateway_address [--vlan vlanid]
```

### Options

- `ip_address`: Specifies the IP address to use for the network interface.
- `--interface`: Specifies the name of the network interface to use.
- `--netmask`: Specifies the IP netmask to use for the interface.
- `--gateway`: Specifies the IP network gateway address to use for the interface.
- `--vlan`: Optionally, specifies the VLAN ID to use for the interface.

## Usage Notes

Use the `--setup-ip` command to plumb a network interface on a KVM host, which you can use for testing and troubleshooting purposes. Use the `--cleanup-ip` command to remove and clean up the network interface.

### Example 7-2 Setting Up a Network Interface

The following shows an example of using the `--setup-ip` command plumb a network interface and use it to access the network gateway.

```
ping 192.0.2.1
PING 192.0.2.1 (192.0.2.1) 56(84) bytes of data.
From 10.32.106.12 icmp_seq=1 Destination Host Unreachable
^C
--- 192.0.2.1 ping statistics ---
2 packets transmitted, 0 received, +1 errors, 100% packet loss, time 999ms
/opt/exadata_ovm/vm_maker --setup-ip 192.0.2.6 --interface eth3 --netmask
255.255.255.128 --gateway 192.0.2.1 --vlan 144
[INFO] Interface eth3 currently down
[INFO] Bringing up interface eth3
[INFO] Interface eth3.144 is plumbed.
ping 192.0.2.1
PING 192.0.2.1 (192.0.2.1) 56(84) bytes of data.
64 bytes from 192.0.2.1: icmp_seq=2 ttl=255 time=0.317 ms
64 bytes from 192.0.2.1: icmp_seq=3 ttl=255 time=0.302 ms
64 bytes from 192.0.2.1: icmp_seq=4 ttl=255 time=0.307 ms
^C
--- 192.0.2.1 ping statistics ---
4 packets transmitted, 3 received, 25% packet loss, time 3077ms
rtt min/avg/max/mdev = 0.302/0.308/0.317/0.021 ms
ip rule
0: from all lookup local
32757: from 192.0.2.6 to 192.0.2.1 iif eth3.144 lookup main
32758: from all to 10.32.96.5 lookup 220
32759: from 10.32.96.5 lookup 220
32760: from all to 10.32.106.0/23 lookup 210
32761: from 10.32.106.0/23 lookup 210
32762: from all to 192.168.0.4 lookup 181
32763: from 192.168.0.4 lookup 181
32764: from all to 192.168.0.3 lookup 180
32765: from 192.168.0.3 lookup 180
32766: from all lookup main
32767: from all lookup default
ip a s | grep eth3
5: eth3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group
default qlen 1000
67: eth3.144@eth3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue
state UP group default qlen 1000
 inet 192.0.2.6/25 scope global eth3.144
/opt/exadata_ovm/vm_maker --cleanup-ip eth3 --vlan 144
[INFO] Deleting routing rules.
[INFO] Flushing ip-address from interface.
ip a s | grep eth3
```

```
5: eth3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP
group default qlen 1000
```

## 7.23 start-domain Command

Use the `--start-domain` command to create and start an Oracle Linux KVM guest.

### Syntax

```
vm_maker --start-domain { guest_name | XML-config-file | --all } [--no-discovery]
[--keep] [--locked] [--ssh-key key_file] [--console]
```

### Options

- *guest\_name*: The name of an existing guest to start.
- *XML-config-file*: The full file name of the XML configuration file for creating a new guest.
- `--all`: Starts all existing domains on the host.
- `--no-discovery`: Optionally specifies to skip network discovery and bridge creation when creating a new guest.
- `--keep`: Indicates that the guest should not be removed even if the first boot fails.
- `--locked`: If specified in conjunction with the `--ssh-key` option, password-based authentication is disabled for the `root` user, and access to the guest requires key-based authentication. Otherwise, the option forces users to choose a new password when they log in for the first time.
- `--ssh-key`: Optionally specifies the location of a file containing an SSH public key, which is used for authenticating the `root` user.
- `--console`: Optionally causes `vm_maker` to display console messages when starting existing guests.

## 7.24 stop-domain Command

Use the `--stop-domain` command to stop the specified Oracle Linux KVM guest.

### Syntax

```
vm_maker --stop-domain { guest_name | --all }
```

### Options

- *guest\_name*: The name of the guest to stop.
- `--all`: Use this option to stop all guests instead of a specific one.

## 7.25 validate-ip Command

Use the `--validate-ip` command to validate specified network details for an Oracle Linux KVM guest.

### Syntax

```
vm_maker --validate-ip ip_address --netmask netmask
--interface network_interface --gateway gateway_address [--vlan vlanid]
```

```
vm_maker --validate-ip ip_address --netmask netmask
--discover --gateway gateway_address [--vlan vlanid]
```

### Options

- `ip_address`: Specifies the IP address to use in the validation.
- `--netmask`: Specifies the IP netmask to use in the validation.
- `--gateway`: Specifies the IP network gateway address to use in the validation.
- `--vlan`: Optionally, specifies the VLAN ID to use in the validation.
- `--interface`: Instructs `vm_maker` to perform interface validation and specifies the network interface to use in the validation.
- `--discover`: Instructs `vm_maker` to perform discovery validation.

### Usage Notes

Use the `--validate-ip` command to verify if the provided network details are valid to use for a KVM guest. It can be used for troubleshooting purposes during deployment of KVM guests.

Two types of validation are available:

- When you use the `--interface` option, `vm_maker` validates the specified interface by temporarily plumbing the specified network with the provided details and checking that it can reach the specified gateway.
- When you use the `--discover` option, `vm_maker` temporarily plumbs the specified network with the provided details and discovers the interfaces that are able to reach the specified gateway.

### Example 7-3 Interface Validation

The following shows an example of interface validation using the `--interface` option.

```
/opt/exadata_ovm/vm_maker --validate-ip 192.0.2.34 --netmask 255.255.240.0
--interface eth1 --gateway 192.0.2.1
[INFO] Interface eth1 is part of bondeth0. Validating bondeth0.
[INFO] Validating network settings, this can take up to a minute.
[INFO] Interface bondeth0 with ip address 192.0.2.34 and netmask
255.255.240.0 can ping gateway 192.0.2.1
```

### Example 7-4 Discovery Validation

The following shows an example of discovery validation using the `--discover` option.

```
/opt/exadata_ovm/vm_maker --validate-ip 192.0.2.34 --netmask
255.255.240.0 --discover --gateway 192.0.2.1
[INFO] Potential interfaces for discovery: eth0 eth1 eth2 eth3 eth4
[INFO] Discovering interface eth0
[INFO] Discovering interface eth1
[INFO] Discovering interface eth2
[INFO] Discovering interface eth3
[INFO] Discovering interface eth4
[INFO] Interfaces found: eth1 eth2
```

## 7.26 vhostmd Command

Use the `--vhostmd` command to enable, disable, or show the status of the vhostmd service.

### Syntax

```
vm_maker --vhostmd { enable | disable | status }
```

# 8

## Updating Exadata Software

There are different types of software used on an Exadata Database Machine that need be updated regularly.

- [About Updating Exadata Software](#)
- [Planning for Software Maintenance](#)  
Before starting a software update, you should review best practices, determine the version to which you will upgrade, and obtain the proper patching software.
- [About Upgrading to Oracle Linux 7 on Exadata Servers](#)  
When you upgrade to Oracle Exadata System Software release 19.1.0, the Oracle Linux version is upgraded from Oracle Linux 6 to Oracle Linux 7.
- [Overview of Performing Exadata Software Updates](#)  
For each software update of the Exadata components, there are various actions required to complete the software update.
- [Exadata Patchmgr Update Utility](#)  
Patchmgr is the utility used to update software for Exadata infrastructure components.
- [Updating Exadata Database Machine Database Servers](#)  
Use the following information and procedures when updating database servers within Exadata Database Machine.
- [Updating Database Servers Running Oracle Exadata System Software Release 11.2.2.4.2](#)  
Updates for database servers running Oracle Exadata System Software release 11.2.2.4.2 include an update that prepares the servers to use yum.
- [Updating Software on Oracle Exadata Storage Servers](#)  
Use the following information and procedures when updating storage servers within Exadata Database Machine.
- [Upgrading and Downgrading RoCE Network Fabric Switch Firmware](#)  
This topic describes procedures to upgrade and downgrade the firmware on the RoCE Network Fabric switches.
- [Updating InfiniBand Network Fabric Switch Firmware](#)  
This topic describes procedures to upgrade and downgrade the firmware on the InfiniBand Network Fabric switches.
- [Upgrading Oracle Java SE on Oracle Linux](#)  
You can upgrade Oracle Java SE (JDK) running Oracle Linux 6 on database servers and storage servers.
- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.
- [Troubleshooting Software Updates on Exadata Database Machine](#)  
Review these topics if you encounter errors or problems when updating the software on Exadata Database Machine

## 8.1 About Updating Exadata Software

Exadata software updates apply to three major components:

- Exadata storage servers
- Exadata database servers
- Exadata RDMA Network Fabric switches

Exadata storage server and Exadata database server updates generally contain updates for:

- Oracle Linux operating system
- Oracle Exadata System Software
- Firmware (for example: disk, flash, RAID controller, ILOM, HCA)

The updates do not modify the Oracle Grid Infrastructure home, Oracle Database home (other than relinking during the `dbnodeupdate.sh -c` step), or customer-installed software.

While it is generally recommended for components to stay in step with the recommended minimum release, you can choose to update different components at different times. For example, you could update RDMA Network Fabric switches at a later time than Exadata storage servers and Exadata database servers. However, you must check My Oracle Support Doc ID 888828.1 for any dependencies.

It is not mandatory to apply each and every Oracle Exadata System Software update that comes out. For example, you can skip two or three releases and update directly to a newer release. Oracle recommends that you update database servers twice a year.

Upgrading is allowed under the following circumstances:

1. The product version of the target release is higher than the installed software, and
2. The date code of the target release is higher than the installed software.

For example, consider a system that is currently running image version 12.1.1.1.2.150411, which consists of product version 12.1.1.1.2 and date code 150411.

- Upgrading to 12.1.2.1.2.150617.1 is allowed because both rules are satisfied.
- Upgrading to 12.1.2.1.1.150316.2 is not allowed because date code 150411 of the installed software is higher than date code 150316 of the target release.

### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Understanding the Exadata Software Image Version](#)  
The Exadata image version number contains the product version, a date code, and a build number.

## 8.2 Planning for Software Maintenance

Before starting a software update, you should review best practices, determine the version to which you will upgrade, and obtain the proper patching software.

- [Understanding Exadata Database Machine Software and Updates](#)  
Understanding the different types of software updates required for Exadata Database Machine helps you plan an update schedule.
- [Configuration and Operational Best Practices for Software Maintenance](#)  
As part of planning your software update, you should review the different methods of performing updates and the best practices for updating the software.
- [Understanding the Exadata Software Image Version](#)  
The Exadata image version number contains the product version, a date code, and a build number.
- [Rules for Updating to Newer Exadata Version](#)  
When upgrading, you need to consider both the product version and date code of the installed and target releases.

### 8.2.1 Understanding Exadata Database Machine Software and Updates

Understanding the different types of software updates required for Exadata Database Machine helps you plan an update schedule.

- [What Software Do You Update on Exadata Database Machine?](#)  
The software that runs on Exadata Database Machine is divided into two main categories.
- [Software Release Types](#)
- [Software Release Availability](#)
- [Software Update Frequency](#)  
You should plan to update your Exadata software on a regular basis.
- [Software Update Utilities](#)  
When updating software on Exadata Database Machine, you use specific utilities, depending on the component being updated.

#### 8.2.1.1 What Software Do You Update on Exadata Database Machine?

The software that runs on Exadata Database Machine is divided into two main categories.

- Oracle Exadata System Software
- Oracle Grid Infrastructure and Oracle Database software

The following table describes the two primary categories of software that run on an Exadata Database Machine.



Software	Components Installed On	Software Content	Example Software Version
Oracle Exadata System Software	<ul style="list-style-type: none"> <li>Exadata storage servers</li> <li>Exadata database servers</li> <li>Exadata RDMA Network Fabric switches</li> </ul>	<ul style="list-style-type: none"> <li>Oracle Linux operating system</li> <li>Oracle VM Server (on database servers for Oracle VM configurations)</li> <li>Oracle Exadata System Software</li> <li>Firmware (for example: disk, flash, RAID controller, ILOM, HCA, RDMA Network Fabric switch firmware)</li> </ul>	Oracle Exadata System Software on storage servers and database servers: <ul style="list-style-type: none"> <li>12.2.1.1.0</li> <li>12.1.2.3.4</li> </ul> InfiniBand Network Fabric switch firmware: <ul style="list-style-type: none"> <li>2.2.4-3</li> <li>2.1.8-1</li> </ul> RoCE Network Fabric switch firmware: <ul style="list-style-type: none"> <li>7.0(3)I7(6)</li> </ul>
Oracle Grid Infrastructure and Oracle Database software	Exadata database servers	<ul style="list-style-type: none"> <li>Oracle Grid Infrastructure (Grid home)</li> <li>Oracle Database (Database home)</li> </ul>	12.2.0.1.0 12.1.0.2.160117 11.2.0.4.161018

When an Exadata Database Machine is deployed, all of the software described in the table is installed and configured to deliver high performance and availability for Oracle Database. Extensive end-to-end testing ensures all software components supplied with Exadata Database Machine work seamlessly together.

My Oracle Support Doc ID 888828.1 is the primary source of information for software that runs on Exadata Database Machine. It contains the following information:

- List of all current and previous releases
- Minimum requirements for feature usage
- Compatibility requirements between Oracle Exadata System Software version and Oracle Database software version
- Compatibility requirements for specific hardware releases
- Guidelines for related products when used with Exadata Database Machine
- References to other pertinent information sources for Exadata Database Machine software maintenance

#### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

## 8.2.1.2 Software Release Types

The following table describes the release types for Oracle Exadata System Software:

Release type	Description and Content
Feature release	<p>Historically, the first four digits represent an Exadata feature release. For example, 12.2.1.1.</p> <p>Starting with the Oracle Exadata System Software release 18.1, the first two digits represent an Exadata feature release.</p> <p>A feature release contains new features, bug fixes, and security fixes. It may be used to update any prior feature or sustaining release.</p> <p>You can update directly to feature release 12.2.1.1.0 from any prior release from 11.2.3.3.1 to 12.1.2.3.4, inclusive.</p>
Sustaining release	<p>Historically, the fifth digit represents a sustaining release. For example, 12.1.2.3.4.</p> <p>Starting with the Oracle Exadata System Software release 18.1, the third digit represents a sustaining release. For example, 19.3.6.</p> <p>A sustaining release contains bug fixes and security fixes. A sustaining release may be used to update any prior feature or sustaining release.</p> <p>You can update directly to sustaining release 12.1.2.3.4 from any prior release from 11.2.3.3.1 to 12.1.2.3.3, inclusive.</p>
Interim patch	<p>An interim patch is one-off bug fix made available to customers who cannot wait until the fix is included in a sustaining release or feature release. Installation of an interim patch is done on an as-needed basis and is not a regularly scheduled planned maintenance event.</p>

#### Related Topics

- [About Oracle Database Release Numbers](#)

### 8.2.1.3 Software Release Availability

The following table outlines the release cycles for Oracle Exadata System Software, Oracle Grid Infrastructure, and Oracle Database software.

Software	Release type	Release Frequency	Importance of Maintaining Up-to-Date Versions	Recommendation for Adopting New Releases
Oracle Exadata System Software	Feature Release	6-18 months	Medium	Update to adopt new features and get critical fixes and security updates.

Software	Release type	Release Frequency	Importance of Maintaining Up-to-Date Versions	Recommendation for Adopting New Releases
Oracle Exadata System Software	Sustaining Release	Monthly	High	Update regularly to get critical fixes and security updates.  Oracle recommends that you remain within 12 months of the latest sustaining release.
Oracle Grid Infrastructure and Oracle Database software	Long Term Release	24+ months	Medium	Use long term releases for the highest level of stability and the longest length of error correction support.
Oracle Grid Infrastructure and Oracle Database software	Innovation Release	12-24 months	Medium	Use innovation releases to adopt new features as quickly as possible.
Oracle Grid Infrastructure and Oracle Database software	Release Update (RU)	3 months (quarterly)	High	Update quarterly to get critical fixes and security updates.  Oracle recommends that you remain within 12 months of the latest release update.

### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Release Schedule of Current Database Releases \(My Oracle Support Doc ID 742060.1\)](#)

## 8.2.1.4 Software Update Frequency

You should plan to update your Exadata software on a regular basis.

These examples show three high-level software maintenance plans with quarterly updates over a four year cycle using typical intervals between releases. The examples illustrate different update strategies designed to meet specific goals.



### Note:

Actual release frequencies may vary from these examples.

- [Example 8-1](#) Production System Software Maintenance Plan
- [Example 8-2](#) Production System Software Maintenance Plan with Reduced Updates

- [Example 8-3](#) Development and Test System Software Maintenance Plan

**Example 8-1 Production System Software Maintenance Plan**

**Goal** — Minimize risk by regularly applying critical and security fixes, adopting new feature releases only as required, and not adopting a new Exadata feature release and new Oracle Database release simultaneously.

Action	Y1 Q 1	Y1 Q 2	Y1 Q 3	Y1 Q 4	Y2 Q 1	Y2 Q 2	Y2 Q 3	Y2 Q 4	Y3 Q 1	Y3 Q 2	Y3 Q 3	Y3 Q 4	Y4 Q 1	Y4 Q 2	Y4 Q 3	Y4 Q 4	Y5 Q 1
Update to latest Oracle Exadata System Software sustaining release	-	X	X	X	X	-	X	X	X	X	-	X	X	X	X	-	X
Update to latest Oracle Exadata System Software feature release	X	-	-	-	-	X	-	-	-	-	X	-	-	-	-	X	-
Update to latest Oracle Database Release Update (RU)	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Update to latest Oracle Database innovation release	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Update to latest Oracle Database long term release	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

**Example 8-2 Production System Software Maintenance Plan with Reduced Updates**

**Goal** — Manage risk and reduce maintenance time by applying less frequent updates. Adopt new Oracle Database feature releases as required. Minimize risk by not adopting a new Exadata feature release and new Oracle Database feature release simultaneously.

Action	Y1 Q 1	Y1 Q 2	Y1 Q 3	Y1 Q 4	Y2 Q 1	Y2 Q 2	Y2 Q 3	Y2 Q 4	Y3 Q 1	Y3 Q 2	Y3 Q 3	Y3 Q 4	Y4 Q 1	Y4 Q 2	Y4 Q 3	Y4 Q 4	Y5 Q 1
Update to latest Oracle Exadata System Software sustaining release	-	-	-	X	-	X	-	-	-	X	-	X	-	-	-	X	-
Update to latest Oracle Exadata System Software feature release	-	X	-	-	-	-	-	X	-	-	-	-	-	X	-	-	-
Update to latest Oracle Database Release Update (RU)	-	X	-	-	-	X	-	X	-	X	-	X	-	X	-	-	-
Update to latest Oracle Database innovation release	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-
Update to latest Oracle Database long term release	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-

**Example 8-3 Development and Test System Software Maintenance Plan**

**Goal** — Adopt the latest features and software updates on a quarterly basis.

Action	Y1 Q 1	Y1 Q 2	Y1 Q 3	Y1 Q 4	Y2 Q 1	Y2 Q 2	Y2 Q 3	Y2 Q 4	Y3 Q 1	Y3 Q 2	Y3 Q 3	Y3 Q 4	Y4 Q 1	Y4 Q 2	Y4 Q 3	Y4 Q 4	Y5 Q 1
Update to latest Oracle Exadata System Software sustaining release	-	X	X	X	X	-	X	X	X	X	-	X	X	X	X	-	X
Update to latest Oracle Exadata System Software feature release	X	-	-	-	-	X	-	-	-	-	X	-	-	-	-	X	-
Update to latest Oracle Database Release Update (RU)	-	X	X	X	X	-	X	X	X	X	-	X	X	X	X	-	X
Update to latest Oracle Database innovation release	-	-	-	-	-	X	-	-	-	-	X	-	-	-	-	-	-
Update to latest Oracle Database long term release	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-

### 8.2.1.5 Software Update Utilities

When updating software on Exadata Database Machine, you use specific utilities, depending on the component being updated.

Update Type	Utility	Components	Description
Pre-update readiness check	Oracle EXAchk	Exadata Database Machine	Health check tool used to qualify maintenance readiness, identify critical issue exposure, and provide version recommendations for updating software.  Obtain Oracle EXAchk from My Oracle Support Document 1070954.1.

Update Type	Utility	Components	Description
Oracle Exadata System Software	patchmgr	<ul style="list-style-type: none"> <li>Exadata Database Machine storage servers</li> <li>Exadata Database Machine database servers</li> <li>RDMA Network Fabric switches</li> </ul>	Exadata Database Machine software update tool that updates all software (Oracle Linux, Oracle Exadata System Software, firmware) on Exadata Database Machine database servers, storage servers, or RDMA Network Fabric switches. For example, patchmgr updates Exadata Database Machine storage servers and database servers to Oracle Exadata System Software version 12.2.1.1.0, or InfiniBand Network Fabric switches to version 2.2.4-3.
Oracle Grid Infrastructure and Oracle Database software release update	OPatch or opatchauto	Exadata Database Machine database servers	Software update tool that applies a release update to an Oracle Grid Infrastructure home or Oracle Database home. For example, OPatch applies the 12.1.0.2.170117 release update to 12.1.0.2 Oracle Grid Infrastructure and Oracle Database homes. Obtain OPatch from My Oracle Support Document 274526.1.

Update Type	Utility	Components	Description
Oracle Grid Infrastructure and Oracle Database software patch set, major release, or release update.	Oracle Universal Installer	Exadata Database Machine database servers	<p>Software installation tool that installs an Oracle Grid Infrastructure and Oracle Database patch set, major, or release update, and upgrades Oracle Grid Infrastructure.</p> <p>For example, Oracle Universal Installer installs Oracle Database 12.2.0.1, or installs and upgrades Oracle Grid Infrastructure from 12.1.0.2 to 12.2.0.1.</p> <p>Use the version of Oracle Universal Installer bundled with the Oracle Grid Infrastructure and Oracle Database release to which you are upgrading.</p>
Upgrade of an Oracle Database	Database Upgrade Assistant (DBUA)	Exadata Database Machine database servers	<p>Database tool that upgrades a database to a later patch set, major release, or release update. For example, DBUA upgrades a database from 12.1.0.1 to 12.1.0.2, or from 12.1.0.2 to 12.2.0.1.</p> <p>Use DBUA bundled with the Oracle Grid Infrastructure and Oracle Database release to which you are upgrading.</p>

Update Type	Utility	Components	Description
Oracle Exadata System Software Oracle Grid Infrastructure and Oracle Database software	Fleet Patching and Provisioning (previously named Rapid Home Provisioning (RHP))	<ul style="list-style-type: none"> <li>• Exadata storage servers</li> <li>• Exadata database servers</li> <li>• RDMA Network Fabric switches</li> </ul>	<p>Software maintenance tool that provisions, patches, and upgrades Oracle Grid Infrastructure and Oracle Database software on one or more clusters from a centralized server.</p> <p>Use Fleet Patching and Provisioning as a feature of Oracle Clusterware.</p> <p>Starting with Oracle Grid Infrastructure release 18c, in addition to Oracle Grid Infrastructure and Oracle Database homes, Fleet Patching and Provisioning supports patching the Oracle Exadata components: database nodes, storage cells, and RDMA Network Fabric switches.</p>
Oracle Exadata System Software Oracle Grid Infrastructure and Oracle Database software	Oracle Enterprise Manager Cloud Control	<ul style="list-style-type: none"> <li>• Exadata storage servers</li> <li>• Exadata database servers</li> <li>• RDMA Network Fabric switches</li> </ul>	<p>Oracle Enterprise Manager Cloud Control provides lifecycle management capability to apply software updates to Exadata Database Machine.</p> <p><b>Note:</b> Facilities for updating Oracle Exadata System Software are deprecated in Oracle Enterprise Manager Cloud Control release 13.4.</p>

### Related Topics

- [Oracle Exadata Database Machine exachk or HealthCheck \(My Oracle Support document 1070954.1\)](#)
- [How To Download And Install The Latest OPatch \(6880880\) Version \(My Oracle Support document 274526.1\)](#)
- “Fleet Patching and Provisioning and Maintenance” in *Oracle Clusterware Administration and Deployment Guide*



- “Patching Oracle Exadata Software” in *Oracle Clusterware Administration and Deployment Guide*
- “Performing Engineered System Software Updates” in *Oracle Enterprise Manager Lifecycle Management Administrator’s Guide 13c Release 3*

## 8.2.2 Configuration and Operational Best Practices for Software Maintenance

As part of planning your software update, you should review the different methods of performing updates and the best practices for updating the software.

- [Understanding Rolling and Non-Rolling Updates](#)  
Software updates can be performed in rolling manner while the database remains online and available, or non-rolling manner where Oracle Clusterware and Oracle Database are shutdown.
- [Online Updates for Oracle Linux Kernel and Oracle Database Interim Fixes](#)  
Online updates of qualified fixes are supported for Oracle Linux and Oracle Database.
- [Configuration Practices for Optimal Software Maintenance](#)  
When configuring an Exadata Database Machine, it is important to adopt features that will lessen the impact and risk of performing software updates.
- [Operational Practices for Optimal Software Maintenance](#)
- [Version Compatibility and Mixed Version Support](#)  
It is recommended that software for all components be updated regularly to stay in step with the recommended minimum release in order to maintain the most stable and secure system.
- [Ordering of Updates](#)  
In general, when performing software updates, the updates may be applied to components in any order based on business needs and maintenance window requirements.

### 8.2.2.1 Understanding Rolling and Non-Rolling Updates

Software updates can be performed in rolling manner while the database remains online and available, or non-rolling manner where Oracle Clusterware and Oracle Database are shutdown.

The manner an update is performed does not affect how often it should be done, but does determine how long it will take. In general there are two methods:

- **Rolling software updates** — A rolling software update is one that is performed to one of a particular component at a time while the others remain online servicing requests. Note the following key points about rolling software updates:
  - Rolling updates have less application downtime compared to non-rolling updates.
  - The overall length of time to complete the update is longer because one component is offline and updated at a time, while all other components remain online and operational.
  - The impact to existing database connections differs depending on the component being updated.

- \* Rolling updates for Exadata Database Machine storage servers or RDMA Network Fabric switches — All databases remain fully online and available for the duration of the rolling update. There is no disruption to database connections.
  - \* Rolling updates for Exadata Database Machine database servers, Oracle Grid Infrastructure, or Oracle Database — Multi-instance databases using Oracle Real Application Clusters (Oracle RAC) remain available for the duration of the update. However, database connections on the server being updated will be disrupted when the local database instance is shutdown. Use Oracle RAC features for client high availability to minimize application disruption, as described in [Table 8-1](#).
- **Non-Rolling software updates** — A non-rolling software update is one that is performed to all of a particular component while those components are offline. Note the following key points about non-rolling software updates:
    - Non-rolling updates are faster than rolling updates for overall maintenance time because multiple components are updated in parallel.
    - Because all of a particular component are offline for the duration of the update, a database (or all databases on a system) will also be offline for the duration of the update, resulting in a complete outage to the application
    - Important applications serviced by the Exadata Database Machine being updated in a non-rolling manner are often moved to a standby system in environments using Oracle Data Guard.
  - **Combination of rolling and non-rolling software updates** — When multiple components are updated in the same maintenance window, it is possible to use a combination of rolling and non-rolling methods to achieve the desired balance of application downtime and maintenance time. One typical combination used in the situation where an application does not handle connection disruption efficiently is to perform Exadata Database Machine storage server and RDMA Network Fabric switch updates in a rolling manner, then performing Oracle Grid Infrastructure, Oracle Database, and Exadata Database Machine database server updates in a non-rolling manner.

The `patchmgr` update utility manages the update orchestration in a rolling or non-rolling manner for Exadata Database Machine infrastructure components (storage servers, database servers, and RDMA Network Fabric switches).

The following table describes which update method is supported for each component type.

**Table 8-1 Rolling versus Non-Rolling Upgrades**

Method	Component Support	Database Availability Impact During Update
Rolling	Exadata Database Machine storage servers	No impact. Oracle RAC and single instance databases remain fully available on all nodes in the cluster.
Rolling	RDMA Network Fabric switches	No impact. Oracle RAC and single instance databases remain fully available on all nodes in the cluster.

**Table 8-1 (Cont.) Rolling versus Non-Rolling Upgrades**

Method	Component Support	Database Availability Impact During Update
Rolling	Exadata Database Machine database servers	Local connections are disconnected and all local instances are shutdown.  Oracle RAC databases remain available through other nodes in the cluster.
Rolling	Oracle Grid Infrastructure	Local connections are disconnected and all local instances are shutdown.  Oracle RAC databases remain available through other nodes in the cluster.
Rolling	Oracle Database - Release Update (RU)	For the database home being updated, local connections are disconnected and local instances are shutdown.  Oracle RAC databases remain available through other nodes in the cluster.  Databases running from other Oracle Database software homes are unaffected.
Non-Rolling	<ul style="list-style-type: none"> <li>• Exadata Database Machine storage servers</li> <li>• Exadata Database Machine database servers</li> <li>• Oracle Grid Infrastructure</li> <li>• Oracle Database - Release Update (RU)</li> <li>• Oracle Database - Patch Set</li> <li>• Oracle Database - Release</li> </ul>	Databases unavailable Move workload to Oracle Data Guard or Oracle GoldenGate standby systems to minimize impact.

### 8.2.2.2 Online Updates for Oracle Linux Kernel and Oracle Database Interim Fixes

Online updates of qualified fixes are supported for Oracle Linux and Oracle Database.

Typically, planned software updates require that the component be restarted after the update. For example, an Exadata database server must be restarted after applying an Exadata software release to update the system firmware and to make active a new Oracle Linux kernel. Similarly, an Oracle database instance must be stopped and restarted to apply a software update to the database home.

Oracle supplies some interim fixes, however, that can be applied online, such that a component does not require a restart to apply the fix and make them active. Online updates of qualified fixes are supported for the following components:

- Oracle Linux kernel on Exadata database servers using the Ksplice Offline Client.
- Oracle Database using OPatch Online Patching.

Online updates are typically performed as a temporary measure when a critical fix must be applied to the system before the next scheduled planned software maintenance.

#### Related Topics

- [HOWTO: Install ksplice kernel updates for Exadata Database Nodes \(My Oracle Support Doc ID 2207063.1\)](#)
- [RDBMS Online Patching Aka Hot Patching \(My Oracle Support Doc ID 761111.1\)](#)

### 8.2.2.3 Configuration Practices for Optimal Software Maintenance

When configuring an Exadata Database Machine, it is important to adopt features that will lessen the impact and risk of performing software updates.

Configuration Practice	Use for Software Maintenance	Purpose
Oracle ASM high redundancy disk groups	Rolling updates - Exadata storage servers	<p>When performing storage server updates in a rolling manner with the databases remaining online, it is highly recommended to configure Oracle ASM disk groups with high redundancy. High redundancy disk groups can tolerate the failure of a disk in another storage server during rolling updates.</p> <p>During rolling, or online, storage server updates, the disks for the storage server being updated are taken offline on one storage server by patchmgr while it is updated. After the update completes the disks are resynchronized by Oracle ASM, and then patchmgr starts to update the next storage server.</p> <p>While disks are offline the disk group has reduced redundancy. A normal redundancy disk group with reduced redundancy during a rolling update may dismount and have data loss if a disk fails in another storage server.</p> <p>Oracle ASM disk group redundancy is typically set during initial system configuration. Therefore, consider how you plan to perform storage server updates, rolling or non-rolling, prior to system configuration.</p> <p>An Oracle Data Guard physical standby system can also provide protection against disk failure during rolling storage server update.</p>

Configuration Practice	Use for Software Maintenance	Purpose
Oracle RAC features for client high availability	Rolling updates - Exadata database servers, Oracle Grid Infrastructure or Oracle Database software	When performing Exadata database server, Oracle Grid Infrastructure, or Oracle Database software updates in a rolling manner on one database server at a time while database services remain running on the other servers in the cluster, the database services must be stopped on the Exadata database server being updated. To minimize the impact to client applications connected to database instances that will be stopped during maintenance, configure client applications to use database services, Fast Application Notification (FAN), Fast Connection Failover (FCF), and Application Continuity.
Oracle Data Guard physical standby database	Any rolling or non-rolling update (except Oracle Database patch set and release updates)	Oracle Data Guard Standby-First Patch Apply provides support for different Exadata infrastructure, Oracle Grid Infrastructure, and Oracle Database software between a primary database and its physical standby database for the purpose of applying and validating Oracle patches in rolling fashion with minimal risk to the primary database.
Oracle Data Guard transient logical standby database	Any rolling or non-rolling update	Use Oracle Data Guard logical standby database to reduce database upgrade downtime for patch set and release updates. Database upgrade downtime is reduced by allowing the logical standby database to be upgraded to the new version and kept synchronized while the primary database remains online running the current version.
Oracle GoldenGate	Any rolling or non-rolling update	Use Oracle GoldenGate to reduce database upgrade downtime for patch set and release updates. Downtime during a database upgrade is reduced by allowing the target database to be upgraded to the new version and kept synchronized while the source database remains online running the current version.

### Related Topics

- “Client High Availability with Services, FAN, FCF, and Application Continuity” in *Oracle Database 2 Day + Real Application Clusters Guide*
- [Oracle Patch Assurance - Data Guard Standby-First Patch Apply \(My Oracle Support Doc ID 1265700.1\)](#)
- “Patching, Upgrading, and Downgrading Databases in an Oracle Data Guard Configuration” in *Oracle Data Guard Concepts and Administration*
- “Configuring Oracle GoldenGate to Maintain a Live Standby Database” in *Administering Oracle GoldenGate for Windows and UNIX*

## 8.2.2.4 Operational Practices for Optimal Software Maintenance

The following operational practices enable optimal software maintenance:

- **Run Oracle EXAchk regularly** — As a general health check tool to ensure an Exadata system continues to meet the current and constantly evolving best practices, run Oracle EXAchk monthly. The exachk report should be utilized as follows:
  - Baseline comparison - Compare the current report against an accepted baseline report using the report comparison feature (`-diff` option).
  - Critical issue exposure - Review the report for exposure to critical issues. Take prompt action to resolve critical issues reported by Oracle EXAchk.
  - Version recommendation - Review the report for version recommendation. The MAA Scorecard section evaluates current software versions for consistency, compatibility, and whether or not it is current.
- **Qualify maintenance readiness with Oracle EXAchk** — Prior to performing software maintenance, run Oracle EXAchk to ensure the system is in a healthy state. Correct any `FAIL` or `WARNING` checks before updating any software. After software maintenance is complete, run Oracle EXAchk again to confirm system health.
- **Update software at regular intervals** — All software should be updated regularly. Maintaining software at current or recent releases provides the following benefits: better software security, more stable sustaining releases, continued compatibility with newer related software, better support and faster resolution of issues, and ability to supply fixes for newly discovered issues.
- **Use the latest versions of the update utilities** — Use the latest version of software update utilities Oracle EXAchk, patchmgr, and OPatch.

Oracle EXAchk is updated regularly to contain new features, fixes, best practice health checks, and version recommendations. Patchmgr for database servers is updated regularly to contain new features, fixes, and workarounds for known database server update issues. OPatch is updated regularly to contain new features and fixes.
- **Perform storage server updates separately from RDMA Network Fabric switch updates** — Do not update storage servers and RDMA Network Fabric switches concurrently. RDMA Network Fabric network connections must be stable during some critical stages of storage server updates. The RDMA Network Fabric switch firmware upgrade requires a switch reboot, which disrupts some connections on the RDMA Network Fabric.
- **Avoid unsupported system changes** — Exadata Database Machine is an integrated system and engineered to be the best platform for running Oracle Database. Exadata Database Machine storage servers and RDMA Network Fabric switches contain all

software necessary to run Oracle Database and are configured to run Oracle Database optimally. Software updates to Exadata Database Machine storage servers and RDMA Network Fabric switches are performed using `patchmgr`. Configuration or installed software may not be altered manually (without using `patchmgr`) in any way unless the Exadata Database Machine documentation contains steps to perform the desired change. While making an unsupported manual change may have the desired immediate effect, there are many potential negative consequences, such as:

- Failed future `patchmgr` software updates
- Inability to rescue the system
- Inability to diagnose a software defect efficiently

Contact Oracle Support Services for further guidance if a desired Exadata Database Machine storage server or RDMA Network Fabric switch change is not documented.

- **Minimize database server customization** — Exadata Database Machine is an integrated system and engineered to be the best platform for running Oracle Database. Exadata Database Machine database servers contain all software necessary to run Oracle Database and are configured to run Oracle Database optimally. Software updates to Exadata Database Machine database servers are performed using `patchmgr`. However, it may be necessary to manually install additional, site-specific software, such as monitoring agents or backup agents.

It is supported to manually customize database servers, but note that customizing the operating system by adding or updating packages may require additional actions when applying a future Oracle Exadata System Software update with `patchmgr` (for example, removing customization prior to updating Exadata Database Machine servers and reapplying the customization after the update completes) because the additional software may add new dependencies which are not provided by a future Oracle Exadata System Software update. It is recommended to minimize database server customization. See the topic [Updating Exadata Database Machine Database Servers](#) for additional details.

- **Test supported configuration changes** — Database server site-specific customization must be tested completely, including verifying database servers reboot properly after making configuration changes. Custom configuration changes that prevent successful system reboot will cause future Oracle Exadata System Software updates to fail.

### 8.2.2.5 Version Compatibility and Mixed Version Support

It is recommended that software for all components be updated regularly to stay in step with the recommended minimum release in order to maintain the most stable and secure system.

However, it is supported to update only a subset of components during a software maintenance window while the remaining components remain at an earlier version. For example, the following scenarios are supported:

- Update a subset of Exadata Database Machine storage servers to a higher version of the Oracle Exadata System Software while the remaining storage servers remain at the earlier Oracle Exadata System Software version.



- Update Exadata Database Machine storage servers to a higher version of the Oracle Exadata System Software while the database servers remain at the earlier Oracle Exadata System Software version.
- Update Exadata Database Machine database servers to a higher version of the Oracle Exadata System Software while the storage servers remain at the earlier Oracle Exadata System Software version.
- Update Exadata Database Machine storage servers to a higher version of the Oracle Exadata System Software while RDMA Network Fabric switches remain at the earlier Oracle Exadata System Software version.
- Update RDMA Network Fabric switches to a higher version of the Oracle Exadata System Software while Exadata Database Machine storage servers remain at the earlier Oracle Exadata System Software version.
- Update Exadata Database Machine storage servers to a higher version of the Oracle Exadata System Software while Oracle Grid Infrastructure or Oracle Database software remains at an earlier release and update level.

While mixed versions are supported (within the same component or across different components), it is highly recommended that this be only a temporary configuration that exists for the purpose and duration of rolling upgrade.

If mixing versions the following rules and considerations must be observed:

- A specific generation of Exadata Database Machine hardware will have a minimum required Oracle Exadata System Software version. For example, Exadata Database Machine X6 hardware requires Oracle Exadata System Software release 12.1.2.3.1 or higher.
- A specific Oracle Database release requires a minimum Oracle Exadata System Software release to fully support Exadata Database Machine features. For example, Oracle Database 12c Release 2 (12.2) requires Oracle Exadata System Software release 12.2.1.1.0 or higher on storage servers to support all Exadata Database Machine offload features.
- Features supplied with a new Oracle Exadata System Software release may require a minimum Oracle Grid Infrastructure or Oracle Database software release or release update. For example, the Exadata Database Machine feature Smart Scan Offload for Compressed Index Scan requires Oracle Database 12.2 and Oracle Exadata System Software 12.2 on storage servers.
- Oracle Grid Infrastructure supports mixed versions only during a rolling update (for example, while updating from release 12.1.0.2 to 12.2.0.1, or updating from release 12.1.0.2.161018 to 12.1.0.2.170117). Some Oracle Clusterware functionality is restricted while the cluster is in rolling upgrade mode.
- Oracle Database supports mixed versions only during a rolling update (for example, while updating from release 12.1.0.2.161018 to 12.1.0.2.170117).

### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support document 888828.1\)](#)
- [What's New in Oracle Exadata Database Machine 12c Release 2 \(12.2.1.1.0\)](#)



## 8.2.2.6 Ordering of Updates

In general, when performing software updates, the updates may be applied to components in any order based on business needs and maintenance window requirements.

The recommended order for applying software updates is:

1. Oracle Grid Infrastructure and Oracle Database software
2. Exadata Database Machine database server software
3. Exadata Database Machine storage server software
4. RDMA Network Fabric switch software

There are exceptions where specific update ordering is required, which are documented in the Oracle Exadata System Software supplemental README for the version being updated to. Some examples of required update ordering are:

- Oracle Grid Infrastructure and Oracle Database 18c require Oracle Exadata System Software version 18.1.4 or higher. When updating a system with Oracle Exadata System Software release 12.2 and Oracle Grid Infrastructure 12c to Oracle Exadata System Software release 18 and Oracle Grid Infrastructure 18c, respectively, all Exadata Database Machine components must be updated first to Oracle Exadata System Software release 18.1.4 or higher before updating Oracle Grid Infrastructure and Oracle Database to release 18c.
- When updating a system with Oracle Advanced Cluster File System (Oracle ACFS) configured, the Oracle Grid Infrastructure home must contain the fix for bug 22810422 before updating the database server (non-Oracle VM or Oracle VM) to Oracle Exadata System Software release 12.2.
- When updating an Oracle VM configuration from Oracle Exadata System Software release 12.1 to release 12.2, all user domains must be updated to Oracle Exadata System Software release 12.2 before updating the management domain to Oracle Exadata System Software release 12.2.
- When updating an Oracle VM configuration, the Oracle Grid Infrastructure home in the user domain must be running release 12.1.0.2.161018 (October 2016) or later before updating a user domain to Oracle Exadata System Software release 12.2.

In general, software update ordering requirements occur when updating one or more components to a later version that contains new features, for example, when updating Oracle Exadata System Software from release 12.1 to release 12.2. Refer to the Oracle Exadata System Software supplemental README for the software version being updated to for additional details.

### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

## 8.2.3 Understanding the Exadata Software Image Version

The Exadata image version number contains the product version, a date code, and a build number.

The Exadata version installed on an Exadata storage server or an Exadata database server is determined with the `imageinfo` command. The image version of an Exadata release consists of three components. For example, if the command `imageinfo -ver` returns `12.1.2.1.1.150316.2`, then:

- The product version is 12.1.2.1.1
- The date code is 150316
- The Oracle-internal build number is 2. This component is not always used.

In most circumstances it is sufficient to reference a release by the product version only, for example, 12.1.2.1.1. However, when upgrading, you need to consider both the product version and date code of the installed and target releases.

Starting with Oracle Exadata System Software release 18.1, the product version uses a three field format consisting of: *Year.Update.Revision*, for example 18.1.0.

## 8.2.4 Rules for Updating to Newer Exadata Version

When upgrading, you need to consider both the product version and date code of the installed and target releases.

In most circumstances it is sufficient to reference a release by the product version only, for example, 12.1.2.1.1.

Upgrading to a specific target version must adhere to the following rules, which are enforced by `patchmgr`:

1. The product version of the target release must be higher than the installed software and
2. The date code of the target release must be higher than the installed software.

For example, consider an Exadata Database Machine that is currently running image version 12.1.1.1.2.150411, which consists of product version 12.1.1.1.2 and date code 150411.

- Upgrading to 12.1.2.1.2.150617.1 is allowed because both rules are satisfied.
- Upgrading to 12.1.2.1.1.150316.2 is not allowed because date code 150411 of the installed software is higher than date code 150316 of the target release.

## 8.3 About Upgrading to Oracle Linux 7 on Exadata Servers

When you upgrade to Oracle Exadata System Software release 19.1.0, the Oracle Linux version is upgraded from Oracle Linux 6 to Oracle Linux 7.

Oracle Exadata System Software release 19.1.0 provides support for Oracle Database 19c, which runs on Oracle Linux 7. When you upgrade to Oracle Exadata System Software release 19.1.0 the operating system is upgraded at the same time.

If you use quorum disks on your Exadata Database Machine, then refer to My Oracle Support note 2453054.1.

### Related Topics

- [Exadata Oracle Linux 6 to Oracle Linux 7 Quorumdisk Migration FAQ \(Doc ID 2453054.1\)](#)

## 8.4 Overview of Performing Exadata Software Updates

For each software update of the Exadata components, there are various actions required to complete the software update.

- [Actions to Perform Before Any Software Maintenance](#)  
Perform you update any software on your Exadata Database Machine, you should perform these actions.
- [Overview of Performing Exadata Storage Server Updates](#)  
A high level look at the process for updating Exadata storage servers.
- [Overview of Performing Exadata Database Servers Updates](#)  
When updating Exadata database servers, there are different steps for virtualized environments.
- [Overview of Performing RoCE Network Fabric Switch Updates](#)  
Patchmgr manages the update orchestration in a rolling manner for RoCE Network Fabric switches
- [Overview of Performing InfiniBand Network Fabric Switch Updates](#)  
Patchmgr manages the update orchestration in a rolling manner for InfiniBand Network Fabric switches
- [Overview of Performing Oracle Grid Infrastructure and Oracle Database Updates](#)  
The steps for updating Oracle Grid Infrastructure or Oracle Database software depend on the type of software update you are performing.
- [Using sudo When Performing Software Updates](#)  
You can use `sudo` when running `dbnodeupdate.sh` and `patchmgr` (and `dbnodeupdate` Orchestration).

### 8.4.1 Actions to Perform Before Any Software Maintenance

Perform you update any software on your Exadata Database Machine, you should perform these actions.

- **Identify the target release**  
The target release can be determined by one of the following:
  - Recommendation in My Oracle Support Document 888828.1.
  - Version recommendations in the MAA Scorecard section of the Oracle ExaCHK report.
- **Run Oracle ExaCHK**  
Ensure software maintenance readiness of the system by running the latest version of Oracle ExaCHK (see My Oracle Support Document 1070954.1). Review the report for the following:
  1. Correct any FAIL or WARNING checks that deviate from your baseline.
  2. Review the MAA Scorecard section for version recommendations. Oracle ExaCHK evaluates current software versions for consistency, compatibility, and whether or not it is current.
  3. Review the MAA Scorecard section for exposure to critical issues. Target version choice should resolve any critical issue exposure.

It is recommended to run the latest version of Oracle ExaCHK in the following times:

- Every month to maintain a system that continues to adhere to Exadata best practices
  - The week before the software update to ensure maintenance readiness
  - The day before the software update to ensure maintenance readiness
  - Immediately after completing the software update
- **Run prerequisite checks**
    1. Download the target software release(s) and the latest versions of the software update utilities (patchmgr, OPatch, and so on) from My Oracle Support.
    2. Run the prerequisite checks and correct any issues. The initial prerequisite check run should be performed days or weeks in advance of the maintenance window to allow time to correct issues. Run prerequisite checks again immediately before the update.
  - **Review the latest documentation**
    1. Review My Oracle Support Document 1270094.1 for recently published critical issues that are not yet automatically checked by Oracle ExaCHK.
    2. Review the Supplemental README in My Oracle Support for the target software version for known issues discovered after the software was released.

#### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Oracle Exadata Database Machine exachk or HealthCheck \(My Oracle Support Doc ID 1070954.1\)](#)
- [Exadata Critical Issues \(My Oracle Support Doc ID 1270094.1\)](#)

## 8.4.2 Overview of Performing Exadata Storage Server Updates

A high level look at the process for updating Exadata storage servers.

Patchmgr manages the update orchestration in a rolling or non-rolling manner for storage servers. When using the `-rolling` option, Oracle recommends that you use Oracle ASM high redundancy disk groups to tolerate the failure of a disk in another storage server during the update.

1. Download the target Exadata software from My Oracle Support and stage it on the driving system. See My Oracle Support document 888828.1.
2. Create a file that contains the list of storage servers to update. This file will be specified as the *component\_list\_file*.
3. Configure SSH equivalence from the user that will run patchmgr to root on all storage servers in the *component\_list\_file*.
4. Run patchmgr prerequisite check and correct any issues.
5. If storage servers will be updated non-rolling, then stop Oracle Clusterware and all databases accessing the storage servers.
6. Run patchmgr to update storage servers in a non-rolling (default) or rolling (`-rolling` option) manner.
7. If storage servers were updated non-rolling, then restart Oracle Clusterware and all databases.

### Related Topics

- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

## 8.4.3 Overview of Performing Exadata Database Servers Updates

When updating Exadata database servers, there are different steps for virtualized environments.



### Note:

Patchmgr manages the update orchestration in a rolling or non-rolling manner for Exadata database servers. When using the `-rolling` option, Oracle recommends that you use Oracle RAC features for client high availability to minimize the impact to client applications connected to database instances that will be stopped during maintenance.

- [Steps for Updating Non-Virtualized Configurations](#)  
An overview of the steps involved in updating an Exadata database server that does not have Oracle VM configured.
- [Steps for Updating Virtualized Configurations](#)  
The management domains (dom0s) and user domains (domUs) are updated independently using separate patchmgr executions.

### 8.4.3.1 Steps for Updating Non-Virtualized Configurations

An overview of the steps involved in updating an Exadata database server that does not have Oracle VM configured.

1. Download the latest patchmgr from My Oracle Support Patch 21634633 and stage it on the driving system.
2. Download the target Exadata software from My Oracle Support and stage it on the driving system. See My Oracle Support Document 888828.1.
3. Create the *component\_list\_file* containing the list of database servers to update.
4. Configure SSH equivalence from the user that will run patchmgr to root on all database servers in the *component\_list\_file*.
5. Run patchmgr backup to backup database server root file system and operating system.
6. Run patchmgr prerequisite check and correct any issues.
7. Run patchmgr to update database servers in a non-rolling (default) or rolling (`-rolling` option) manner.

### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)

- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

### 8.4.3.2 Steps for Updating Virtualized Configurations

The management domains (dom0s) and user domains (domUs) are updated independently using separate patchmgr executions.

#### Updating the Management domain (dom0)

- Download the latest patchmgr from My Oracle Support Patch 21634633 and stage it on the driving system.
- Download the target dom0 Exadata software from My Oracle Support and stage it on the driving system. See My Oracle Support Patch Document 888828.1.
- Create the *component\_list\_file* containing the list of database server dom0s to update.
- Configure SSH equivalence from the user that will run patchmgr to root on all database server dom0s in the *component\_list\_file*.
- Run patchmgr backup to backup database server dom0 root file system and operating system.
- Run patchmgr prerequisite check and correct any issues.
- Run patchmgr to update database server dom0s in a non-rolling (default) or rolling (`-rolling` option) manner.

#### Updating the User domains (domU)

1. Download the latest patchmgr from My Oracle Support Patch 21634633 and stage it on the driving system.
2. Download the target domU Exadata software from My Oracle Support and stage it on the driving system. See My Oracle Support document 888828.1.
3. Create the *component\_list\_file* containing the list of database server domUs to update.
4. Configure SSH equivalence from the user that will run patchmgr to root on all database server domUs in the *component\_list\_file*.
5. Run patchmgr backup to backup database server domU root file system and operating system.
6. Run patchmgr prerequisite check and correct any issues.
7. Run patchmgr to update database server domUs in a non-rolling (default) or rolling (`-rolling` option) manner.

#### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

## 8.4.4 Overview of Performing RoCE Network Fabric Switch Updates

Patchmgr manages the update orchestration in a rolling manner for RoCE Network Fabric switches



### Note:

Perform storage server updates separately from RDMA Network Fabric switch updates. Do not update storage servers and RDMA Network Fabric switches concurrently. RDMA Network Fabric network connections must be stable during some critical stages of storage server updates. The RDMA Network Fabric switch firmware upgrade requires a switch reboot, which disrupts some connections on the RDMA Network Fabric.

The general steps for using patchmgr are:

1. Download the target Exadata software from My Oracle Support and stage it on the driving system. See My Oracle Support document 888828.1.
2. Create the *component\_list\_file* containing the list of RoCE Network Fabric switches to update.
3. Configure SSH equivalence from the user that will run patchmgr to `root` on all RoCE Network Fabric switches in the *component\_list\_file*.
4. Run patchmgr prerequisite check and correct any issues.
5. Run patchmgr to update RoCE Network Fabric switches. RoCE Network Fabric switches are always updated in a rolling manner.

### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

## 8.4.5 Overview of Performing InfiniBand Network Fabric Switch Updates

Patchmgr manages the update orchestration in a rolling manner for InfiniBand Network Fabric switches

 **Note:**

Perform storage server updates separately from RDMA Network Fabric switch updates. Do not update storage servers and RDMA Network Fabric switches concurrently. RDMA Network Fabric network connections must be stable during some critical stages of storage server updates. The RDMA Network Fabric switch firmware upgrade requires a switch reboot, which disrupts some connections on the RDMA Network Fabric.

The general steps for using `patchmgr` are:

1. Download the target Exadata software from My Oracle Support and stage it on the driving system. See My Oracle Support document 888828.1.
2. Create the `component_list_file` containing the list of InfiniBand Network Fabric switches to update.
3. Configure SSH equivalence from the user that will run `patchmgr` to `root` on all InfiniBand Network Fabric switches in the `component_list_file`.
4. Run `patchmgr` prerequisite check and correct any issues.
5. Run `patchmgr` to update InfiniBand Network Fabric switches. InfiniBand Network Fabric switches are always updated in a rolling manner.

**Related Topics**

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

## 8.4.6 Overview of Performing Oracle Grid Infrastructure and Oracle Database Updates

The steps for updating Oracle Grid Infrastructure or Oracle Database software depend on the type of software update you are performing.

**Updating to a Release Update (RU)**

Updates to a new Release Update (RU) may be performed in-place or out-of-place. Both in-place and out-of-place methods support rolling and non-rolling updates.

- **In-place**

The update is applied to the current software home using OPatch while the Oracle Grid Infrastructure or Oracle Database software is shutdown on the node being updated.

This is the default method of applying an RU for Oracle Grid Infrastructure or Oracle Database software, as described in the RU README. The steps to apply the RU must be performed on each node.

- **Out-of-place (recommended)**

A new software home is prepared and updated while the Oracle Grid Infrastructure and Oracle Database software remains running. Once the new homes are prepared, Oracle



Grid Infrastructure and Oracle Database are quickly stopped, switched to the new home, and restarted.

Out-of-place has significant advantage over in-place updates:

- There is less risk because the new home can be prepared while the Oracle Grid Infrastructure and Oracle Database software remains online.
- There is less downtime because switching to the new software home is faster than applying an update in-place.
- Rollback is faster because it is possible to simply switch back to the original software home.

The recommended way to adopt out-of-place updates is to use Fleet Patching and Provisioning (previously called Rapid Home Provisioning), which provides the following advantages:

- Distributes software updates to all nodes in the cluster.
- Orchestrates updates across the cluster in a rolling or non-rolling manner with a single command.
- Provides control over database service relocation to maintain application availability.

Out-of-place software updates are also supported by Oracle Enterprise Manager Cloud Control.

Out-of-place software updates without Fleet Patching and Provisioning or Oracle Enterprise Manager Cloud Control may be accomplished by following My Oracle Support Document 2087150.1.

### **OJVM Update**

The OJVM update is a separate software update for database homes that addresses OJVM security vulnerabilities. It is installed separately from a standard software update. The OJVM update may be installed in a rolling manner under certain situations.

### **Updating to a New Release**

To update to a higher Oracle Grid Infrastructure or Oracle Database release, follow the step-by-step release instructions in My Oracle Support.

### **Related Topics**

- [Minimal downtime patching via cloning 12c ORACLE\\_HOME directories with OPlan \(My Oracle Support Doc ID 2087150.1\)](#)
- [RAC Rolling Install Process for the "Oracle JavaVM Component Database PSU" \(OJVM PSU\) Patches \(My Oracle Support Doc ID 2217053.1\)](#)
- [12.2 Grid Infrastructure and Database Upgrade steps for Exadata Database Machine running 11.2.0.3 and later on Oracle Linux \(My Oracle Support Doc ID 2111010.1\)](#)
- [11.2.0.2, 11.2.0.3, 11.2.0.4 or 12.1.0.1 to 12.1.0.2 Grid Infrastructure and Database Upgrade on Exadata Database Machine running Oracle Linux \(My Oracle Support Doc ID 1681467.1\)](#)
- [Exadata Database Machine 11.2.0.4 Grid Infrastructure and Database Upgrade for 11.2.0.2 BP12 and later \(My Oracle Support Doc ID 1565291.1\)](#)

- [Exadata Database Machine 11.2.0.4 Grid Infrastructure and Database Upgrade for 11.2.0.2 BP11 and earlier \(My Oracle Support Doc ID 1555036.1\)](#)

## 8.4.7 Using sudo When Performing Software Updates

You can use `sudo` when running `dbnodeupdate.sh` and `patchmgr` (and `dbnodeupdate` Orchestration).

- [Running patchmgr \(and dbnodeupdate Orchestration\) Using sudo](#)  
You can run `patchmgr` (which is packaged in `dbserver.patch.zip`) using `sudo` to perform any of `patchmgr`'s functionalities, such as patching cells, patching RDMA Network Fabric switches, or orchestrating `dbnodeupdate.sh` execution.
- [Running dbnodeupdate.sh Using sudo](#)  
Before using `dbnodeupdate.sh` using `sudo`, you configure the `/etc/sudoers` file.

### 8.4.7.1 Running patchmgr (and dbnodeupdate Orchestration) Using sudo

You can run `patchmgr` (which is packaged in `dbserver.patch.zip`) using `sudo` to perform any of `patchmgr`'s functionalities, such as patching cells, patching RDMA Network Fabric switches, or orchestrating `dbnodeupdate.sh` execution.

`Patchmgr` is packaged in `dbserver.patch.zip`. Perform the following steps to set up the `/etc/sudoers` file for running `patchmgr` using `sudo`:

1. Log in as the `root` user and edit `/etc/sudoers` using `visudo`.

```
visudo
```

2. Add the following entry to the bottom of the `sudoers` file to allow non-root users, such as the `oracle` user, to run `patchmgr` as `root`.

#### Note:

The first field in the line specifies the non-root user who is granted `sudo` access for the `patchmgr` command. The line below uses the `oracle` user as an example. You can specify a different user if necessary.

```
oracle ALL=(ALL) NOPASSWD:SETENV: /u01/stage/patch/dbserverpatch/
patchmgr
```

3. As `root`, create the `/u01/stage/patch/dbserverpatch` directory and unzip `dbserver.patch.zip`:

```
mkdir -p /u01/stage/patch/dbserverpatch/
cp dbserver.patch.zip /u01/stage/patch/dbserverpatch/
cd /u01/stage/patch/dbserverpatch/
unzip dbserver.patch.zip
```

4. Move everything under the `/u01/stage/patch/dbserverpatch/dbserver_patch_x.yymmdd` directory to `/u01/stage/patch/dbserverpatch/`.

```
mv /u01/stage/patch/dbserverpatch/dbserver_patch_x.yymmdd/* /u01/
stage/patch/dbserverpatch/
```

#### Note:

- Patchmgr expects root SSH equivalence on all database nodes that will be updated, even when run using `sudo`.
- The above setup requires that the entire contents of `/u01/stage/patch/dbserverpatch` be owned by `root`.
- If you update `dbserver.patch.zip`, then you must place the new version in the same location as specified by `sudoers`.

To verify that the setup is correct, run `patchmgr` in `prereq check` mode as the `oracle` user:

```
[oracle]$ cd /u01/stage/patch/dbserverpatch/

[oracle]$ sudo ./patchmgr -dbnodes dbgroup -precheck \
 -yum_repo http://my-yum-repo/yum/EngineeredSystems/exadata/dbserver/12.1.2.1.3/base/
x86_64/ \
 -target_version 12.1.2.1.3.151021
```

## 8.4.7.2 Running `dbnodeupdate.sh` Using `sudo`

Before using `dbnodeupdate.sh` using `sudo`, you configure the `/etc/sudoers` file.

1. Log in as the `root` user and edit `/etc/sudoers` using `visudo`.

```
visudo
```

2. Add the following entry (all on one line) to the bottom of the `sudoers` file to allow non-root users, such as the `oracle` user, to run `dbnodeupdate.sh` as `root`:

#### Note:

The first field in the line specifies the non-root user who is granted `sudo` access for the `dbnodeupdate.sh` command. The line below uses the `oracle` user as an example. You can specify a different user if necessary.

```
oracle ALL=(ALL) NOPASSWD:SETENV: /u01/stage/patch/dbnodeupdate/
dbnodeupdate.sh
```

3. As root, create the `/u01/stage/patch/dbnodeupdate` directory and unzip `dbnodeupdate.zip`:

```
mkdir -p /u01/stage/patch/dbnodeupdate
cp dbnodeupdate.zip /u01/stage/patch/dbnodeupdate
cd /u01/stage/patch/dbnodeupdate
unzip dbnodeupdate.zip
```

To verify that the setup is correct, run `dbnodeupdate.sh` in `prereq` check mode as the `oracle` user:

```
[oracle]$ cd /u01/stage/patch/dbnodeupdate

[oracle]$ sudo ./dbnodeupdate.sh -u -l http://my-yum-repo/yum/
EngineeredSystems/exadata/dbserver/12.1.2.1.3/base/x86_64/ -v
```

`dbnodeupdate` exits if it is run without `root` privileges.



#### Note:

- The above setup requires that everything in `/u01/stage/patch/dbnodeupdate` be owned by the `root` user.
- If you update the `dbnodeupdate` utility, then you must place the new version in the same location as specified by `sudoers`.

## 8.5 Exadata Patchmgr Update Utility

Patchmgr is the utility used to update software for Exadata infrastructure components.

- [About the Exadata Patchmgr Update Utility](#)  
Patchmgr is designed to simplify the software update process.
- [Obtaining Patchmgr](#)  
You can download the patchmgr utility from My Oracle Support.
- [Patchmgr Syntax](#)  
Patchmgr is a utility used to update software for Exadata Database Machine infrastructure components.

### 8.5.1 About the Exadata Patchmgr Update Utility

Patchmgr is designed to simplify the software update process.

Patchmgr has the following capabilities:

- With a single invocation updates all Exadata Database Machine storage servers, database servers, or RDMA Network Fabric switches.
- Orchestrates the software update across components one at a time when performing a rolling update.

- Parallelizes the software update to all components at the same time when performing a non-rolling update.
- Updates all component software, as required, such as firmware, operating system, and Oracle Exadata System Software.
  - Database servers contain all software necessary to run Oracle Database and are configured to run Oracle Database optimally. However, it may be necessary to manually install additional, site-specific software, such as monitoring agents or backup agents. It is supported to manually customize database servers, but note that customizing the operating system by adding or updating packages may require additional actions when applying a future Oracle Exadata System Software update with patchmgr.
  - Storage server or RDMA Network Fabric switch configuration or installed software may not be altered manually (update without using patchmgr) in any way unless Exadata Database Machine documentation contains steps to perform the desired change.
- Patchmgr may be run from an Engineered System running Oracle Linux, as the `root` user or a `non-root` user.
- Multiple invocations of patchmgr may be run from the same software directory.

When updating database servers, patchmgr manages the following as needed:

- Stops and starts databases and clusterware
- Stops and starts user domains (domUs)
- Stops and starts Oracle Enterprise Manager Cloud Control agents
- Unmounts remote network mounts
- Performs a root file system operating system backup that can be used for rollback
- Relinks database home and Oracle Grid Infrastructure home binaries
- Applies updated best practices configuration changes and workarounds for known issues

## 8.5.2 Obtaining Patchmgr

You can download the patchmgr utility from My Oracle Support.

Before Oracle Exadata System Software release 19.3.0, the server and RDMA Network Fabric switch updates use the version of patchmgr bundled with the Oracle Exadata System Software release.

Starting with Oracle Exadata System Software release 19.3.0, there are separate patchmgr distributions for the servers and RDMA Network Fabric switches.

By default, the appropriate patchmgr distribution is included with the relevant server or switch update. However, it is recommended to *always* use the latest patchmgr from My Oracle Support when updating Exadata Database Machine servers or switches. The patchmgr utility is updated frequently to address known issues and best practices, and to support new hardware. See My Oracle Support document 888828.1.

### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(Doc ID 888828.1\)](#)

## 8.5.3 Patchmgr Syntax

Patchmgr is a utility used to update software for Exadata Database Machine infrastructure components.

### Prerequisites

Patchmgr runs on the "driving system", which is an Exadata Database Machine database server or a non-Exadata Database Machine system running Oracle Linux. This allows patchmgr to run from a central server to update multiple Exadata Database Machine systems. If patchmgr runs from an Exadata Database Machine database server, then that database server cannot be in the group file supplied to patchmgr.

### Syntax

Patchmgr syntax differs depending on the component to be updated, however, the general patchmgr syntax is:

```
./patchmgr -component component_list_file -action required_arguments
[optional_arguments]
```

### Options

Option	Description
<i>component</i>	Represents the component to update. Valid values are <code>-cells</code> , <code>-dbnodes</code> , <code>-ibswitches</code> , or <code>-roceswitches</code>
<i>component_list_file</i>	A text file containing the host names of components to update. For example, if updating Exadata Database Machine storage servers (using <code>-cells</code> ), then <i>component_list_file</i> will contain the list of all host names of the storage servers.
<i>action</i>	The action to perform. In most cases the action is specific to the component being updated.
<i>required_arguments</i>	The additional arguments required for certain actions.
<i>optional_arguments</i>	Any optional arguments required for certain actions.

### Usage Notes

- Starting with Oracle Exadata System Software release 19.3, the options are prefixed with `--`. Prior to this release, the options were prefixed with `-`.
- Multiple invocations of `patchmgr` may be run concurrently from the same software directory by using the `-log_dir` option. This allows `patchmgr` to update multiple Exadata Database Machine systems concurrently from the same software directory.
- Patchmgr may be run as the `root` user or as a non-root user. To run `patchmgr` as a non-root user, the `-log_dir` option must be used.
- The user running `patchmgr` must have `root`-level SSH equivalence configured to the servers or switches that `patchmgr` will update. The SSH equivalence must be bi-

directional. That is, the user running `patchmgr` must be able to use SSH without a password to access the target servers or switches as the `root` user, and the `root` user on target servers or switches must be able to use SSH without a password to access the driving system as the user running `patchmgr`.

- Prior releases used the `dbnodeupdate.sh` utility to update database servers. `dbnodeupdate.sh` has been integrated with and is replaced by `patchmgr`.
- [Patchmgr Syntax for Storage Servers](#)  
You can use `patchmgr` to update software for Exadata Database Machine storage servers.
- [Patchmgr Syntax for Database Servers](#)  
You can use `patchmgr` to update software for Exadata Database Machine database servers.
- [Patchmgr Syntax for RoCE Network Fabric Switches](#)  
You can use `patchmgr` to update software for RoCE Network Fabric switches.
- [Patchmgr Syntax for InfiniBand Network Fabric Switches](#)  
You can use `patchmgr` to update software for InfiniBand Network Fabric switches.
- [Patchmgr Syntax for the Management Network Switch](#)  
You can use `patchmgr` to perform firmware upgrades on the 9000 series Management Network Switch found on Oracle Exadata Database Machine X7-2 and later systems.

### 8.5.3.1 Patchmgr Syntax for Storage Servers

You can use `patchmgr` to update software for Exadata Database Machine storage servers.

#### Prerequisites

`Patchmgr` is run on the "driving system", which is an Exadata Database Machine database server or a non-Exadata Database Machine system running Oracle Linux. This allows `patchmgr` to run from a central server to update multiple Exadata Database Machine systems

#### Patchmgr Syntax for Storage Servers

```
./patchmgr --cells cell_host_file
{ { --patch_check_prereq | --rollback_check_prereq } [--rolling] [--
unkey] [--ignore_alerts] [--ignore_date_validations] |
 { --patch | --rollback } [--rolling] [--unkey] [--ignore_alerts] |
 --cleanup [--unkey] }
[--log_dir { log_directory | auto }]
```

## Main Arguments

Argument	Description
<code>--cells <i>cell_host_file</i></code>	The <i>cell_host_file</i> is a text file containing the host names of the storage servers to update. The file contains one storage server host name or IP address per line. The storage server patching will fail if the list file is not specified.
<code>--patch</code>	Apply the patch, including firmware updates, wherever possible (BIOS, Disk Controller and if possible disk drives) to all storage servers in the storage server list file.
<code>--rollback</code>	Rolls back the patch.
<code>--cleanup</code>	Cleans up all patch files and temporary content on all storage servers. Before cleaning up, collects logs and information for problem diagnostics and analysis. Cleaning up patch files can be done manually if patch fails by removing the directory <code>/root/_cellupd_dpullec_</code> on each storage server.

## Supported Options

The following options are supported for storage server patching and rollback:

**Table 8-2 Patchmgr Options for Storage Servers**

Option	Description
<code>--get <i>property</i></code>	Get information about specific property. Property can be: <code>log_dir</code> - assigned directory for all log files.
<code>--ignore_alerts</code>	Ignore any active hardware alerts on the Exadata storage server and proceed with the patching.
<code>--ignore_date_validations</code>	Ignore date version validation. This allows upgrades to later releases with an earlier date stamp, for example <code>19.1.2.0.0.190306 -&gt; 19.2.0.0.0.1</code> .
<code>--log_dir ( <i>log_directory</i>   auto )</code>	Specifies the absolute path to the log directory, or you can specify <code>auto</code> to use a log directory that is based on the launch directory and the content of the nodes list file.  Specifying the <code>-log_dir</code> option enables you to run multiple patch manager invocations and also to run patch manager as a non-root user.
<code>--patch_check_prereq</code>	Runs prerequisite check on all the storage servers to determine if the patch can be applied to the storage servers.
<code>--rollback_check_prereq</code>	Runs prerequisite check on all the storage servers to determine if the storage servers can be rolled back for this specific patch.



Table 8-2 (Cont.) Patchmgr Options for Storage Servers

Option	Description
--rolling	Specifies that the update is to be done in a rolling fashion. If not specified, the update is done in a non-rolling fashion.  Environment variable EXA_PATCH_ACTIVATE_TIMEOUT_SECONDS controls the timeout value waiting for the grid disks to be activated. The default is set to 36000 (10 hours).  <b>Note:</b> Prerequisite checks and cleanups are always done in a non-rolling fashion, even if the --rolling option is specified.
--smtp_from "email_addr"	Specifies the <b>from</b> email address for the patchmgr notification.
--smtp_to "email_addr1 email_addr2 email_addr3 ..."	Specifies the <b>to</b> email addresses for the patchmgr notification.
--smtp_set_envelope_sender	Specifies that the same <b>from</b> address in Return-Path: mail header should be used.
--unkey	Removes passwordless SSH access to the cells before exit.

### Usage Notes

- Starting with Oracle Exadata System Software release 19.3, the options are prefixed with --. Prior to this release, the options were prefixed with -.
- Multiple invocations of patchmgr may be run concurrently from the same software directory by using the --log\_dir option. This allows patchmgr to update multiple Exadata Database Machine systems concurrently from the same software directory.
- Patchmgr may be run as the root user or as a non-root user. The user running patchmgr must have root-level SSH equivalence configured to the servers or switches that patchmgr will update. To run patchmgr as a non-root user, the --log\_dir option must be used.

### Example 8-4 Run storage server update pre-requisite checks, then update storage servers

```
./patchmgr -cells cell_group -patch_check_prereq
./patchmgr -cells cell_group -patch
```

### Example 8-5 Update storage servers in a rolling manner

The following command updates storage servers in a rolling manner, receive email notifications as the update proceeds, and remove passwordless SSH access to cells after the update is complete.

```
./patchmgr -cells cell_group -patch -rolling -unkey -smtp_from
"dbm01@example.com" -smtp_to "admin1@example.com,admin2@example.com"
```

### Example 8-6 Update multiple storage servers at one time

The following commands update storage servers on three Exadata Database Machine systems simultaneously from the same software directory; one update using a rolling update, the others updates are non-rolling.

In this example:

- Each patchmgr command must be run from a separate terminal window.
- Each patchmgr execution uses a unique logging directory name automatically generated based on the content of the *component\_list\_file*.

```
(terminal1) ./patchmgr -cells cell_group_exa01 -patch -rolling -log_dir auto
(terminal2) ./patchmgr -cells cell_group_exa02 -patch -log_dir auto
(terminal3) ./patchmgr -cells cell_group_exa03 -patch -log_dir auto
```

To have a subsequent patchmgr execution use an altered *component\_list\_file* with different content, yet use the same logging directory as a prior patchmgr execution, use the `-get log_dir` option to obtain the logging directory. For example:

1. The logging directory for the initial patchmgr execution is generated automatically.

```
./patchmgr -cells cell_group -patch -log_dir auto
```

2. Assume the last cell failed to update and patchmgr will be re-run for the last cell only, using the same logging directory as the initial patchmgr execution. Use the `-get log_dir` option to obtain the logging directory using the original *component\_list\_file*.

```
./patchmgr -cells cell_group -patch -log_dir auto -get log_dir
log_dir=/tmp/patch_12.2.1.1.1.170419/log/dm01cel01_dm01cel02_cacce4ee
```

3. Update the *cell\_group* file to contain only the last cell, or use a different file that contains only the last cell. Specify the logging directory from the initial patchmgr execution so all logs for this group of cells are created in the same logging directory.

```
./patchmgr -cells cell_group -patch -log_dir /tmp/
patch_12.2.1.1.1.170419/log/dm01cel01_dm01cel02_cacce4ee
```

#### Related Topics

- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

## 8.5.3.2 Patchmgr Syntax for Database Servers

You can use patchmgr to update software for Exadata Database Machine database servers.

#### Prerequisites

Patchmgr is run on the "driving system", which is an Exadata Database Machine database server or a non-Exadata Database Machine system running Oracle Linux. This allows patchmgr to run from a central server to update multiple Exadata Database Machine systems. If patchmgr is run from an Exadata Database Machine database server, then that database server cannot be in the file that lists the servers to patch.

## Patchmgr Syntax for Database Servers

```
./patchmgr --dbnodes database_node_file
{ --backup [--rolling] [--unkey] |
 --precheck { --yum_repo base_URL | --iso_repo zipped_iso_file } [--
target_version version] [--unkey] |
 --upgrade { --yum_repo base_URL | --iso_repo zipped_iso_file } [--
target_version version] [--rolling] [--unkey] |
 --complete [--target_version version] [--unkey] |
 --rollback [--rolling] [--unkey] |
 --cleanup [--unkey] }
[--log_dir { log_directory | auto }]
```

## Main Arguments

Argument	Description
<code>--dbnodes database_node_file</code>	The <i>database_node_file</i> is a text file containing the host names of the database servers to update. The file contains one database server host name or IP address per line. The database server patching will fail if the list file is not specified.
<code>--backup</code>	Perform backup for the database nodes specified in the host list. Runs backup in non-rolling fashion unless the <code>--rolling</code> option is specified.
<code>--precheck</code>	Runs the pre-upgrade validation checks in the database nodes specified in the host list in non-rolling fashion.
<code>--upgrade</code>	Upgrades the database nodes specified in the host list. Runs upgrade in non-rolling fashion unless the <code>--rolling</code> option is specified.
<code>--complete</code>	Runs 'completion-steps' only. In normal cases there is no need to run this separately as this already runs as part of <code>--upgrade</code> . <b>Note:</b> If the database stack or user domains are up, they will be shut down and restarted.
<code>--rollback</code>	Rollback the database nodes specified in the host list. Runs rollback in non-rolling fashion unless the <code>--rolling</code> option is specified.
<code>--cleanup</code>	Cleans up all temporary content on the database servers specified in the host list in non-rolling fashion.

## Supported Options

The following options are supported for storage server patching and rollback:

Table 8-3 Patchmgr Options for Database Servers

Option	Description
<code>--allow_active_network_mounts</code>	Allows <code>dbnodeupdate</code> to run with active NFS or SMB mounts. This is equivalent to the <code>dbnodeupdate.sh -A</code> command.
<code>--allow_non_signed_repo</code>	Allow <code>dbnodeupdate</code> to run with a non-signed repository. (Oracle Exadata System Software release 19.2 and later)
<code>--dbnode_patch_base</code>	User preferred location on target database servers where patch iso image and <code>dbnodeupdate</code> archive files are to be unzipped. <b>Note:</b> Provided location must be an absolute path of local file system and it should have sufficient amount of free space and inodes.
<code>--force_remove_custom_rpms</code>	Remove any custom RPMs when database server upgrades from Oracle Linux 5 to Oracle Linux 6.
<code>--ignore_alerts</code>	Ignore any active hardware alerts on the Exadata server and proceed with the patching.
<code>--iso_repo repo_location</code>	The path to a zipped ISO file. It is recommended that the file be in the same directory as the <code>dbnodeupdate.zip</code> file. This option or the <code>-yum_repo</code> option must be specified for <code>-backup</code> , <code>-precheck</code> and <code>-upgrade</code> actions. This option cannot be used with the <code>-yum_repo</code> option.
<code>--log_dir ( log_directory  auto )</code>	The absolute path to the log directory, or you can specify <code>auto</code> to use a log directory that is based on the directory you started <code>patchmgr</code> from and the content of the nodes list file. Specifying the <code>--log_dir</code> option enables you to run multiple patch manager invocations and also to run patch manager as a non-root user.
<code>--no_connection_draining</code>	Disables database connection draining for Fleet Patching and Provisioning, formerly known as Rapid Home Provisioning (RHP). The connection draining availability depends on the Oracle Grid Infrastructure release. This option is applicable to only rolling updates.
<code>--nobackup</code>	Do not backup the database servers before an upgrade.

**Table 8-3 (Cont.) Patchmgr Options for Database Servers**

Option	Description
<code>--rolling</code>	Specifies that the update is to be done in a rolling fashion, one server at a time. If not specified, the update is done in a non-rolling fashion.  Environment variable <code>EXA_PATCH_ACTIVATE_TIMEOUT_SECONDS</code> controls the timeout value waiting for the grid disks to be activated. The default is set to 36000 (10 hours).  <b>Note:</b> Prerequisite checks and cleanups are always done in a non-rolling fashion, even if the <code>--rolling</code> option is specified.
<code>--rolling_backups</code>	Directs patchmgr to backup each node prior to updating that node in a rolling manner. If this option is not specified (default), then patchmgr completes the backups of all nodes in parallel before updating the first node.  This option can only be used in conjunction with the <code>--rolling</code> option. Otherwise, this option is ignored and the default backup method is used.  This option is ignored if <code>--nobackup</code> is included in the command.
<code>--skip_gi_db_validation</code>	Skip certification of Oracle Grid Infrastructure and Oracle Database home compatibility with Oracle Linux 7.
<code>--smtp_from "email_addr"</code>	Specifies the <b>from</b> email address for the patchmgr notification.
<code>--smtp_to "email_addr1 email_addr2 email_addr3 ..."</code>	Specifies the <b>to</b> email addresses for the patchmgr notification.
<code>--smtp_set_envelope_sender</code>	Specifies that the same <b>from</b> address in Return-Path: mail header should be used.
<code>--target_version version</code>	The patch version, for example 12.1.2.3.0.160207.3.
<code>--unkey</code>	Removes passwordless SSH access to the cells before exit.
<code>--yum_repo base_URL</code>	Base URL for the Exadata update repository. This option must be specified for <code>--backup</code> , <code>--precheck</code> and <code>--upgrade</code> actions if the <code>--iso_repo</code> option is not specified. This option cannot be combined with the <code>--iso_repo</code> option.

### Usage Notes

- Starting with Oracle Exadata System Software release 19.3, the options are prefixed with `--`. Prior to this release, the options were prefixed with `-`.
- Prior releases used the `dbnodeupdate.sh` utility to update database servers. `dbnodeupdate.sh` has been integrated with and is replaced by `patchmgr`.

## Examples

### Example 8-7 Backup database servers then perform the update

The following commands backup the Exadata Database Machine database servers, run the prerequisite checks for the database server update, and then update the database servers in a rolling manner.

```
./patchmgr -dbnodes dbnode_group -backup -iso_repo /tmp/
p25640941_122111_Linux-x86-64.zip
-target_version 12.2.1.1.1.170419
```

```
./patchmgr -dbnodes dbnode_group -precheck -iso_repo /tmp/
p25640941_122111_Linux-x86-64.zip
-target_version 12.2.1.1.1.170419
```

```
./patchmgr -dbnodes dbnode_group -upgrade -nobackup -iso_repo /tmp/
p25640941_122111_Linux-x86-64.zip
-target_version 12.2.1.1.1.170419 -rolling
```

## Related Topics

- [Setting up SSH Equivalence](#)

Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

## 8.5.3.3 Patchmgr Syntax for RoCE Network Fabric Switches

You can use patchmgr to update software for RoCE Network Fabric switches.

### Prerequisites

Patchmgr is run on the "driving system", which is an Exadata Database Machine database server or a non-Exadata Database Machine system running Oracle Linux. This allows patchmgr to run from a central server to update multiple Exadata Database Machine systems.

#### Note:

Prior to Oracle Exadata System Software release 19.3.9, you must run `patchmgr` as a non-root user for patching RoCE Network Fabric switches.

### Patchmgr Syntax for RoCE Network Fabric Switches

```
./patchmgr --roceswitches [roceswitch_list_file]
{ --upgrade [--roceswitch-precheck] [--unkey] [--force] |
 --downgrade [--roceswitch-precheck] [--unkey] [--force] |
 --apply-config [--unkey] [--force] |
 --verify-config [--newswitchlist new_list_file] [--unkey] }
[-log_dir { absolute_path_to_log_directory | AUTO }]
```

## Main Arguments

Argument	Description
<code>--roceswitches</code> <code>[roceswitch_list_file]</code>	<p>Specifies that patchmgr is acting on the RoCE Network Fabric switches.</p> <p>If specified, the switch list file identifies the RoCE Network Fabric switches.</p> <p>In its simplest form, the file has one switch host name or IP address on each line. In this case, each switch is assumed to be a leaf switch in a single-rack Exadata environment. For example:</p> <pre>switch123-rocea0 switch123-roceb0</pre> <p>To specify the configuration type for each switch, append a colon (:) and tag to each switch host name or IP address in the switch list file. The following tags are supported:</p> <ul style="list-style-type: none"><li>• <code>leaf</code> - Identifies a leaf switch in a single rack system. This configuration type is assumed if no tag is specified.</li><li>• <code>mleaf</code> - Identifies a leaf switch in a multi-rack system.</li><li>• <code>sflleaf</code> - Identifies a leaf switch in a single rack system that is enabled to support Exadata Secure RDMA Fabric Isolation.</li><li>• <code>msflleaf</code> - Identifies a leaf switch in a multi-rack system that is enabled to support Exadata Secure RDMA Fabric Isolation.</li><li>• <code>m spine</code> - Identifies a spine switch. Note that one spine switch configuration supports all spine switches on single and multi-rack systems, with and without Exadata Secure RDMA Fabric Isolation.</li><li>• <code>leaf23</code> - Identifies a leaf switch in a single rack system that is configured with 23 host ports. This configuration is required only for 8-socket systems (X8M-8 and later) with 3 database servers and 11 storage servers.</li><li>• <code>mleaf23</code> - Identifies a leaf switch in a multi-rack system that is configured with 23 host ports. This configuration is required only for 8-socket X8M-8 systems with 3 database servers and 11 storage servers.</li><li>• <code>mleaf_u14</code> - Identifies a leaf switch in a multi-rack system that is configured with 14 inter-switch links. This is the typical multi-rack leaf switch configuration for X9M systems.</li></ul>

Argument	Description
	<ul style="list-style-type: none"><li>• msfleaf_u14 - Identifies a leaf switch in a multi-rack system that is enabled to support Exadata Secure RDMA Fabric Isolation and is configured with 14 inter-switch links. This configuration is required for X9M systems with Secure Fabric enabled.</li><li>• mleaf23_u13 - Identifies a leaf switch in a multi-rack system that is configured with 23 host ports and 13 inter-switch links. This configuration is required only for 8-socket X9M-8 systems with three database servers and 11 storage servers.</li></ul>
	For example:
	<pre>switch123-rocea0:leaf switch123-roceb0:leaf</pre>
	For multi-rack configurations only, also specify a unique loopback octet for each switch.
	The loopback octet is the last octet of the switch loopback address, which uniquely identifies a switch.
	To specify the loopback octet for each switch, append a period (.) and numeric loopback octet value to each tagged switch entry in the switch list file. The range of valid loopback octet values is:
	<ul style="list-style-type: none"><li>• 101-118 for leaf switches</li><li>• 201-208 for spine switches</li></ul>
	For example, the switch list file for a 2-rack system might contain:
	<pre>rack1sw-rocea0:mleaf.101 rack1sw-roceb0:mleaf.102 rack1sw-roces0:mspine.201 rack2sw-rocea0:mleaf.103 rack2sw-roceb0:mleaf.104 rack2sw-roces0:mspine.202</pre>
	If no file name is provided, then the command acts on all RoCE Network Fabric switches discovered from the host that is running patchmgr.



Argument	Description
<code>--upgrade</code>	<p>Upgrade the firmware on the RoCE Network Fabric switches.</p> <p>If required, this option also installs or upgrades the client software component that propagates switch alerts to Oracle Auto Service Request (ASR).</p> <p><b>Note:</b> Commencing with the August 2022 <code>patchmgr</code> release, <code>patchmgr</code> performs an additional series of checks on the RoCE Network Fabric. The checks occur immediately before any firmware upgrade and also during prerequisite checking using the <code>--roceswitch-precheck</code> option. These checks mitigate the risks of failure associated with unexpected problems in the RoCE Network Fabric. For example, if one of the RoCE Network Fabric ports on a storage server is down, the storage server would become unavailable if the switch connected to the only operational port is taken offline for an upgrade. If any check fails, <code>patchmgr</code> reports the problem and ends immediately. In this case, you must correct the problem with the RoCE Network Fabric before you can perform the upgrade.</p>
<code>--downgrade</code>	<p>Downgrade the firmware on the RoCE Network Fabric switches.</p> <p><b>Note:</b> Commencing with the August 2022 <code>patchmgr</code> release, <code>patchmgr</code> performs an additional series of checks on the RoCE Network Fabric. The checks occur immediately before any firmware downgrade and also during prerequisite checking using the <code>--roceswitch-precheck</code> option. These checks mitigate the risks of failure associated with unexpected problems in the RoCE Network Fabric. For example, if one of the RoCE Network Fabric ports on a storage server is down, the storage server would become unavailable if the switch connected to the only operational port is taken offline for an upgrade. If any check fails, <code>patchmgr</code> reports the problem and ends immediately. In this case, you must correct the problem with the RoCE Network Fabric before you can perform the downgrade.</p>
<code>--apply-config</code>	<p>Applies the golden configuration template to each switch.</p> <p>This option relies on tags in the switch list file, which specify the configuration type for each switch.</p>

Argument	Description
<code>--verify-config [ --newswitchlist new_list_file ]</code>	<p>Verify each switch configuration against the golden configuration.</p> <p>Verification is performed automatically when using <code>--upgrade</code> or <code>--downgrade</code>. You can use this option to perform verification as a standalone operation.</p> <p>If specified, the <code>--newswitchlist</code> option generates a new switch list file with entries that match the current configuration of each switch.</p>

### Supported Options

The following options are supported for RoCE Network Fabric switch configuration and firmware update:

**Table 8-4 Patchmgr Options for RoCE Network Fabric Switches**

Option	Description
<code>--roceswitch-precheck</code>	Performs switch firmware upgrade or downgrade simulation on the RoCE Network Fabric switches in the list file but does not perform the actual install. Use this option with <code>--upgrade</code> or <code>--downgrade</code> .
<code>--unkey</code>	In conjunction with <code>--upgrade</code> or <code>--downgrade</code> , this option removes the configuration settings that enable passwordless SSH access to the RoCE Network Fabric switch.
<code>--force</code>	In conjunction with <code>--upgrade</code> or <code>--downgrade</code> , this option proceeds with the upgrade or downgrade even if the switch is already on target firmware version or the RoCE Network Fabric switch is experiencing non-critical failures.  In conjunction with <code>--apply-config</code> , this option bypasses the check that determines if the current switch configuration matches the configuration type specified in the switch list file.
<code>-log_dir ( absolute_path_to_log_directory   AUTO )</code>	<p>When running patchmgr as a non-root user, use <code>-log_dir</code> to specify the absolute path to the log directory or use the keyword <code>AUTO</code>. If you specify <code>AUTO</code>, then patchmgr generates and sets the path to the log directory based on the directory patchmgr is launched from and the content of the nodes list file.</p> <p><b>Note:</b> Specifying <code>-log_dir</code> enables multiple patch manager invocations and is required when running patch manager as a non-root user.</p>

### Usage Notes

- Starting with Oracle Exadata System Software release 19.3, the options are prefixed with `--`. Prior to this release, the options were prefixed with `-`.

**Example 8-8 Using patchmgr to Upgrade Firmware on RoCE Network Fabric Switches**

This example runs the upgrade prerequisite checks on all detected switches, then upgrades the switches.

```
$./patchmgr --roceswitches --upgrade --roceswitch-precheck
$./patchmgr --roceswitches --upgrade
```

**Example 8-9 Using patchmgr to Apply Golden Configurations to RoCE Network Fabric Switches**

This example applies golden configuration settings to RoCE Network Fabric switches as specified in the `switches.lst` file. Logs for the operation are written to `/tmp/switchlogs`.

```
$ cat switches.lst
switch456-rocea0:leaf
switch456-roceb0:leaf

$./patchmgr --roceswitches switches.lst --apply-config -log_dir /tmp/switchlogs
```

### 8.5.3.4 Patchmgr Syntax for InfiniBand Network Fabric Switches

You can use patchmgr to update software for InfiniBand Network Fabric switches.

**Prerequisites**

Patchmgr is run on the "driving system", which is an Exadata Database Machine database server or a non-Exadata Database Machine system running Oracle Linux. This allows patchmgr to run from a central server to update multiple Exadata Database Machine systems.

**Patchmgr Syntax for InfiniBand Network Fabric Switches**

```
./patchmgr --ibswitches [ibswitch_list_file]
{ --upgrade | --downgrade } [--ibswitch_precheck] [--unkey] [--force
[yes | no]]
```

**Main Arguments**

Argument	Description
<code>--ibswitches <i>ibswitch_list_file</i></code>	Specifies the name of the InfiniBand Network Fabric switch list file. The file has one switch host name or IP address per line. If no file name is provided, then it runs the command on all InfiniBand Network Fabric switches discovered from this host.

Argument	Description
--upgrade	Upgrade the InfiniBand Network Fabric switches in the list file to \$EXADATA_IMAGE_IBSWITCH_UPGRADE_VERSION.
--downgrade	Downgrade the InfiniBand Network Fabric switches in the list file to \$EXADATA_IMAGE_IBSWITCH_DOWNGRADE_VERSION.

### Supported Options

The following options are supported for InfiniBand Network Fabric switch configuration and firmware update:

**Table 8-5 Patchmgr Options for InfiniBand Network Fabric Switches**

Option	Description
--force [yes   no]	Specifies to proceed with the upgrade or downgrade even on non-critical failures. The value <code>yes</code> specifies to proceed by auto-answering prompts with <code>yes</code> . The value <code>no</code> specifies to proceed by auto-answering prompts with <code>no</code> . The default value is <code>yes</code> .
--ibswitch_precheck	Runs the pre-update validation checks on the InfiniBand Network Fabric switches in the list file.
--smtp_from "email_addr"	Specifies the <b>from</b> email address for the patchmgr notification.
--smtp_to "email_addr1 email_addr2 email_addr3 ..."	Specifies the <b>to</b> email addresses for the patchmgr notification.
--smtp_set_envelope_sender	Specifies that the same <b>from</b> address in Return-Path: mail header should be used.
--unkey	Removes passwordless SSH access to the InfiniBand Network Fabric switches before exit.

### Usage Notes

- Starting with Oracle Exadata System Software release 19.3, the options are prefixed with `--`. Prior to this release, the options were prefixed with `-`.

### Example 8-10 Using patchmgr for InfiniBand Network Fabric Switches

This example runs the update prerequisite checks on all switches, then upgrades the switches specified in the `ib_group` file.

```
./patchmgr --ibswitches --upgrade --ibswitch_precheck
./patchmgr --ibswitches ib_group --upgrade
```

### 8.5.3.5 Patchmgr Syntax for the Management Network Switch

You can use patchmgr to perform firmware upgrades on the 9000 series Management Network Switch found on Oracle Exadata Database Machine X7-2 and later systems.

#### Prerequisites

Patchmgr is run on the "driving system", which is an Exadata Database Machine database server or a non-Exadata Database Machine system running Oracle Linux. This allows patchmgr to run from a central server to update multiple Exadata Database Machine systems.

#### Patchmgr Syntax for Management Network Switch

```
./patchmgr --adminswitches [adminswitch_list_file]
{ --upgrade | --downgrade } [--adminswitch-precheck] [--unkey] [--
force]
[-log_dir { absolute_path_to_log_directory | AUTO }]
```

#### Main Arguments

Argument	Description
--adminswitches [adminswitch_list_file]	Specifies that patchmgr is acting on the Management Network Switches.  If specified, the switch list file identifies the Management Network Switches. The file has one switch host name or IP address on each line.  If no file name is provided, then the command acts on all Management Network Switches discovered from the host that is running patchmgr.
--upgrade	Upgrade the firmware on the Management Network Switches.
--downgrade	Downgrade the firmware on the Management Network Switches.

#### Supported Options

The following options are supported for Management Network Switch firmware update:

**Table 8-6 Patchmgr Options for Management Network Switches**

Option	Description
--adminswitch-precheck	Performs switch firmware upgrade or downgrade simulation on the Management Network Switches in the list file but does not perform the actual install. Use this option with --upgrade or --downgrade.

**Table 8-6 (Cont.) Patchmgr Options for Management Network Switches**

Option	Description
<code>--unkey</code>	In conjunction with <code>--upgrade</code> or <code>--downgrade</code> , this option removes the configuration settings that enable passwordless SSH access to the Management Network Switch.
<code>--force</code>	In conjunction with <code>--upgrade</code> or <code>--downgrade</code> , this option proceeds with the upgrade or downgrade even if the switch is already on target firmware version or the Management Network Switch is experiencing non-critical failures.
<code>-log_dir</code> ( <i>absolute_path_to_log_directory</i>   <code>AUTO</code> )	When running <code>patchmgr</code> as a non-root user, use <code>-log_dir</code> to specify the absolute path to the log directory or use the keyword <code>AUTO</code> . If you specify <code>AUTO</code> , then <code>patchmgr</code> generates and sets the path to the log directory based on the directory <code>patchmgr</code> is launched from and the content of the switch list file.  <b>Note:</b> Specifying <code>-log_dir</code> enables multiple patch manager invocations and is required when running patch manager as a non-root user.

**Example 8-11 Using patchmgr to Upgrade Firmware on Management Network Switches**

This example runs the upgrade prerequisite checks on the switch specified in the `switches.lst` file, then upgrades the switch.

```
$ cat switches.lst
dbm0sw-adm0

$./patchmgr --adminswitches switches.lst --upgrade --adminswitch-precheck

$./patchmgr --adminswitches switches.lst --upgrade
```

## 8.6 Updating Exadata Database Machine Database Servers

Use the following information and procedures when updating database servers within Exadata Database Machine.

- [Overview of Exadata Database Machine Database Server Updates](#)  
When updating database servers, there is more than one software that needs to be updated.
- [Update Paths for Exadata Database Servers](#)

- [Installing, Updating, and Managing Non-Oracle Software](#)  
Installing and updating non-Oracle branded RPMs on Exadata Database Machine database servers are allowed as long as kernel and RDMA Network Fabric packages remain untouched.
- [Customization Levels and Impact](#)
- [Update Utility for Exadata Database Servers](#)  
Patchmgr is the update utility for updating Oracle Exadata database servers.
- [Update Tool Execution Host](#)
- [Running the Update Utility as a Non-root User and Running Multiple Invocations Concurrently](#)
- [Recommended Timeline for Updating Exadata Database Servers](#)
- [Preparing and Populating the YUM Repository with the Oracle Exadata Channel Content](#)
- [Managing Exadata Obsolete Packages](#)
- [Updating Individual Packages](#)
- [Running Prerequisite Checks](#)
- [Backing up Exadata Database Servers Before Planned Maintenance](#)  
Oracle recommends making a backup of the Exadata database servers before making any changes when updating to a next release.
- [Running the Update](#)  
You can perform the actual update of Exadata database servers in a rolling (using the `-rolling` flag) or non-rolling fashion. The default is non-rolling.
- [Rolling Back Updates](#)

## 8.6.1 Overview of Exadata Database Machine Database Server Updates

When updating database servers, there is more than one software that needs to be updated.

Exadata Database Machine database server release updates contain updates for the following components within a database server:

- Oracle Linux operating system
- Firmware (Disk, RAID controller, ILOM, HCA)
- Oracle Exadata System Software

The software and firmware components that are updated for a specific release depend on the current Oracle Exadata System Software release the database server is running and the release you are updating to. Oracle Linux operating system packages and Oracle Exadata System Software are always updated while firmware may be updated for only a small selection of the components or not at all.

Updates for Exadata Database Machine database servers can be applied independently from the storage servers or RDMA Network Fabric switches unless otherwise specified in My Oracle Support Note 888828.1.

Updating database servers is always performed in-place. This means the active operating system is updated. The actual update is performed using YUM, but the

command to do this is wrapped in an Exadata Database Machine utility called patchmgr.

The YUM command is wrapped in the update utility to maintain strict ordering of validation and preparation steps during the update process. Also by using a utility Oracle can enforce application of fixes for known issues and best practices.

Oracle Clusterware processes and Oracle Real Application Clusters (Oracle RAC) database instances must not be running on a database server that is being updated. To reduce application-level impact, follow the client failover best practices described in the "Client Failover Best Practices for Highly Available Oracle Databases" technical reference paper.

If you cannot afford cluster-wide downtime, you can update database servers in a rolling fashion. This means updating one database server at a time. If you can afford cluster-wide downtime, you can update all database servers in parallel. Non-rolling updates reduce the overall time required to complete the upgrade at the expense of having a full database outage.

**Related Topics**

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Client Failover Best Practices for Highly Available Oracle Databases](#)

## 8.6.2 Update Paths for Exadata Database Servers

To update Exadata database servers running Oracle Linux 5.5 or later and Oracle Exadata software release 11.2.2.4.2 or later, you must use the "update utility" for applying Exadata software updates.

To update Exadata database servers running Exadata software releases older than 11.2.2.4.2, you should first update the servers to Oracle Linux 5.5 and Exadata Software release 11.2.2.4.2 or later before you can use the update utility. See the table below for details.

**Table 8-7 Update Paths**

Currently Installed Oracle Exadata Software Release	Currently Installed Oracle Linux Release	Recommended Action
Release 11.2.2.4.2 or later	Release 5.5. or later	Update to the target release using the update utility.
Releases earlier than 11.2.2.4.2	Release 5.5. or later	<ol style="list-style-type: none"> <li>1. Use patch 13513611 to update the database servers to release 11.2.2.4.2. Refer to the steps in section 8 of the patch 13513611 README.</li> <li>2. Update to a later release using the update utility.</li> </ol>



Table 8-7 (Cont.) Update Paths

Currently Installed Oracle Exadata Software Release	Currently Installed Oracle Linux Release	Recommended Action
Releases earlier than 11.2.2.4.2	Releases earlier than 5.5	<ol style="list-style-type: none"> <li>1. Update to release 11.2.3.2.0. Refer to My Oracle Support note 1284070.1 for details.</li> <li>2. Update to a later release using the Update utility.</li> </ol>

**Related Topics**

- [Updating key software components on database hosts to match those on the cells \(My Oracle Support Doc ID 1284070.1\)](#)

### 8.6.3 Installing, Updating, and Managing Non-Oracle Software

Installing and updating non-Oracle branded RPMs on Exadata Database Machine database servers are allowed as long as kernel and RDMA Network Fabric packages remain untouched.

Note that customizing the operating system by adding or updating packages may introduce problems when applying an Oracle Exadata System Software update because the non-Oracle software may add new dependencies which will not be provided by the Oracle Exadata System Software update. For this reason it is recommended to stay close to the Exadata Database Machine image and customize as little as possible.

 **Note:**

Exadata Database Machine does not ship with 32-bit software. This means any customization that introduces 32-bit RPMs on the system will break the Oracle Exadata System Software update and will need to be removed.

In case you have added or updated packages supplied by the Exadata Database Machine image, you can remove the `exadata-sun-computenode-exact` RPM when needed as follows:

```
[root@dm01]# rpm -e exadata-sun-computenode-exact
```

If you have installed customized packages, it is recommended that you have scripts to automate the removal (to run before updating Oracle Exadata System Software) and installation (to run after the Oracle Exadata System Software update) of those packages. After the Oracle Exadata System Software update, verify that the customized packages are still compatible and are still needed, before re-installing them.



**Note:**

Packages should not be forced in using the `rpm -Uvh --nodeps` command, unless directed by Oracle Support Services.

The same principle applies for customizations other than installation of additional software. Oracle recommends customizing the operating system as little as possible. Be aware that while changes on the database server are allowed within the earlier described boundaries, it is not possible for Oracle to anticipate each and every customization on every configuration item of the operating system.

There are certain common standards that Exadata Database Machine utilities depend on. These standards are based on best practices and have been validated over time. By moving away from Exadata Database Machine standards, systems might be missing optimal settings, and risk is introduced because it is difficult for Oracle to anticipate and validate software for one-off configurations. This is why customizations can cause unexpected results. As a result of adding software or customizing system configuration, you might require changes or additional steps to the standard Exadata Database Machine update process. With or without customizations, it is highly recommended to validate Oracle Exadata System Software updates on test systems before doing them on production systems.

**Related Topics**

- [Running Prerequisite Checks](#)

## 8.6.4 Customization Levels and Impact

The table below provides some guidance in the risks associated with customizing Exadata database servers. While the prerequisite check checks for most customizations, certain configuration changes might need to be rolled back in order to make updating the database server possible. If still needed, the changes can be restored after the update process.

**Table 8-8 Customizations and Their Risks**

Item	Customization Level	Risk
Default provided Engineered System Software Stack	None	None
Setting up interactive shell profiles for system users (root/oracle)	Low	High
Using all free space in VGExaDb	Low	Low
Installing additional (non-Exadata) rpm packages	Medium	Low
Customizing file system with different mount points	Medium	Low
Updating packages shipped with the current Exadata image	Medium	Low
Customizing configuration files, changing basic operating system functionality	High	Medium

**Table 8-8 (Cont.) Customizations and Their Risks**

Item	Customization Level	Risk
Installing additional (non-Exadata) non-rpm packages	High	Medium
Changing LVM layout	High	High

## 8.6.5 Update Utility for Exadata Database Servers

Patchmgr is the update utility for updating Oracle Exadata database servers.

Starting with release 12.2.1.1.0, Exadata software updates for the Exadata database server can be applied only through the patchmgr utility.

### Note:

The patchmgr utility for updating Exadata database servers is not the same as the patchmgr script shipped with the Exadata software update.

The patchmgr utility supports all hardware generations and Exadata storage server releases starting with 11.2.3.1.0, Exadata database servers running Oracle Virtual Server (dom0), and Exadata Virtual Machines (domU). The README files for the Oracle Exadata System Software updates specify whether the update itself is applicable for a particular hardware generation or not. The README is not shipped with `dbserver_patch.zip` but with the Oracle Exadata System Software update zip file.

The utility also supports Exadata software updates from Oracle Linux 5 to Oracle Linux 6 or updates from Oracle Linux 6 to Oracle Linux 7.

The utility takes care of the orchestration. You can perform the update in a rolling or non-rolling fashion across one or multiple Exadata database servers.

The update utility performs the following tasks:

- Automates all preparation, update, and validation steps, including:
  - Stopping the databases, Grid Infrastructure stack or domU's
  - Stopping Oracle Enterprise Manager Cloud Control agents,
  - Un-mounting remote network mounts (when required)
- Uses the built-in `dbserver_backup.sh` script to perform a backup of the file system hosting the operating system before updating the Exadata database server.
- Applies Oracle best practices and fixes for the latest known issues.
- Verifies that the update was successful, relinks the Oracle binaries, and starts the Oracle stack and domU's

### Related Topics

- [Obtaining Patchmgr](#)  
You can download the patchmgr utility from My Oracle Support.

- [Backing up Exadata Database Servers Before Planned Maintenance](#)  
Oracle recommends making a backup of the Exadata database servers before making any changes when updating to a next release.
- [dbnodeupdate.sh and dbserver.patch.zip: Updating Exadata Database Server Software using the DBNodeUpdate Utility and patchmgr \(My Oracle Support Doc ID 1553103.1\)](#)

## 8.6.6 Update Tool Execution Host

If you are planning to update all Exadata database servers at once, it is a requirement to run the update utility from a Linux node outside the group of Exadata database servers being updated. This is because the update utility cannot update the Exadata database server it is currently running on. If you have no other systems running Oracle Linux or Oracle Solaris you can run the update utility from one of the Exadata database servers. In such cases be sure that the Exadata database server where the update utility is running is not listed in the `dbns_group` file you specify.

You need to set up ssh equivalence for the root user from the driving node to the root user of all Exadata database servers that will be updated.

### Related Topics

- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

## 8.6.7 Running the Update Utility as a Non-root User and Running Multiple Invocations Concurrently

By default the update utility assumes you want to run as root. It is however possible to run the update utility as a non-root user from a remote host. It is also possible to run multiple invocations at the same time. This allows you to update multiple logical groups of Exadata Database Servers concurrently. To do this, you run the update utility with the `-log_dir` flag.

Ensure that ssh equivalence is set up for the current user to the root user of the Exadata database servers to be updated.

The following example shows the option of running as a non-root user and running multiple invocations.

```
[oracle@nonExadataHost]$./patchmgr -dbnodes ~/dbns_group -upgrade -
iso_repo /u01/ /p23557378_121223_Linux-x86-64.zip
-target_version 12.1.2.4.0.160710 -log_dir auto
```

`-log_dir` specifies the absolute path to the log directory or the keyword `auto` for the utility to generate and set a path to the log directory that is based on the launch directory and content of nodes list file. If you want to make sure you are using the same log directory in later invocations where the list of Exadata database servers changes, use the `-get log_dir` flag to obtain the `-log_dir` location used on previous sessions. For example, if the following command:

```
[oracle@nonExadata]$./patchmgr -dbnodes ~/dbns_group_test -log_dir auto -
get log_dir
```

This command returns output similar to the following:

```
log_dir=/u01/test/dbserver_patch_5.160715/log/dbm01_dbm02_e8f1f75
```

Use the `log_dir` value in subsequent commands. For example:

```
[oracle@nonExadata]$./patchmgr -dbnodes ~/dbs_group_test -precheck
-log_dir /u01/test/dbserver_patch_5.160715/log/dbm01_dbm02_e8f1f75
-iso_repo /u01/test/dbserver_patch_5.160715/p23557378_121223_Linux-
x86-64.zip
-target_version 12.1.2.2.3.160720 -allow_active_network_mounts
```

### Related Topics

- [Update Utility for Exadata Database Servers](#)

## 8.6.8 Recommended Timeline for Updating Exadata Database Servers

Oracle recommends the following timeline for updating Exadata database servers. By following this approach, you allow yourself time to perform any necessary corrective actions.



### Note:

Before making any changes, including prerequisite check, you should make a backup.

**Table 8-9 Timeline for Performing Updates**

When	Tasks
Weeks to days before the update	<ul style="list-style-type: none"> <li>• Download the latest <code>dbserver.patch.zip</code> from My Oracle Support note 1553103.1.</li> <li>• Download the latest <code>Exachk</code> from My Oracle Support note 1070954.1.</li> <li>• Research release-specific My Oracle Support note for known issues.</li> <li>• Research Exadata Critical Issues from My Oracle Support note 1270094.1.</li> <li>• Run <code>Exachk</code>.</li> <li>• Perform a first prerequisite check. For details, see <a href="#">Running Prerequisite Checks</a>.</li> </ul>
Just before the update	<ul style="list-style-type: none"> <li>• Download the latest <code>dbserver.patch.zip</code> from My Oracle Support note 1553103.1.</li> <li>• Download the latest <code>Exachk</code> from My Oracle Support note 1070954.1.</li> <li>• Research release-specific My Oracle Support note for known issues.</li> <li>• Research Exadata Critical Issues from My Oracle Support note 1270094.1.</li> <li>• Run <code>Exachk</code> and perform corrective actions as needed.</li> </ul>

**Table 8-9 (Cont.) Timeline for Performing Updates**

When	Tasks
At update time	<ul style="list-style-type: none"> <li>Perform a “backup only” run using the <code>-backup</code> flag.                      See <a href="#">Backing up Exadata Database Servers Before Planned Maintenance</a> for details.</li> <li>Perform a second prerequisite check.</li> <li>Remove any blocking rpms and re-run the prerequisite check to validate if all changes are complete.</li> <li>Perform the update. Use the <code>-nobackup</code> flag to skip the backup because you already made a “backup only” run.                      See <a href="#">Running the Update</a> for details.</li> </ul>
After the update	<ul style="list-style-type: none"> <li>Run <code>Exachk</code>.</li> <li>Reinstall any non-Exadata rpms that you removed before the Exadata update.</li> </ul>

### 8.6.9 Preparing and Populating the YUM Repository with the Oracle Exadata Channel Content

The Exadata software update procedure for Exadata database servers uses the Unbreakable Linux Network (ULN) for distribution. The updates are distributed as a set of packages (rpms) grouped in a channel. You need to ensure that these packages are available locally before performing the prerequisite check or update.

A channel can be subscribed to by a local non-Exadata database server. Channel content is then downloaded to the local server and the repository is made available as a YUM repository which is used to update Exadata database servers. If no Internet connection is available from the data center or if ULN synchronization is not possible, then a ready-to-use YUM repository can be downloaded as an ISO image.



**Note:**

Each Exadata Storage Server Update comes with its own channel/ISO image. This ISO image and ULN channel are not intended to be a complete replacement of a generic Linux ULN channel. The Exadata ISO and ULN channel contain only the Oracle Exadata channel content. If your environment uses other Linux packages for which updates are not contained in the Oracle Exadata ULN channel or ISO image, you should update those packages using Linux ULN channels as needed.

Use the following methods to build a YUM repository for hosting the Exadata software update. The update utility uses the local YUM repository to update the Exadata database servers.

- [Downloading the ISO Image as a YUM Repository from My Oracle Support and Passing the Location of the Compressed Image to the Update Utility](#)

- [Setting Up a Local Mirror as a YUM HTTP Repository from the ULN Oracle Exadata Channel on a Separate Server](#)
- [Downloading the ISO Image as a YUM Repository from My Oracle Support and Making It Available as a YUM HTTP Repository on a Web Server](#)

### 8.6.9.1 Downloading the ISO Image as a YUM Repository from My Oracle Support and Passing the Location of the Compressed Image to the Update Utility

The Oracle Exadata channel holding the Oracle Exadata software update is available as a downloadable image (a compressed ISO file) from My Oracle Support and can be used in two ways:

- Use the file location of the compressed file as an argument for the update utility using the `-iso_repo` flag
- Publish using HTTP and use the URL for the update utility using the `-yum_repo` flag.

The file location of the compressed file can be passed directly to the update utility which is recommended for the following conditions:

- There is no separate server available to be used as a repository YUM HTTP server.
- The Exadata database servers do not have customized software that requires updates from additional ULN channels.
- The simplest method is preferred.

#### Note:

The compressed ISO image should only exist on the node running the update utility. The distribution of the compressed ISO image is handled by the update utility

You can perform a prerequisite check (`-precheck`) to verify the repository is usable.

### 8.6.9.2 Setting Up a Local Mirror as a YUM HTTP Repository from the ULN Oracle Exadata Channel on a Separate Server

This method is recommended for the following conditions:

- There is a large number of Exadata database servers to update.
- The single repository server is accessible to all the Exadata database servers.
- An infrastructure exists for building a local ULN mirror on a separate Linux server.
- Exadata database servers have customized software that requires updates from other ULN channels.

The mirror server is a separate Linux server that holds the downloaded Exadata software update for Exadata database servers as downloaded from ULN. The Exadata

database servers connect to this local mirror (YUM HTTP repository) to retrieve updates.

 **Note:**

Do not use an Exadata database server as the local ULN mirror. Doing so may lead to dependency conflicts between the packages required by ULN to construct the repository, and the packages installed or updated on the Exadata database server by the Exadata software update. If no separate server is available, then use the ISO image method instead. See "[Downloading the ISO Image as a YUM Repository from My Oracle Support and Passing the Location of the Compressed Image to the Update Utility](#)".

When setting up a local repository for the first time, you may need additional Linux subscriptions, depending on the Linux release being used for the server that will be hosting the YUM repository. The server should subscribe to the following additional ULN channels in order for the repositories to be built:

- For 64-bit systems: Enterprise Linux 5 systems: `el5_x86_64_addons` are required.
- For 64-bit systems: Oracle Linux 5 systems: `el5_x86_64_addons` are required.
- For 64-bit systems: Oracle Linux 6 systems: `ol6_x86_64_addons` are required.

 **Note:**

After registration, depending on the operating system running your local ULN mirror, the system server automatically subscribes to `el5_x86_64_latest`, `ol5_x86_64_latest`, or `ol6_x86_64_latest`.

Review the prerequisites and server setup procedure as described in "How to create a local Unbreakable Linux Network mirror" at <http://www.oracle.com/technetwork/articles/servers-storage-admin/yum-repo-setup-1659167.html> to create and maintain a local ULN mirror as YUM repository. For step 5 in the above instructions, in addition to the listed channels, perform the following step for Exadata:

- Add the channels listed in the Oracle Exadata patch README, such as `exadata_dbserver_12.1.2.3.2_x86_64_base`.

The Oracle Exadata channel to subscribe to depends on the Oracle Exadata software release you want to update to. See the patch README for the release for details.

The YUM mirror creation procedure mentioned in the link above provides a script (`uln-yum-mirror`) which you can use to populate the local mirror. Running the script as root starts the rpm download from ULN to a local directory holding the repository.

You need to install a Web server and configure its "document root" directory to point to this repository location on disk so that the Oracle Exadata channel content is made available to the Oracle Exadata database servers using the HTTP Web server. This is described in step 7 from the mentioned instructions.

When you run the update utility, you specify the URL to the repository location. This URL should always be at the level of the "repdata" directory. You can run a prerequisite check (-



precheck) to verify that the repository works. See the prerequisite check example using YUM repository in "Running Prerequisite Checks."

**Note:**

Exadata database servers must be running Exadata release 12.1.2.2.0 or later to access YUM repositories listening on IPv6 addresses only.

### 8.6.9.3 Downloading the ISO Image as a YUM Repository from My Oracle Support and Making It Available as a YUM HTTP Repository on a Web Server

You can download, uncompress, and place the ISO image channel holding the Oracle Exadata software update on a web server that has HTTP connectivity to every Oracle Exadata database server. This method is recommended for the following conditions:

- There is a large number of database servers to update.
- The single repository server is accessible from all the database servers.

**Note:**

If you run the update utility on a non-Exadata system against this ISO image, this is taken care of locally on each database server. In such cases there is no need to set up a local mirror.

- The database servers do not have customized software that requires updates from additional ULN channels.

**Note:**

If the database servers have custom software installed it is better to use an HTTP YUM repository and not the ISOs, because this improves flexibility in adding back/installing rpms.

Use the following procedure to make the ISO image of the Oracle Exadata channel content available on a Linux-based server running an HTTP server. You can adapt the instructions for other operating systems. In this procedure, the ISO image is copied to the `/u01/app/oracle/stage` directory. The "Document Root" entry in the web server document is `/var/www/html`. The procedure uses release 12.1.1.1.0 as an example.

**Note:**

Oracle recommends that a separate server (non Exadata database server) be used as the web server.

1. Create the ISO image mount point as the root user using the following commands:

```
[root@nonExadataHost]# mkdir -p /var/www/html/yum/unknown/EXADATA/dbserver/12.1.1.1.0/base
```

```
[root@nonExadataHost]# mkdir -p /u01/app/oracle/stage
```

2. Copy the compressed ISO image to the web server into `/u01/app/oracle/stage`.
3. Uncompress and mount the ISO image as the root user using the following commands:

```
[root@nonExadataHost]# unzip p17997668_121110_Linux-x86-64.zip
```

```
[root@nonExadataHost]# mount -o loop /u01/app/oracle/stage /121110_base_repo.iso /var/www/html/yum/unknown/EXADATA/dbserver/12.1.1.1.0/base
```

4. Start the `httpd` service as the root user. This assumes `httpd` is installed.

```
[root@nonExadataHost]# service httpd start
```

5. Identify and test the YUM HTTP repository by connecting to it using a web browser. The following is an example of the repository URL.

```
http://yum-repo/yum/unknown/EXADATA/dbserver/12.1.1.1.0/base/x86_64/
```

6. Run the prerequisite check as the root user to verify that the repository is set up properly. See [Running Prerequisite Checks](#).

## 8.6.10 Managing Exadata Obsolete Packages

If you are updating to release 11.2.3.3.0 or later, some packages on the Exadata database server become obsolete. While updating an Exadata database server, the update utility prints the `exclude rpm list` and `obsolete rpm list` in the log file.

The following example shows the exclusion and obsolete lists from the log file. In this example, an exclusion list has not yet been created by the user.

```
RPM exclusion list : Not in use (add rpms to /etc/exadata/yum/exclusion.lst
and restart dbnodeupdate.sh)
RPM obsolete list : /etc/exadata/yum/obsolete.lst (lists rpms to be
removed by the update)
 : RPM obsolete list is extracted from exadata-sun-
computenode-11.2.3.3.0.131014.1-1.x86_64.rpm
```

To see which packages will become obsolete, review the contents of the `obsolete.lst` file. This file lists the packages defined to be obsolete by Exadata; these packages will be removed during the update when no action is taken. Packages manually added to this list are ignored. The following is a small sample of the `obsolete.lst` file:

```
[root@dm01]# cat /etc/exadata/yum/obsolete.lst
Generated by dbnodeupdate.sh runid: 021213024645
at.x86_64
java-*-openjdk
rhino.noarch
jline.noarch
jpackage-utils.noarch
```

To prevent a package listed in the `obsolete.lst` file from being removed, create the `/etc/exadata/yum/exclusion.lst` file, and put in the rpm name (wildcards are allowed) for

the packages you want to keep. Place the `/etc/exadata/yum/exclusion.lst` file on all Exadata database servers where you want to use it.

The following example shows a package added to the exclusion list:

```
[root@dm01]# cat /etc/exadata/yum/exclusion.lst
java-*-openjdk
```

After you have added an entry to the `exclusion.lst` file and re-run the update utility, the utility detects the exclusion list. The rpm packages on the exclusion list are still shown in the `obsolete.lst` file, but the listed packages in the `exclusion.lst` file will not be removed during the update.

## 8.6.11 Updating Individual Packages

Due to security findings or customization, you may need to update individual (general purpose) Linux packages supplied by the Exadata release. You can do this by removing the `exadata-sun-computenode-exact` rpm first.

Removing this rpm does not impact any functionality, but it removes a “lock”. By removing this “lock”, you are allowed to update specific individual Linux rpms.

If needed, you can remove the `exadata-sun-computenode-exact` rpm as follows:

```
[root@dm01]# rpm -e exadata-sun-computenode-exact
```

When updating to a new release, the update utility tries to “restore” the “lock” (`exadata-sun-computenode-exact`). For example, this is possible if newer packages are shipped in the newer release.

If the `exadata-sun-computenode-exact` rpm cannot be restored, the update utility falls back to `exadata-sun-computenode-minimum`.



### Note:

Do not force in packages using the `rpm -Uvh --nodeps` command, unless directed by Oracle Support Services.

## 8.6.12 Running Prerequisite Checks

You should always run prerequisite checks before doing the actual update. The prerequisite checks do not require downtime and execute important validations such as:

- Validation of the Exadata release (minimum is 11.2.2.4.2 running Oracle Linux 5.5)
- Validation of user input
- Validation of the installation media (YUM repository, either ISO or HTTP)
- Validation of disk space and snapshots
- Validation of YUM settings that are important for the update to finish successfully
- Known issues / best practices

The most important validation executed by the update utility is the YUM dependency check. The YUM dependency check is a YUM update dry-run command (introduced in 11.2.3.3.0) that does not do the actual YUM update but does validate dependencies. This is a final test in determining whether or not the update can proceed. It is often due to customizations that prevent successful updates. For example, installation of additional RPMs might require dependent packages that are not in the YUM repository. If this happens, you need to take corrective action to resolve the conflict.

The YUM dependency check (dry-run) is validated against minimum and exact dependencies. These dependencies are enforced by non-functional Exadata RPMs and help administrators stay exactly at (or close to) the original Exadata release when customizing the system. The update utility uses the `exadata-sun-computenode-exact` and the `exadata-sun-computenode-minimum` RPMs as follows:

- The `exadata-sun-computenode-exact` rpm ensures that only a specific release of Oracle Exadata branded packages is allowed during the update. (release = x)
- The `exadata-sun-computenode-minimum` rpm ensures that a specific or later release of Oracle Exadata branded packages is allowed during the update. (release >= x)

With `exadata-sun-computenode-exact` rpm, the system appears as if it were freshly imaged to the newer release because all the Oracle Exadata packages are exactly the same as on a vanilla installation. The `exadata-sun-computenode-minimum` rpm, however, sets the minimum dependencies, and enforces that all packages are installed, but it also allows packages to be at a later version. A vanilla installation always starts with both RPMs. To allow customization or updates, you need to remove `exadata-sun-computenode-exact`.

By default, the update utility attempts to match the exact dependencies when updating to a later Exadata release. When exact dependencies conflict and cannot be enforced, the utility falls back and attempts to apply the `exadata-sun-computenode-minimum` rpm to enforce minimum dependencies. In such cases the `exadata-sun-computenode-exact` rpm is not installed.

Missing or not updating with exact dependencies is allowed and not a problem. If a system needs to be updated to the exact dependencies, then the conflict needs to be resolved first. Check the log file to see what packages conflict, remove them cautiously, and then re-run the update utility in prerequisite check mode.

If the prerequisite check fails due to dependency issues, you can view the errors on-screen. The update utility's log file has more details and shows which dependencies failed. When both exact and minimum dependencies do not match, the update cannot proceed.

For such cases, check the log file to determine what caused the dependencies to fail. After removing the failed dependencies, re-run the update utility to ensure that at least the minimum dependencies can be enforced.

When dependency errors occur during the prerequisite check or before the update starts, do the following to resolve the problem:

- Analyze the YUM errors in the log file. Search for `Error`.
- Depending on the issue, you may need to de-install, install, or update the rpm packages causing the dependency issue or conflict. The log file lists the failed dependencies.

After the update, you may re-install custom rpm packages that you de-installed, assuming you still require the packages and the packages are compatible with the updated system.

By default, prerequisite check warnings such as active NFS mounts result in a failed check. If you want to allow active NFS mounts, then starting in release 12.1.2.1.1 you can use the `-allow_active_network_mounts` flag.

For more options, see the update utility built-in help.

### Prerequisite Check Examples

The following command shows an example of a prerequisite check using ISO. The command is run as root.

```
[root@dm01]# ./patchmgr -dbnodes dbs_group -precheck -
iso_repo /u01/exa/p22750145_121230_Linux-x86-64.zip
-target_version 12.1.2.3.0.160207.3
```

`-dbnodes` specifies the list of database nodes to be updated.

`-precheck` specifies the prerequisite check action.

`-iso_repo` specifies the location of the ISO (YUM) repository on the node running the updating utility. This can be replaced by `-yum_repo <location>`.

`-target_version` specifies the target release the database servers are being updated to.

### Related Topics

- [Troubleshooting Exadata Database Server Updates](#)  
You can use the log files generated by the update utility to troubleshoot updates.
- [Update Utility for Exadata Database Servers](#)  
Patchmgr is the update utility for updating Oracle Exadata database servers.

## 8.6.13 Backing up Exadata Database Servers Before Planned Maintenance

Oracle recommends making a backup of the Exadata database servers before making any changes when updating to a next release.

This means running the update utility in backup-only mode before making any other (manual) changes to make the prerequisite / dependency check pass. The backup-only action backs up the active root and `/boot` file system only and is sufficient for rolling back (failed) updates.

### Note:

The update utility's built-in backup (`dbserver_backup.sh`) is insufficient for recovering from non-booting system failures. It is recommend having additional (validated) backup/restore procedures in place to recover the entire database server from failures such as double disk failures.

For regular and virtualized Exadata database servers (domU), when the active system image is running from a file system on `/dev/mapper/VGExaDb-LVDbSys1`, the backup is

made onto `/dev/mapper/VGExaDb-LVDbSys2` (and vice versa). For Exadata database servers running dom0 with `/dev/mapper/VGExaDb-LVDbSys2` as the active image, the backup goes to `/dev/mapper/VGExaDb-LVDbSys3` (and vice versa).

This means that upon rollback of an update the active system image for regular database servers and domU will become `/dev/mapper/VGExaDb-LVDbSys2`, and for dom0 will become `/dev/mapper/VGExaDb-LVDbSys3`.

The process of backing up an active system partition requires the default LVM scheme for the Sys\* LVM's and both LVM's to be the same size. This is because it is impossible to back up a larger LVM partition to a smaller LVM partition. Resizing LVM `VGExaDb-LVDbSys1` partition is allowed as long as `VGExaDb-LVDbSys2` is resized to the same size.

While the backup is running, an LVM snapshot takes care of a consistent view of the file system. This LVM snapshot is maintained by the backup script and will always claim 1G of free VG space in `VGExaDb`.

Space for the snapshot is guaranteed by Oracle by using a placeholder LVM called `/dev/VGExaDb/LVDoNotRemoveOrUse`. When the backup script runs, it removes this placeholder, making sure there is always 1G of free space available for a snapshot. After the backup is completed, the snapshot is removed and the `/dev/VGExaDb/LVDoNotRemoveOrUse` LVM is re-created.

You can perform the backup while all system services are up and running. Starting with release 12.1.2.1.1, the utility supports active network mounts (NFS) with the `-allow_active_network_mounts` flag.

The time it takes to run the backup depends on how busy the system is and on the size and type of data that is backed up. For example, backing up millions of small files can take significantly longer than backing up a couple of larger files. For this reason it is recommended to make sure directories holding database `.aud` files are not found on the root file system. Note the following:

- You can have only one backup. Running a new backup means overwriting the existing backup.
- Re-running the update utility with the `-backup` flag (or in default updating mode) will overwrite existing backups.
- All files in `/boot` and found on the active Sys LVM's are backed up. Files in, for example, `/u01`, are not backed up.

### Examples for “Backup Only”

The following example shows running “backup only” as root:

```
[root@dm01]# ./patchmgr -dbnodes dbs_group -backup -iso_repo /u01/exa/
p22750145_121230_Linux-x86-64.zip
-target_version 12.1.2.3.0.160207.3 -allow_active_network_mounts
```

- `-dbnodes` specifies the list of database nodes to be updated.
  - `-backup` specifies the “backup only” action.
  - `-iso_repo` specifies the location of the ISO (YUM) repository on the node running the update utility. Alternatively, you can use the `-yum_repo` flag, which specifies the HTTP location of the YUM repository.
- `-target_version` specifies the target release for which the backup is run.

- `-allow_active_network_mounts` ensures active network mounts remain active while performing the backup.

The following example shows a “backup only” action run as a non-root user:

```
[oracle@nonExadataHost]$./patchmgr -dbnodes ~/dbs_group_scab -backup
-iso_repo /u01/iso/p23557378_121223_Linux-x86-64.zip
-target_version 12.1.2.2.3.160720 -allow_active_network_mounts -
smtp_from "sender@somedomain.com"
-smtp_to "recipient@example.com"
```

### Related Topics

- [Recovering Oracle Linux Database Server with Uncustomized Partitions](#)

## 8.6.14 Running the Update

You can perform the actual update of Exadata database servers in a rolling (using the `-rolling` flag) or non-rolling fashion. The default is non-rolling.

You can also perform the update as root or as a non-root user (using the `-log_dir` flag), as described in "[Running the Update Utility as a Non-root User and Running Multiple Invocations Concurrently](#)".

The update proceeds only if the “minimum dependency check” succeeds. You may need to remove customizations for the update to proceed.

By default the update creates a backup on the inactive system image. If you have already taken a backup and have not changed the image before running the prerequisite check, you can omit the `-nobackup` flag when performing the update.



### Note:

Use the `-nobackup` flag only if a backup was already made before running the prerequisite check and the image has not changed.

The update action requires the following mandatory flags:

- `-upgrade` to specify the update action
- `-iso_repo` (for ISO image) or `-yum_repo` (for HTTP locations) to point to the YUM repository (see [Example 8-13](#))
- `-target_version` to specify the release you want to update to. The patch README always has this information.

You can specify additional flags to allow active remote network mounts during backup and updating (`-allow_active_network_mounts`) and specify mail recipients for updating status notification (`-smtp_from "addr" and -smtp_to "addr1 addr2 addr3 ..."`)

**Example 8-12 Running Update Using ISO Image for YUM Repository**

The following example shows an update action run as root and using an ISO image for the YUM repository. Active network mounts are allowed, and mail information is specified for status notification:

```
[root@dm01]# ./patchmgr -dbnodes ~/dbs_group -upgrade -iso_repo /u01/iso/
p23557378_121223_Linux-x86-64.zip
-target_version 12.1.2.2.3.160720 -allow_active_network_mounts -smtp_from
"sender@somedomain.com"
-smtp_to "receiver@somedomain.com" -nobackup
```

The following example shows an update action run as a non-root user from a remote host using an ISO YUM repository. Active network mounts are allowed, and mail information is specified for status notification:

```
[oracle@nonExadataHost]$./patchmgr -dbnodes ~/dbs_group -upgrade -
iso_repo /u01/iso/p23557378_121223_Linux-x86-64.zip
-target_version 12.1.2.2.3.160720 -allow_active_network_mounts -log_dir auto
-smtp_from "sender@somedomain.com"
-smtp_to "receiver@somedomain.com" -nobackup
```

**Example 8-13 Running Update Using HTTP Location for YUM Repository**

The following example shows an update action run as root using HTTP for the YUM repository. Active network mounts are allowed, and mail information is specified for status notification:

```
[root@dm01]$./patchmgr -dbnodes ~/dbs_group -upgrade -yum_repo http://yum-
repo/yum/ol6/EXADATA/dbserver/12.1.2.2.3/base/x86_64/
-target_version 12.1.2.2.3.160720 -allow_active_network_mounts -smtp_from
"sender@somedomain.com" -smtp_to "receiver@somedomain.com"
-nobackup
```

The following example shows an update action run as a non-root user from a remote host using HTTP for the YUM repository. Active network mounts are allowed, and mail information is specified for status notification:

```
[oracle@nonExadataHost]$./patchmgr -dbnodes ~/dbs_group -upgrade -yum_repo
http://yum-repo/yum/ol6/EXADATA/dbserver/12.1.2.2.3/base/x86_64/
-target_version 12.1.2.2.3.160720 -allow_active_network_mounts -log_dir auto
-smtp_from "sender@somedomain.com"
-smtp_to "receiver@somedomain.com" -nobackup
```

**Related Topics**

- [Running Prerequisite Checks](#)
- [Installing, Updating, and Managing Non-Oracle Software](#)
- [Backing up Exadata Database Servers Before Planned Maintenance](#)



## 8.6.15 Rolling Back Updates

A backup enables you to roll back updates, regardless of whether the update failed or succeeded. This backup is stored on the inactive system partition, as described in "[Backing up Exadata Database Servers Before Planned Maintenance](#)".

When rolling back an update, the update utility performs the following actions:

- Shuts down the stack and domU's.
- Deactivates the active system partition, and activates the inactive system partition.
- Restores `/boot` from the inactive partition.
- Updates the grub boot-loader.

Having only one inactive system partition limits the rollback options to only the previous active image.

### Note:

- For systems being updated to Oracle Linux 6, a backup must be performed before proceeding with the update. The backup is automatic when updating LVM-enabled systems from Oracle Linux 5 to Oracle Linux 6.
- Database servers running as Oracle VM Server (dom0) switch between `LVDbSys2` and `LVDbSys3` as the active system partition when rolling back.
- Database servers running as Oracle VM (domU) have smaller sizes for `LVDbSys1` compared to physical hardware deployments.

### Example 8-14 Rolling back an update using `patchmgr`

```
[root@dm01]# ./patchmgr -dbnodes dbs_group -rollback
```

`-dbnodes` specifies the list of database nodes to be updated.

`-rollback` specifies the rollback action.

For more options see the update utility's built-in help.

 **Note:**

Firmware updates are not rolled back when rolling back to a previous image. The Oracle Exadata System Software releases support later firmware releases. After rolling back, run the following commands to apply older firmware versions when needed:

```
/etc/init.d/lssdiag stop

/etc/init.d/lssi_mrdsnmpd stop

/opt/oracle.cellos/CheckHWnFWProfile -action updatefw -mode exact
```

The last command only applies to releases 11.2.3.3.0 or later.

## 8.7 Updating Database Servers Running Oracle Exadata System Software Release 11.2.2.4.2

Updates for database servers running Oracle Exadata System Software release 11.2.2.4.2 include an update that prepares the servers to use yum.

The steps to prepare the servers are done using the `dbnodeupdate.sh` utility. For this reason, updates from Oracle Exadata System Software release 11.2.2.4.2 require two runs of the `dbnodeupdate.sh` utility with different arguments. The utility provides instructions about which command to run and when to it.

When updating the database servers, the following assumptions are made:

- The Oracle Exadata System Software release is 11.2.2.4.2.
- The database servers are running Oracle Linux 5.5 or later.
- The database server update is available as a local ULN mirror or as a local ISO image.
- [Preparing to Use the dbnodeupdate.sh Utility on Database Servers with Release 11.2.2.4.2](#)  
This procedure describes how to download and prepare the utility on the server.
- [Running the dbnodeupdate.sh Utility](#)  
You can use the `dbnodeupdate.sh` utility on database servers running release 11.2.2.4.2.

### 8.7.1 Preparing to Use the dbnodeupdate.sh Utility on Database Servers with Release 11.2.2.4.2

This procedure describes how to download and prepare the utility on the server.

The `dbnodeupdate.sh` utility is available from My Oracle Support from Doc ID 1553103.1. The `dbnodeupdate.sh` utility can update releases running 11.2.2.4.2 directly to Oracle Linux 6.

1. Download the `dbnodeupdate.sh` utility from My Oracle Support.

2. Log in as the `root` user on the database server.
3. Put the compressed file in the `/u01/dbnodeupdate` directory on the database server. Create this directory if it does not exist.
4. Uncompress the `p16486998_12xxxx_Linux-x86-64.zip` package file in the `/u01/dbnodeupdate` directory using the following command:

```
[root@dm01]# unzip p16486998_12xxxx_Linux-x86-64.zip
```

 **Note:**

- When using a compressed ISO, upload it to the database server and place it in the `/u01/dbnodeupdate` directory, provided there is free space available. Create this directory if it does not exist.
- When using a local ULN mirror, make sure the HTTP location is available.
- When making ISO contents available on a web server, upload the contents to the web server, mount it, and verify the URL.
- A user with `sudo` privileges can use `sudo` to run the `dbnodeupdate.sh` utility.

**Related Topics**

- [dbnodeupdate.sh and dbserver.patch.zip: Updating Exadata Database Server Software using the DBNodeUpdate Utility and patchmgr \(My Oracle Support Doc ID 1553103.1\)](#)
- [Preparing and Populating the YUM Repository with the Oracle Exadata Channel Content](#)

## 8.7.2 Running the `dbnodeupdate.sh` Utility

You can use the `dbnodeupdate.sh` utility on database servers running release 11.2.2.4.2.

1. Log in as the `root` user on the database server.
2. Run the `dbnodeupdate.sh` utility using the following command, where `repo` is the HTTP location of the ULN mirror or the location of the compressed ISO image:

```
[root@dm01]# ./dbnodeupdate.sh -u -l repo
```

The database server restarts automatically.

 **Note:**

Just before the restart, instructions for the next step are provided onscreen.

3. Run the following command after the restart mentioned in step 2. The utility automatically remounts the ISO image when a compressed ISO image is used.

```
[root@dm01]# ./dbnodeupdate.sh -u -p 2
```

4. Run the following command to complete the update.

```
[root@dm01]# ./dbnodeupdate.sh -c
```

During the completion process, the utility performs post-update checks, applies best practices, relinks the Oracle homes, and starts the stack.

 **Note:**

The update process might still be running when the `dbnodeupdate.sh -c` command is entered. When this happens, the utility waits until it can determine the image status. The system may restart while waiting to make pending updates effective. If this happens, then re-enter the `dbnodeupdate.sh -c` command when the system is back online.

## 8.8 Updating Software on Oracle Exadata Storage Servers

Use the following information and procedures when updating storage servers within Exadata Database Machine.

- [Overview of Exadata Database Machine Storage Server Updates](#)  
When updating storage servers, there is more than one type of software that needs to be updated and different methods of performing the updates.
- [Scheduling Automated Updates of Storage Servers](#)  
Starting with Oracle Exadata System Software release 18.1.0.0.0, you can schedule software updates for the storage servers.
- [Update Utility for Oracle Exadata Storage Server](#)
- [Recommended Timeline for Updating Oracle Exadata Storage Server](#)
- [Preparing Exadata Storage Servers for Update](#)  
Perform these preparation steps before updating the Exadata storage servers.
- [Running the Update for Exadata Storage Servers](#)
- [Rolling Back Updates for Exadata Storage Servers](#)

### 8.8.1 Overview of Exadata Database Machine Storage Server Updates

When updating storage servers, there is more than one type of software that needs to be updated and different methods of performing the updates.

Oracle Exadata System Software release updates contain updates for the following components within an Exadata Database Machine storage server:

- Oracle Linux operating system
- Firmware (Flash, Disk, RAID controller, Integrated Lights Out Manager (ILOM), HCA)

- Oracle Exadata System Software

What software and firmware that is updated depends on the current Oracle Exadata System Software release the storage server is on and the release it is updated to. Oracle Linux operating system packages and Oracle Exadata System Software are always updated, while firmware updates are applied only on a small selection of the components, or not at all.

Updates for Exadata Database Machine storage server can be applied independently from the updates to Exadata Database Machine database server or RDMA Network Fabric switches unless specified otherwise. It is not mandatory to apply each and every Oracle Exadata System Software update that comes out. For example, you can skip two or three releases and update directly to a newer release.

Updating the Oracle Exadata System Software is always performed “out of place”. This means that a new version of the operating system including the Oracle Exadata System Software is installed on the inactive system partition. The utility to update the Oracle Exadata System Software ships with the update itself.

If you cannot afford cluster-wide downtime, you can update the storage servers in a rolling fashion. Rolling means updating one storage server at a time. If you can afford the cluster-wide downtime, you can update all storage servers in parallel. Non-rolling updates reduce the overall time required.

Starting with Oracle Exadata System Software release 18c (18.1.0), there is a more scalable alternative to using `patchmgr` for software updates. The storage servers automatically verify preconditions and download the update software from a URL that you specify. Each storage server downloads the software to its active partition, and then loads the software on its passive partition. At a specified time, the storage servers reboot to the new version. The storage servers use the Oracle Automatic Storage Management (Oracle ASM) disk deactivation status to determine when it is safe to deactivate the disks and reboot the storage server to the new software version. These scheduled updates invoke the same scripts that are currently used by the `patchmgr` process.

## 8.8.2 Scheduling Automated Updates of Storage Servers

Starting with Oracle Exadata System Software release 18.1.0.0.0, you can schedule software updates for the storage servers.

Perform the following steps from an external server, for example, an Oracle Exadata Database Server.

If you are accessing the Software Update store using the HTTPS protocol, then TLS certificate checks are required by default. If the certificate for the web server that hosts the software update cannot be validated, then the following error is returned:

```
CELL-00076: An error occurred during download of software update:
source https://example.com:port is not available.
CELL-00092: The store's TLS certificate cannot be authenticated with
known CA certificates
```

1. Copy the software update ZIP file to a directory which is hosted by a web server.
2. If you are using Oracle Exadata System Software 18c (18.1.0) or 18c (18.1.1), then the patch file must have a name like `18.1.1.0.0.171018.patch.zip`.

If the downloaded patch has a name like `p26875767_181100_Linux-x86-64.zip`, then rename the file to `18.1.1.0.0.17018.patch.zip`. Rename the ZIP file to use the release number and date string that is used for the directory name within the ZIP file. For example, when you unzip the patch `p26875767_181100_Linux-x86-64.zip` it extracts the directory `patch_18.1.1.0.0.171018`.

If you are using Oracle Exadata System Software release 18.1.2 or higher, then the patch downloaded from My Oracle Support is automatically renamed for you during the validation step.

3. Determine approximate time to perform the software update of the storage servers.
4. Create a file on the external server that contains the list of cells to be updated. Name this file `cells`.
5. Use `dcli` to schedule the update of the cells.
  - a. Setup passwordless access, if needed.

```
$ dcli -g cells -k
```

- b. Specify the location of the software update ZIP file to use during the update.

```
$ dcli -g cells cellcli -e 'alter softwareUpdate store="https://host/
exa-updates/cell"'
```

- c. Specify the time to start the update of the Oracle Exadata System Software on the storage servers.

If you specify the time before providing the software store location, then the software update download might start before the proper store location has been set.

```
$ dcli -g cells cellcli -e 'alter softwareUpdate time="1 AM
Thursday"'
```

6. Wait for the updates to occur.

Management Server (MS) will start file download and run pre-checks up to one week before the scheduled update. MS will generate an alert if any cell does not update as scheduled.

#### Related Topics

- ALTER SOFTWAREUPDATE

## 8.8.3 Update Utility for Oracle Exadata Storage Server

You use the `patchmgr` update utility for updating Oracle Exadata storage servers. For Exadata storage server updates the utility is packaged (and shipped) with the update itself and is available for download from My Oracle Support as storage server update.

#### Note:

The `patchmgr` used for updating Exadata storage servers is not the same as the `patchmgr` used for applying the Exadata database server software update.

Whether or not the utility supports the Exadata hardware you have depends on the Exadata storage server release you are trying to update to. The utility orchestrates the update process across the specified Exadata storage servers. The utility allows running the update in a rolling or non-rolling fashion. You can run the update utility from Exadata database servers or from other servers running Oracle Linux or Oracle Solaris.

The update utility performs the following tasks:

- Automates the preparation, update, and validation steps
- Automates rollbacks

The update utility supports multiple sessions: you can run multiple updates concurrently from the same server starting with release 12.1.2.3.2 for Exadata storage servers and starting with release 11.2.3.1.0 for Exadata database servers. This means multiple racks can be updated concurrently from the same server. The update utility can be run as root or as a non-root user. By default the update utility assumes it should run as the root user. If however you want to enable multiple session support or run as a non-root user, then you need to use the `-log_dir` flag. The `-log_dir` flag supports two types of arguments: either a location on disk or the keyword `auto`. If you specify `auto`, the update utility creates its own log directory based on the storage servers listed in the `cell_group` file. This behavior causes the update utility to create new directories for each run of updates in the same cluster where one or more clusters were added or removed from the `cell_group` file. In order to obtain (and reuse) such a directory, the update utility provides the `-get` flag to determine the log directory for your session. The `-get` flag scans the working directory for directories in the log directory and returns the directory for your `cell_group`. For example, the following command:

```
[oracle@nonExadataHost]# ./patchmgr -dbnodes ~/cell_group -log_dir
auto -get log_dir
```

The previous comment might return output similar to the following:

```
log_dir=/u01/test/patch_12.1.2.4.0.160802/log/
dbm02celadm01_dbm02celadm02_9cfbc690
```

In a subsequent update session, you can re-use the log directory location:

```
[oracle@nonExadataHost]# ./patchmgr -cells ~/cell_group -
patch_check_prereq -log_dir /u01/test/patch_12.1.2.4.0.160802/log
/dbm02celadm01_dbm02celadm02_9cfbc690
```

## 8.8.4 Recommended Timeline for Updating Oracle Exadata Storage Server

### Note:

It is highly recommended to validate Oracle Exadata System Software updates on test systems before doing them on production systems.

By following the approach below, you allow yourself time to perform any necessary corrective actions.

**Table 8-10 Timeline for Performing Updates**

When	Tasks
Weeks to days before the update	<ul style="list-style-type: none"> <li>Download the Oracle Exadata System Software update you require from My Oracle Support note <a href="#">888828.1</a></li> <li>Download the latest Oracle EXAchk from My Oracle Support note <a href="#">1070954.1</a></li> <li>Review the release-specific My Oracle Support note for known issues</li> <li>Run Oracle EXAchk</li> <li>Perform prerequisite check. See <a href="#">Preparing Exadata Storage Servers for Update</a> for details.</li> <li>Correct findings that need attention and re-run the above steps as needed.</li> </ul>
At update time	<ul style="list-style-type: none"> <li>Download the Oracle Exadata System Software update from My Oracle Support note <a href="#">888828.1</a></li> <li>Download the latest Oracle EXAchk from My Oracle Support note <a href="#">1070954.1</a></li> <li>Review the release-specific My Oracle Support note for known issues</li> <li>Run Oracle EXAchk and perform corrective actions as needed</li> <li>Perform prerequisite check and corrective actions as needed.</li> <li>Perform the update. See <a href="#">Running the Update for Exadata Storage Servers</a> for details.</li> </ul>
After the update	<ul style="list-style-type: none"> <li>Run Oracle EXAchk.</li> </ul>

## 8.8.5 Preparing Exadata Storage Servers for Update

Perform these preparation steps before updating the Exadata storage servers.

You can perform the update (and rollback) action in a rolling or non-rolling method. You can also perform a prerequisite check in a rolling or non-rolling method. The default is non-rolling.

1. Set up SSH equivalence for the user that is driving the update utility.
2. Download and run Oracle ExaCHK. Review and address any open issues. See My Oracle Support note [1070954.1](#).
3. Review the release-specific My Oracle Support note for any known issues and workarounds.
4. Check prerequisites for your method of update or rollback.
 

**Prerequisites for performing a rolling update:**

  - a. Verify that your Grid Infrastructure home and Database home software versions and patch levels meet the minimum required for Exadata storage server rolling cell update as documented in My Oracle Support note [888828.1](#)



- b. Verify `failgroup_repair_time` or `disk_repair_time` for each Oracle ASM disk group.

When applying the update in a rolling manner, the update utility updates one server at a time, first taking all grid disks and ASM disks offline, then applying the update to the server, then bringing all ASM disks and grid disks back online. The Oracle ASM repair timeout attributes, `disk_repair_time` and `failgroup_repair_time`, need to be set to a value large enough to allow a single storage server update to complete. The default values of 3.6h and 24h, respectively, are the recommended values. Note that during a rolling storage server update disk groups with `compatible.asm >= 12.1.0.2.0` will use the value of `failgroup_repair_time`, and disk groups with `compatible.asm < 12.1.0.2.0` will use the value of `disk_repair_time`.

Verify ASM repair timeout attributes are set to default or higher values. Use the following command to check repair time attributes for all mounted disk groups in the Oracle ASM instance.

```
SQL> col attribute format a30
SQL> col value format a10
SQL> select dg.name as diskgroup, a.name as attribute, a.value
 from v$asm_diskgroup dg, v$asm_attribute a
 where dg.group_number=a.group_number
 and (a.name like '%repair_time' or a.name =
 'compatible.asm');
```

DISKGROUP	ATTRIBUTE
VALUE	
DATA 3.6h	disk_repair_time
DATA 24.0h	failgroup_repair_time
DATA 12.1.0.2.0	compatible.asm
RECO 3.6h	disk_repair_time
RECO 24.0h	failgroup_repair_time
RECO 12.1.0.2.0	compatible.asm

If the Oracle ASM repair timer for any disk group is lower than the default value, then set the repair timer to the default value for the duration of the update. You may set it back to its current value after the update successfully finishes for all storage servers.

**Prerequisites for performing a non-rolling update:**

- a. Shut down and stop the Oracle components on each Exadata database server using the following commands, where *Grid\_home* is the directory where the Oracle Grid Infrastructure software is installed:

```
[root@dm01]# dcli -g dbs_group -l root "Grid_home/bin/crsctl stop crs"
```

If Oracle Clusterware was not stopped using the preceding command, then use the following command to force it to stop:

```
[root@dm01]# crsctl stop crs -f
```

- b. Use the following command to check Oracle Clusterware status, where *Grid\_home* is the directory where the Oracle Grid Infrastructure software is installed:

```
[root@dm01]# dcli -g dbs_group -l root "Grid_home/bin/crsctl check crs"
```

All Oracle Clusterware components must be offline. If you are performing a non-rolling update in a configuration running Oracle VM, then you must check the Oracle Clusterware state in all VM clusters.

5. Check if your system meets the following conditions:
  - Oracle Exadata Database Machine hardware uses Sun servers.
  - The installed Oracle Exadata Software release is earlier than release 11.2.2.2.0.
  - The installed Linux `ofa` package is earlier than 1.5.1-4.0.28.

If the system meets all of the preceding conditions, then Exadata storage servers running Oracle Linux may encounter a file system corruption that results in the root file system mounted as read-only after reboot. Follow the instructions in My Oracle Support note 1589868.1 before updating the storage servers and database servers.

6. Unzip the update. It will extract into the `patch_release.date_code` directory. Change to this patch directory.
7. Download any patchmgr plug-ins attached to the My Oracle Support note for your target release and install them as documented in the My Oracle Support note. Oracle recommends reviewing the My Oracle Support notes, issues, and workarounds listed in the release note just before starting to actually apply the update.
8. Clean up any previous update utility runs using the `-cleanup` flag.

 **Note:**

The first time the storage servers are updated the `-reset_force` flag should be used before running cleanup.

Example using `-reset_force` as the root user:

```
[root@dm01]# ./patchmgr -cells ~/cell_group -reset_force
```

Example using `-reset_force` as a non-root user:

```
[oracle@nonExadataHost]# ./patchmgr -cells ~/cell_group -
log_dir auto -reset_force
```

Example using `-cleanup` as the root user:

```
[root@dm01]# ./patchmgr -cells ~/cell_group -cleanup
```

Example using `-cleanup` as a non-root user:

```
[oracle@nonExadataHost]$./patchmgr -cells ~/cell_group -log_dir
auto -cleanup
```

#### 9. Run prerequisite check.

Example running prerequisite check for a rolling update as the root user from an Exadata database server:

```
[root@dm01]# ./patchmgr -cells ~/cell_group -patch_check_prereq -
rolling -smtp_from "sender@example.com"
-smtp_to receiver@example.com
```

Example running prerequisite check for a non-rolling update as a non-root user from a non-Exadata database server:

```
[oracle@nonExadataHost]$./patchmgr -cells ~/cell_group -log_dir
auto -patch_check_prereq -smtp_from "sender@example.com"
-smtp_to "receiver@example.com"
```

#### Related Topics

- [Overview of Exadata Database Machine Storage Server Updates](#)  
When updating storage servers, there is more than one type of software that needs to be updated and different methods of performing the updates.
- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.
- [Oracle Exadata Database Machine exachk or HealthCheck \(My Oracle Support Doc ID 1070954.1\)](#)

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- *Oracle Automatic Storage Management Administrator's Guide*
- [Procedure to check for corrupted root file system on Exadata Storage Servers and Linux database servers \(My Oracle Support Doc ID 1589868.1\)](#)

## 8.8.6 Running the Update for Exadata Storage Servers

After performing the prerequisite steps in [Preparing Exadata Storage Servers for Update](#), you can perform the actual update step.

Note the following when applying the update to the Exadata storage servers:

- Do not use the serial console or the ILOM web-based console to start the update utility. There is a known issue of a system halt on the serial console when a write is attempted to stderr or stdout. If an update is started from the serial console, then it may halt. You are using the serial console if the output from the following command is serial.

```
[root@dm01]# ./echo $consoletype
```
- When needed, use the ILOM web-based console to monitor the storage server during the update. You will need to use the ILOM web-based console in case troubleshooting is required.
- To obtain ILOM and serial console access for the storage servers, use SSH to the ILOM host name or IP address as the root user. Do the following to start the serial console:

```
start /SP/console
```

To stop it press the Escape key (ESC) followed by (.
  - Start a new login session for each update or rollback procedure. Do not run a rollback procedure from the same login session where an update was applied. Do not run an update from a login session where a rollback procedure was run.
  - Do not interrupt the update process.
  - If you must use a storage server as the patchmgr utility launch system, then do not use `/opt/oracle` as the staging area for the update. This causes the update to fail and corrupt the storage server. Use the `/tmp` directory as the staging area, that is, unzip the files for the update in `/tmp`.
  - Storage servers automatically reboot, as needed, during the update process. Do not reboot or power cycle storage servers while applying updates.
  - Do not edit or open log files in writable mode. You may use any of the following to view a log: `view`, `less`, `more`, or `tail`. You may cause the update process to be interrupted if you edit the log files during the update.
  - At the end of the patchmgr session, the `patchmgr.stdout` log file is divided into individual storage server log files with names in the format of `cell_name.log`. In addition, the `/var/log/cellos` content from the inactive cell partition is copied to the `/var/log/cellos/inactive_partition` directory. To locate the inactive partition, use the following command:

```
[root@dm01]# ./imageinfo -inactive -sys
```

**Examples running update:**

Running the update in a rolling fashion as root from an Exadata database server:

```
[root@dm01]# ./patchmgr -cells ~/cell_group -patch -rolling -smtp_from
"sender@somedomain.com" -smtp_to receiver@somedomain.com
```

Running the update in a non-rolling fashion as a non-root user from a non-Exadata database server:

```
[oracle@nonExadataHost]$./patchmgr -cells ~/cell_group -log_dir auto -patch -
smtp_from "sender@somedomain.com" -smtp_to "receiver@somedomain.com"
```

After the update is done, clean up the storage servers using the `-cleanup` option to clean up all the temporary update or rollback files. This option cleans the stale update and rollback states as well as cleaning up to 1.5 GB of disk space on the storage server. Use this option before retrying a halted or failed run of the update utility. See [step 8](#) for details.

## 8.8.7 Rolling Back Updates for Exadata Storage Servers

You can roll back updated Exadata storage servers only when they are updated successfully. This means the `imageinfo` command must return `success` for active image status. Storage servers with incomplete or failed updates cannot be rolled back. Rollbacks can be done in a rolling or non-rolling fashion.

1. Check the version that the storage servers will be rolled back to and the `flashCacheMode` setting with the following commands:

```
[root@dm01]# dcli -l root -g cell_group imageinfo -ver -inactive

[root@dm01]# dcli -l root -g cell_group cellcli -e 'list cell attributes
flashCacheMode'
```

### Note:

If you need to roll back storage servers to releases earlier than release 11.2.3.2.0 with writeback flash cache enabled, you need to convert the flash cache to writethrough flash cache before performing the rollback action. Disable the writeback flash cache using the script in My Oracle Support note 1500257.1. Storage servers being rolled back to release 11.2.3.2.0 or later retain the flash cache mode that is currently set.

2. Check the prerequisites for rollback using the following command:

```
[root@dm01]# ./patchmgr -cells cell_group -rollback_check_prereq [-rolling]
```

3. Perform the rollback.

Example of a non-rolling rollback run as root:

```
[root@dm01]# ./patchmgr -cells ~/cell_group -rollback
```

Example of a rolling rollback run as a non-root user:

```
[oracle@nonExadataHost]# ./patchmgr -cells ~/cell_group -rollback -rolling -
log_dir auto
```

 **Note:**

Firmware updates are not rolled back when rolling back to a previous image. After rolling back, run the following command to apply older firmware versions when needed:

```
/etc/init.d/lssidiag stop

/etc/init.d/lssi_mrdsnmpd stop

/opt/oracle.cellos/CheckHWnFWProfile -action updatefw -mode exact
```

4. Clean up the Exadata storage servers using the `-cleanup` option to clean up all the temporary update or rollback files. This option cleans the stale update and rollback states as well as cleaning up to 1.5 GB of disk space on the Exadata storage servers. Use this option before retrying a halted or failed run of the `patchmgr` utility.

```
[root@dm01]# ./patchmgr -cells cell_group -cleanup
```

## 8.9 Upgrading and Downgrading RoCE Network Fabric Switch Firmware

This topic describes procedures to upgrade and downgrade the firmware on the RoCE Network Fabric switches.

Note the following when updating the RoCE Network Fabric switch firmware:

- Upgrading and downgrading the RoCE Network Fabric switch firmware is performed using the `patchmgr` utility.
- Starting with Oracle Exadata System Software release 19.3.0, there are separate `patchmgr` distributions for the servers and RDMA Network Fabric switches. Ensure that you use the `patchmgr` distribution included in the patch ZIP file for the RoCE Network Fabric switches.
- Download the appropriate patch ZIP file to any machine with access to the RoCE Network Fabric switches. Refer to My Oracle Support note 888828.1 for the patch information.
- `patchmgr` configures `ssh` passwordless-access to each switch, which requires you to provide the password of the `admin` user for the switch.
- You must use a non-`root` user to perform the patching, and must include the `-log_dir` option with `patchmgr`.
- Switch firmware is always upgraded in a rolling manner (one switch at a time).
- Perform storage server updates separately from RDMA Network Fabric switch updates. Do not update storage servers and RDMA Network Fabric switches concurrently. RDMA Network Fabric network connections must be stable during some critical stages of storage server updates. The RDMA Network Fabric switch firmware upgrade requires a switch reboot, which disrupts some connections on the RDMA Network Fabric.
- Commencing with the August 2022 `patchmgr` release, `patchmgr` performs an additional series of checks on the RoCE Network Fabric. The checks occur immediately before any firmware upgrade or downgrade and also during prerequisite checking using the `--`

`roceswitch-precheck` option. These checks mitigate the risks of failure associated with unexpected problems in the RoCE Network Fabric. For example, if one of the RoCE Network Fabric ports on a storage server is down, the storage server would become unavailable if the switch connected to the only operational port is taken offline for an upgrade. If any check fails, `patchmgr` reports the problem and ends immediately. In this case, you must correct the problem with the RoCE Network Fabric before you can perform the upgrade or downgrade.

Use the following procedures to upgrade and downgrade the firmware on the RoCE Network Fabric switches:

- [Preparing for RoCE Network Fabric Switch Firmware Upgrades or Downgrades](#)  
You must follow a specific order when upgrading the RoCE Network Fabric switches.
- [Upgrading RoCE Network Fabric Switch Firmware Software](#)  
Upgrade the RoCE Network Fabric switches using the `patchmgr` command.
- [Downgrading RoCE Network Fabric Switch Firmware](#)  
Downgrading firmware means applying older firmware.

## 8.9.1 Preparing for RoCE Network Fabric Switch Firmware Upgrades or Downgrades

You must follow a specific order when upgrading the RoCE Network Fabric switches.

1. Log in to a server that has access to the RoCE Network Fabric switches.
2. Download the appropriate patch file to the server.

Starting with Oracle Exadata System Software release 19.3.0, the updates for the switches are in a separate patch. Refer to My Oracle Support note 888828.1 for patch information.

3. Unzip the patch file.

The files are unzipped to the `patch_switch_release` directory.

4. Change to the directory that contains the `patchmgr` utility.

For example:

```
cd patch_switch_19.3.0.0.0.190915
```

5. Create a switch list file to drive the update of the RoCE Network Fabric switches.
  - a. Create a file that contains the host name or IP address of the switches that you want to upgrade. Place each switch on a separate line.

For example, create a file named `switches.lst`, which contains the host name of each switch on separate lines. On a single rack system, the file might contain:

```
switch123-rocea0
switch123-roceb0
```

- b. Tag each line to specify the configuration type for each switch.

To specify the configuration type for each switch, append a colon (:) and tag to each switch host name or IP address in the switch list file. The following tags are supported:

- `leaf` - Identifies a leaf switch in a single rack system. This configuration type is assumed if no tag is specified.
- `mleaf` - Identifies a leaf switch in a multi-rack system.
- `sflleaf` - Identifies a leaf switch in a single rack system that is enabled to support Exadata Secure RDMA Fabric Isolation.
- `msflleaf` - Identifies a leaf switch in a multi-rack system that is enabled to support Exadata Secure RDMA Fabric Isolation.
- `mspine` - Identifies a spine switch. Note that one spine switch configuration supports all spine switches on single and multi-rack systems, with and without Exadata Secure RDMA Fabric Isolation.
- `leaf23` - Identifies a leaf switch in a single rack system that is configured with 23 host ports. This configuration is required only for 8-socket systems (X8M-8 and later) with 3 database servers and 11 storage servers.
- `mleaf23` - Identifies a leaf switch in a multi-rack system that is configured with 23 host ports. This configuration is required only for 8-socket X8M-8 systems with 3 database servers and 11 storage servers.
- `mleaf_u14` - Identifies a leaf switch in a multi-rack system that is configured with 14 inter-switch links. This is the typical multi-rack leaf switch configuration for X9M systems.
- `msflleaf_u14` - Identifies a leaf switch in a multi-rack system that is enabled to support Exadata Secure RDMA Fabric Isolation and is configured with 14 inter-switch links. This configuration is required for X9M systems with Secure Fabric enabled.
- `mleaf23_u13` - Identifies a leaf switch in a multi-rack system that is configured with 23 host ports and 13 inter-switch links. This configuration is required only for 8-socket X9M-8 systems with three database servers and 11 storage servers.

For example:

```
switch123-rocea0:leaf
switch123-roceb0:leaf
```

- c. For multi-rack configurations only, specify a unique loopback octet for each switch. The loopback octet is the last octet of the switch loopback address, which uniquely identifies a switch.

To specify the loopback octet for each switch, append a period (.) and numeric loopback octet value to each tagged switch entry in the switch list file. The range of valid loopback octet values is:

- 101-118 for leaf switches
- 201-208 for spine switches

For example, the switch list file for a 2-rack system might contain:

```
rack1sw-rocea0:mleaf.101
rack1sw-roceb0:mleaf.102
```



```
rack1sw-roces0:mspine.201
rack2sw-rocea0:mleaf.103
rack2sw-roceb0:mleaf.104
rack2sw-roces0:mspine.202
```

6. Run the prerequisite check prior to either upgrading or downgrading the firmware.

```
./patchmgr --roceswitches switches.lst [--upgrade | --downgrade]
--roceswitch-precheck [--force]
[-log_dir {absolute_path_to_log_directory | AUTO}]
```

In the `patchmgr` command:

- `--roceswitch-precheck` instructs `patchmgr` to perform a firmware upgrade or downgrade simulation on the switch.
- `--force` optionally proceeds with the operation even if the switch is already on the target firmware version or the RoCE Network Fabric is experiencing non-critical failures.
- `-log_dir` specifies the absolute path to the log directory, or `AUTO` instructs `patchmgr` to automatically create the log directory. This option is required when running `patchmgr` as a non-root user.

 **Note:**

Prior to Oracle Exadata System Software release 19.3.9, you must run `patchmgr` as a non-root user for patching RoCE Network Fabric switches.

 **Note:**

The current user is expected to have SSH equivalency configured prior to running `patchmgr`. If it is not configured, then `patchmgr` will give you the option to setup keys and key exchange for SSH equivalency.

If the output from the command shows overall status is `SUCCESS`, then proceed with the upgrade. If the output from the command shows overall status is `FAIL`, then review the error summary in the output to determine which checks failed, and then correct the errors. After correcting all the errors, rerun the prerequisite checks until it is successful.

 **Note:**

You may see a `Config validation failed` error because the golden configuration settings have not been applied to the switch. In this case, apply the golden configuration settings from the current patch and then repeat the prerequisite check. See [Applying Golden Configuration Settings on Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches](#).

### Related Topics

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Applying Golden Configuration Settings on Cisco Nexus 9336C-FX2 RoCE Network Fabric Switches](#)  
The golden configuration settings are normally applied on the RoCE Network Fabric switches during initial deployment.

## 8.9.2 Upgrading RoCE Network Fabric Switch Firmware Software

Upgrade the RoCE Network Fabric switches using the `patchmgr` command.

Before starting this task, review the notes in [Upgrading and Downgrading RoCE Network Fabric Switch Firmware](#), and complete the steps in [Preparing for RoCE Network Fabric Switch Firmware Upgrades or Downgrades](#).

1. On the server where you downloaded the patch, use the `patchmgr` command to upgrade the switches.

```
./patchmgr --roceswitches switches.lst --upgrade [--force]
[-log_dir {absolute_path_to_log_directory | AUTO}]
```

In the `patchmgr` command:

- `--force` optionally proceeds with the operation even if the switch is already on the target firmware version or the RoCE Network Fabric is experiencing non-critical failures.
- `-log_dir` specifies the absolute path to the log directory, or `AUTO` instructs `patchmgr` to automatically create the log directory. This option is required when running `patchmgr` as a non-root user.

### Note:

Prior to Oracle Exadata System Software release 19.3.9, you must run `patchmgr` as a non-root user for patching RoCE Network Fabric switches.

2. Verify the upgrade.

Use the `show version` command to verify the firmware version on the switch:

For example:

```
show version
Cisco Nexus Operating System (NX-OS) Software
TAC support: http://www.cisco.com/tac
Copyright (C) 2002-2019, Cisco and/or its affiliates.
All rights reserved.
...

Software
 BIOS: version 05.33
 NXOS: version 7.0(3)I8(1)
 BIOS compile time: 09/08/2018
```

```
NXOS image file is: bootflash:///nxos.7.0.3.I8.1.bin
NXOS compile time: 3/5/2019 13:00:00 [03/05/2019 22:04:55]
```

#### Hardware

```
cisco Nexus9000 C9336C-FX2 Chassis
Intel(R) Xeon(R) CPU D-1526 @ 1.80GHz with 24571632 kB of memory.
Processor Board ID FD023040CS1
```

```
Device name: dbm01sw-rocea0
bootflash: 115805356 kB
Kernel uptime is 17 day(s), 20 hour(s), 50 minute(s), 25 second(s)
```

```
Last reset at 188268 usecs after Mon Aug 12 17:14:40 2019
Reason: Module PowerCycled
System version:
Service: HW check by card-client
```

#### plugin

```
Core Plugin, Ethernet Plugin
```

```
Active Package(s):
```

## 8.9.3 Downgrading RoCE Network Fabric Switch Firmware

Downgrading firmware means applying older firmware.



### Note:

The current storage server release determines what release you can downgrade to. This may be different for each release and may not be the firmware you were on before the upgrade. For more information on the older firmware shipped with the release you are upgrading to, see the patch README file.

Before starting this task, review the notes in [Upgrading and Downgrading RoCE Network Fabric Switch Firmware](#), and complete the steps in [Preparing for RoCE Network Fabric Switch Firmware Upgrades or Downgrades](#).

1. Use the `patchmgr` command to downgrade the firmware on the RoCE Network Fabric switches.

```
./patchmgr --roceswitches switches.lst --downgrade [--force]
[-log_dir {absolute_path_to_log_directory | AUTO}]
```

In the `patchmgr` command:

- `--force` optionally proceeds with the operation even if the switch is already on the target firmware version or the RoCE Network Fabric is experiencing non-critical failures.

- `-log_dir` specifies the absolute path to the log directory, or `AUTO` instructs `patchmgr` to automatically create the log directory. This option is required when running `patchmgr` as a non-root user.

 **Note:**

Prior to Oracle Exadata System Software release 19.3.9, you must run `patchmgr` as a non-root user for patching RoCE Network Fabric switches.

2. Use the `show version` command to verify that the switch firmware is downgraded.

```
show version
Cisco Nexus Operating System (NX-OS) Software
TAC support: http://www.cisco.com/tac
Copyright (C) 2002-2019, Cisco and/or its affiliates.
All rights reserved.
...

Software
 BIOS: version 05.33
 NXOS: version 7.0(3)I7(6)
 BIOS compile time: 09/08/2018
 NXOS image file is: bootflash:///nxos.7.0.3.I7.6.bin
 NXOS compile time: 3/5/2019 13:00:00 [03/05/2019 22:04:55]

Hardware
 cisco Nexus9000 C9336C-FX2 Chassis
 Intel(R) Xeon(R) CPU D-1526 @ 1.80GHz with 24571632 kB of memory.
 Processor Board ID FDO23040CS1

 Device name: dbm01sw-rocea0
 bootflash: 115805356 kB
 Kernel uptime is 17 day(s), 20 hour(s), 50 minute(s), 25 second(s)

 Last reset at 188268 usecs after Mon Aug 12 17:14:40 2019
 Reason: Module PowerCycled
 System version:
 Service: HW check by card-client

plugin
 Core Plugin, Ethernet Plugin

Active Package(s):
```

## 8.10 Updating InfiniBand Network Fabric Switch Firmware

This topic describes procedures to upgrade and downgrade the firmware on the InfiniBand Network Fabric switches.

Note the following when updating the InfiniBand Network Fabric switch firmware:

- The upgrade of the InfiniBand Network Fabric switches is executed using `patchmgr`.

- The minimum switch firmware release that can use the update utility is release 1.3.3-2. Switch firmware is always updated in a rolling manner.
- Download the appropriate patch ZIP file to any machine with access to the InfiniBand Network Fabric switches. Refer to My Oracle Support note 888828.1 for the patch information.
- Switch firmware is always upgraded in a rolling manner.

 **Note:**

Before Oracle Exadata System Software release 19.3.0, the storage server and RDMA Network Fabric switch updates use the version of patchmgr bundled with the Oracle Exadata System Software release. Starting with Oracle Exadata System Software release 19.3.0, there is a separate patchmgr download for the RDMA Network Fabric switches and the storage servers.

Perform storage server updates separately from RDMA Network Fabric switch updates. Do not update storage servers and RDMA Network Fabric switches concurrently. RDMA Network Fabric network connections must be stable during some critical stages of storage server updates. The RDMA Network Fabric switch firmware upgrade requires a switch reboot, which disrupts some connections on the RDMA Network Fabric.

- [Preparing for InfiniBand Network Fabric Switch Firmware Updates](#)  
You must follow a specific order when updating the InfiniBand Network Fabric switches.
- [Updating InfiniBand Network Fabric Switch Firmware Software](#)  
Update the InfiniBand Network Fabric switches using the `patchmgr` command.
- [Downgrading InfiniBand Network Fabric Switch Firmware](#)  
Downgrading firmware means reapplying the older firmware update, which is shipped with the Oracle Exadata System Software update.

## 8.10.1 Preparing for InfiniBand Network Fabric Switch Firmware Updates

You must follow a specific order when updating the InfiniBand Network Fabric switches.

- If a spine switch is present in the rack, it needs to be updated first.
- If a spine switch is not in the rack, then update the switch running the subnet manager first.
- If the subnet manager is not running on the switches, then perform the update in any order.

To update the InfiniBand Network Fabric switches, the switch firmware must be at release 1.3.3-2 or later. If the switch firmware is at an earlier release, then update the firmware to release 1.3.3-2 using the instructions in My Oracle Support note 888828.1. :

1. Log in as the `root` user to an Exadata Database Machine database server that has `root` user SSH access to the switches.

The database server must be on the same InfiniBand Network Fabric network as the switches.

2. Use the `version` command to determine the version of the software the InfiniBand Network Fabric switch is running.

```
[root@dbm01-ibs0 ~]# version

SUN DCS 36p version: 2.1.8-1
Build time: Sep 18 2015 10:26:47
SP board info:
Manufacturing Date: 2015.07.01
Serial Number: "NCDLD0049"
Hardware Revision: 0x0200
Firmware Revision: 0x0000
BIOS version: SUN0R100
BIOS date: 06/22/2010
```

3. Download the appropriate patch file to the database server.  
Refer to My Oracle Support note 888828.1 for patch information.

4. Uncompress the update.

The files are uncompressed to the `patch_release.date` directory.

5. Create a file listing all the InfiniBand Network Fabric switches that need to be updated, with one switch per line.

Use the command `ibswitches` to identify the switches in your rack. Note that switches from Engineered Systems other than Exadata Database Machine might be visible on the same fabric but should probably not be updated at this time.

The following is an example of the file constructed after running the `ibswitches` command:

```
[root@dm01]# cat ibswitches.lst
myibswitch-01
myibswitch-02
```

 **Note:**

If no file name is provided, then the command will be executed on all InfiniBand Network Fabric switches discovered from this host by running `ibswitches` command.

6. Change to the `patch_release.date` directory.

7. Run the prerequisite check.

```
[root@dm01]# ./patchmgr --ibswitches ibswitches.lst --upgrade --
ibswitch_precheck [--force] [--unkey]
```

 **Note:**

Starting with Oracle Exadata System Software release 19.3.0, the `patchmgr` command uses `--` instead of a single hyphen before keywords.

The `--unkey` option removes passwordless SSH access to the InfiniBand Network Fabric switches before exiting.

The `--force` option overrides failures in the InfiniBand Network Fabric topology and connectivity from the servers to the switches. This does not affect the upgrade of the switch.

If the output from the command shows overall status is `SUCCESS`, then proceed with the upgrade. If the output from the command shows overall status is `FAIL`, then review the error summary in the output to determine which checks failed, and then correct the errors. After correcting all the errors, rerun the prerequisite checks until it is successful.

**Related Topics**

- [Exadata Database Machine and Exadata Storage Server Supported Versions \(My Oracle Support Doc ID 888828.1\)](#)
- [Setting up SSH Equivalence](#)  
Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

## 8.10.2 Updating InfiniBand Network Fabric Switch Firmware Software

Update the InfiniBand Network Fabric switches using the `patchmgr` command.

The minimum switch firmware release that can use the `patchmgr` utility is release 1.3.3-2. The switch firmware is upgraded in a rolling manner.

- If a spine switch is present in the rack, then the spine switch is upgraded first.
- If a spine switch is not in the rack, then upgrade the switch that is running the Subnet Manager first.
- If the Subnet Manager is not running on the switches, then perform the upgrade in any order.

You should have completed the steps in [Preparing for InfiniBand Network Fabric Switch Firmware Updates](#) before starting this task.

1. Log in as the `root` user to a database server in the Oracle Exadata Rack that has `root` user SSH access to the switches.

The database server must be on the same InfiniBand Network Fabric network as the switches.

2. Download the appropriate patch file to the database server.  
Refer to My Oracle Support note 888828.1 for patch information.

3. Uncompress the patch files.

The files are uncompressed to the `patch_release.date` directory.

4. Change to the `patch_release.date` directory.

5. Run the prerequisite checks using `patchmgr`.

The `--unkey` option removes passwordless SSH access to the InfiniBand Network Fabric switches before exiting.

The `--force` option overrides failures in the InfiniBand Network Fabric topology and connectivity from the servers to the switches. This does not affect the upgrade of the switch.

```
./patchmgr --ibswitches ibswitches.lst --upgrade --ibswitch_precheck [--force] [--unkey]
```

 **Note:**

Starting with Oracle Exadata System Software release 19.3.0, the `patchmgr` command uses `--` instead of a single hyphen before keywords.

If the output from the command shows overall status is `SUCCESS`, then proceed with the upgrade.

If the output from the command shows overall status is `FAIL`, then review the error summary in the output to determine which checks failed, and then correct the errors.

After the errors have been corrected, rerun the prerequisite checks until it is successful.

6. Upgrade the switches using the following command:

```
[root@dm01]# ./patchmgr --ibswitches ibswitches.lst --upgrade [--force] [--unkey]
```

7. Check the output from the command, and verify the upgrade.

The output should show `SUCCESS`. If there are errors, then correct the errors and run the upgrade command again.

### 8.10.3 Downgrading InfiniBand Network Fabric Switch Firmware

Downgrading firmware means reapplying the older firmware update, which is shipped with the Oracle Exadata System Software update.

The current Oracle Exadata System Software update determines what release you can downgrade to. This may be different for each release and may not be the firmware you were on before the update. For more information on the older firmware shipped with the release you are updating to, see the patch README file.

Complete all steps in this task as the `root` user.

1. Run the `patchmgr` command with the `--precheck` option to verify the switch firmware is ready to be downgraded.

The `ibswitches.lst` file is a file that contains the host names of all the InfiniBand Network Fabric switches that need to be updated, with one switch per line.

```
./patchmgr --ibswitches ibswitches.lst --downgrade --ibswitch_precheck [--force] [--unkey]
```



 **Note:**

Starting with Oracle Exadata System Software release 19.3.0, the `patchmgr` command uses `--` instead of a single hyphen before keywords.

The `--unkey` option removes passwordless SSH access to the InfiniBand Network Fabric switches before exiting.

The `--force` option overrides failures in the InfiniBand Network Fabric topology and connectivity from the servers to the switches. This does not affect the upgrade of the switch.

2. Use the `patchmgr` command to downgrade the firmware on the InfiniBand Network Fabric switches.

```
./patchmgr --ibswitches ibswitches.lst --downgrade [--force] [--unkey]
```

3. Use the `version` command to verify the firmware on the switch has been downgraded.

```
version

SUN DCS 36p version: 2.1.8-1
Build time: Sep 18 2015 10:26:47
SP board info:
Manufacturing Date: 2015.07.01
Serial Number: "NCDLD0049"
Hardware Revision: 0x0200
Firmware Revision: 0x0000
BIOS version: SUN0R100
BIOS date: 06/22/2010
```

## 8.11 Upgrading Oracle Java SE on Oracle Linux

You can upgrade Oracle Java SE (JDK) running Oracle Linux 6 on database servers and storage servers.

Starting with Oracle Exadata System Software release 12.1.2.1.0, the Exadata Database Machine servers include the Java JDK package. Earlier releases of Oracle Exadata System Software do not use the JDK package. If a server running an earlier release of Oracle Exadata System Software has a package installed, such as `java-version-openjdk`, the package is not used by Exadata Database Machine and can be removed. See My Oracle Support note 1405320.1 for details.

To update the JDK package, you must download and update the JDK RPM package, and then reconfigure MS to use the new JDK package. You can update the JDK package by either by configuring YUM to use ULN (currently available only for Oracle Linux 6) or by direct package download (for those not able to use ULN).

- **Stop the MS Process**  
Before updating the JDK package, you must stop the Management Server (MS)

- [Download and Update the Java JDK Package](#)  
Depending on your current environment, you can use one of three procedures to download and update the JDK package.
- [Reconfigure Management Server \(MS\)](#)  
After you update the Java JDK package, you must reconfigure the MS processes to use the updated JDK version.

## 8.11.1 Stop the MS Process

Before updating the JDK package, you must stop the Management Server (MS)



### Note:

These steps are only applicable to on-premise deployments and management domains (dom0). Guest domains (domUs) do not have MS installed.

1. Log in as the `root` user to the server.
2. Stop MS.

- For database servers:

```
dbmcli -e alter dbserver shutdown services ms
```

- For storage servers:

```
cellcli -e alter cell shutdown services ms
```

## 8.11.2 Download and Update the Java JDK Package

Depending on your current environment, you can use one of three procedures to download and update the JDK package.

Using YUM and Unbreakable Linux Network (ULN) is only supported for Exadata Database Machine on-premise database servers running Oracle Linux 6. It is not supported for storage servers or Oracle VM environments, either management domains (dom0) or user domains (domU).

Use one of the following methods to download and update the JDK package:

- [Using YUM and ULN to Update the JDK Package on Database Servers](#)  
If the database servers for your on-premise Exadata Database Machine are running Oracle Linux 6, then you can use YUM and Unbreakable Linux Network (ULN) to simplify the update process.
- [Manually Update the JDK Package on Oracle Exadata System Software versions 12.1.2.1.0 to 12.1.2.2.0](#)  
Update the JDK 7 package to the latest release by downloading the latest version of the package and using the `rpm` utility to install it.
- [Manually Update the JDK Package on Oracle Exadata System Software Release 12.1.2.2.1 and Later](#)  
Update the JDK 8 package to the latest release by downloading the latest version of the package and using the `rpm` utility to install it.

## 8.11.2.1 Using YUM and ULN to Update the JDK Package on Database Servers

If the database servers for your on-premise Exadata Database Machine are running Oracle Linux 6, then you can use YUM and Unbreakable Linux Network (ULN) to simplify the update process.

You can either download the JDK package directly, or use YUM and ULN. Using YUM and ULN for updating the JDK package involves the following tasks:

- [Configuring YUM to Connect to the Oracle Public Repository](#)  
The Oracle Linux YUM server offers a free and convenient way to install the latest Oracle Linux packages. Unbreakable Linux Network (ULN) enables you to update patches and errata for a specific version.
- [Registering an Oracle Linux 6 System with ULN](#)  
After you have configured the access to the YUM repository, you must register your Oracle Linux database servers with ULN.
- [Upgrading JDK on Database Servers Using ULN](#)

### 8.11.2.1.1 Configuring YUM to Connect to the Oracle Public Repository

The Oracle Linux YUM server offers a free and convenient way to install the latest Oracle Linux packages. Unbreakable Linux Network (ULN) enables you to update patches and errata for a specific version.

This procedure is only supported for Exadata Database Machine on-premise database servers running Oracle Linux 6. It is not supported for storage servers or Oracle VM environments, either management domains (dom0) or user domains (domU).

Oracle Linux Yum Server is the Public Repository. For more information on how to obtain updates from the Oracle Linux Yum Server, see <http://yum.oracle.com>. For more information about ULN, see <http://linux.oracle.com>.

1. As the `root` user on a database server, verify that the RHNS-CA-CERT certificate is not expired.

```
yum list installed
```

If the certificate is expired, refer to My Oracle Support Doc ID 2207336.1 to fix the issue.

2. Switch to the `/etc/yum.repos.d` directory.

```
cd /etc/yum.repos.d
```

3. Download the repository configuration file.

Use the `curl` utility to download the repository configuration file that is appropriate for your system. You can alternatively copy the contents of `http://yum.oracle.com/public-yum-release.repo` into a file, where *release* corresponds to the Oracle Linux release, for example `ol6`.

For example, to get the configuration for the Oracle Linux 6 repository using `curl`, use the following command:

```
curl -O http://yum.oracle.com/public-yum-ol6.repo
```

for more information.

4. On a database server, install the `rhn-setup` package to enable `uln_register`.

**⚠ WARNING:**

Do not register your storage servers with ULN or the public yum server.

```
yum install rhn-setup.noarch
```

**⚠ WARNING:**

If you are not using the latest release of Oracle Linux, you will need to edit the file `/etc/yum.repos.d/public-yum-ol6.repo` and enable the correct repository to match the system version. Contact Oracle Support Services and refer to My Oracle Support Doc ID 2241729.1.

#### Related Topics

- [ULN-CA-CERT to be Updated With SHA-2 Certificates on December 15, 2016 My Oracle Support Doc ID 2207336.1](#)
- [Exadata: yum groupinstall "Desktop" failed with Error: Package: exadata-sun-computenode-exact-12.1.2.3.2.160721-1.noarch \(installed\) Requires: libudev\(x86-64\) = 147-2.73.0.1.el6... Updated By: libudev-147-2.73.0.1.el6\\_8.2.x86\\_64 \(ol6\\_latest\) My Oracle Support Doc ID 2241729.1](#)

### 8.11.2.1.2 Registering an Oracle Linux 6 System with ULN

After you have configured the access to the YUM repository, you must register your Oracle Linux database servers with ULN.

**⚠ WARNING:**

Do not register your storage servers with ULN or the public yum server.

1. Run `uln_register` to register your system with ULN.

```
uln_register
```

If you do not have a ULN account, you can register at <https://linux.oracle.com>. Registering for ULN requires a valid customer support identifier (CSI) for Oracle Linux or Oracle VM support.

When you register your system, if a proxy server is required, then use the `--proxy` option to specify the HTTP proxy to use.

```
uln_register --proxy=proxy_hostname:port_number
```

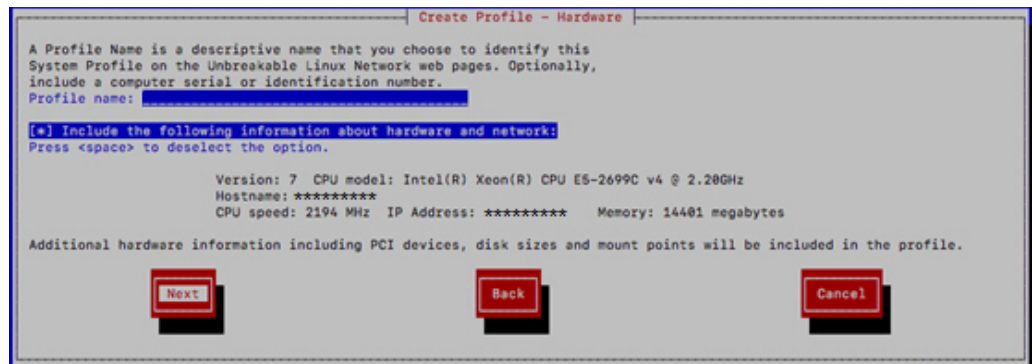
If your proxy requires authentication, then use the additional options `--proxyUser` and `--proxyPassword` to specify the user name and password.

```
uln_register --proxy=proxy_hostname:port_number
--proxyUser=username --proxyPassword=password
```

2. When prompted, enter your ULN user name, password, and CSI.



3. On the Create Profile — Hardware page, enter the required information.
  - a. Enter a name for the system that will allow you to identify it on ULN.
  - b. Choose whether to upload hardware and software profile data that allows ULN to select the appropriate packages for your system.



### 8.11.2.1.3 Upgrading JDK on Database Servers Using ULN

Before starting this step, make sure you have completed the steps in [Stop the MS Process](#).

After you have configured the YUM repository and registered your database server with ULN, you can download the RPM from the ULN channel.

For Oracle Exadata System Software versions 12.1.2.1.0 to 12.1.2.2.0, the Oracle Exadata Database Server includes the JDK 7 package installed as an RPM. To update the RPM, use the ULN channel **Java SE 7 for Oracle Linux**. Make sure you look for JDK 7--do not use JDK 8 or later.

For Oracle Exadata System Software versions 12.1.2.2.1 and later, the Oracle Exadata Database Server includes the JDK 8 package installed as an RPM. To update the RPM, use the ULN channel **Java SE 8 for Oracle Linux**. Make sure you look for the JDK 8.

**⚠ WARNING:**

Do not register your storage servers with ULN.

1. With a web browser, log in to <https://linux.oracle.com/>.
2. Find the Oracle Exadata Database Server on which you want to install Oracle Java SE and click on the name of that system.

In the following image, the server names have been blanked out.

System	OS Release	Registered
*****_oracle.com	6	2018-04-30
*****_oracle.com	6	2018-04-27
*****_oracle.com	6	2018-03-22
*****_oracle.com	6	2018-02-05
hostname	7	2017-11-17

3. Click in **Manage Subscriptions**.
4. Using the arrow buttons, move the desired Java SE channel from the **Available channels** list to the **Subscriber Channels** list, then click **Save Subscriptions**.

5. When prompted, click **Accept** to accept the license agreement.

**6. On the database server, from the `yum.repos.d` directory, upgrade the JDK.**

```
yum check-update jdk
Loaded plugins: rhnplugin, ulninfo
This system is receiving updates from ULN.
ol6_x86_64_JavaSE7_public
By downloading and/or using this software program you agree that
your use is
subject to the applicable license agreement at https://
linux.oracle.com
/licenses.html.

exadata_dbserver_18.1.4.0.0_x86_64_base |
1.2 kB 00:00
exadata_dbserver_18.1.4.0.0_x86_64_base/primary |
436 kB 00:00
exadata_dbserver_18.1.4.0.0_x86_64_base
470/470
ol6_x86_64_JavaSE7_public |
1.2 kB 00:00
ol6_x86_64_JavaSE7_public/primary |
5.9 kB 00:00
ol6_x86_64_JavaSE7_public
16/16
ol6_x86_64_kspllice |
1.2 kB 00:00
ol6_x86_64_kspllice/primary |
160 kB 00:00
ol6_x86_64_kspllice
1436/1436
public_ol6_latest
40132/40132
...
Running transaction
Installing : jdk1.8.0_172-1.8.0_172-
fcs.x86_64 1/1
Unpacking JAR files...
rt.jar...
jsse.jar...
charsets.jar...
localedata.jar...
jfxrt.jar...
Verifying : jdk1.8.0_172-1.8.0_172-
fcs.x86_64 1/1

Installed:
jre.x86_64 0:1.8.0_172-
fcs

Complete!
```

To complete the JDK update, continue with the steps in [Reconfigure Management Server \(MS\)](#).

## 8.11.2.2 Manually Update the JDK Package on Oracle Exadata System Software versions 12.1.2.1.0 to 12.1.2.2.0

Update the JDK 7 package to the latest release by downloading the latest version of the package and using the `rpm` utility to install it.

Before starting this step, make sure you have completed the steps in [Stop the MS Process](#).

With Oracle Exadata System Software versions 12.1.2.1.0 to 12.1.2.2.0, the Oracle Exadata Storage Server includes the JDK 7 package installed as an RPM.

Following any upgrades to the server image (using `dbnodeupdate` or `patchmgr`), check the JDK version. If the JDK package is reverted to an older version during the upgrade, then use the procedures here to update the JDK package to the latest version.

1. Download the latest version of JDK 7 using the links found in My Oracle Support Doc ID 1439822.1. **Do not download JDK 8 or later.**

2. Extract the contents of the ZIP file.

3. Locate the JDK RPM.

The name of the file is similar to `jdk-version-linux-x64.rpm`, for example `jdk-7u91-linux-x64.rpm`.

4. Copy only the RPM file to the target server.

The file can be placed in a temporary directory, such as `/tmp`.

5. As the `root` user, determine the current version of the installed JDK RPM.

```
rpm -q jdk
```

6. If the JDK package is installed and needs to be updated, then use the `rpm` command to install the update.

```
rpm -Uvh /tmp/jdk-version-linux-x64.rpm
```

7. Verify the JDK package was updated.

```
rpm -q jdk
```

8. Removed the staged update file.

```
rm -f /tmp/jdk-version-linux-x64.rpm
```

To complete the JDK update, continue with the steps in [Reconfigure Management Server \(MS\)](#).

### Related Topics

- [All Java SE Downloads on MOS \(My Oracle Support document 1439822.1\)](#)
- [HOWTO: Update JDK on Exadata Database Nodes \(My Oracle Support document 2069987.1\)](#)
- [HOWTO: Update JDK on Exadata Storage Cell Nodes \(My Oracle Support document 2075464.1\)](#)



### 8.11.2.3 Manually Update the JDK Package on Oracle Exadata System Software Release 12.1.2.2.1 and Later

Update the JDK 8 package to the latest release by downloading the latest version of the package and using the `rpm` utility to install it.

Before starting this step, make sure you have completed the steps in [Stop the MS Process](#).

With Oracle Exadata System Software release 12.1.2.2.1 and later, the Oracle Exadata servers includes the JDK 8 package installed as an RPM.

Following any upgrades to the server image (using `dbnodeupdate` or `patchmgr`), check the JDK version. If the JDK package is reverted to an older version during the upgrade, then use the procedures here to update the JDK package to the latest version.

1. Download the latest version of JDK 8 using the links found in My Oracle Support Doc ID 1439822.1. **Download only JDK 8 updates.**

2. Extract the contents of the ZIP file.

3. Locate the JDK RPM.

The name of the file is similar to `jdk-version-linux-x64.rpm`, for example `jdk-8u172-linux-x64.rpm`.

4. Copy only the RPM file to the target server.

The file can be placed in a temporary directory, such as `/tmp`.

5. As the `root` user, determine the current version of the installed JDK RPM.

```
rpm -qa|grep jdk
jdk1.8.0_66-1.8.0_66-fcs.x86_64
```

6. If the JDK package is installed and needs to be updated, then use the `rpm` command to install the update.

```
rpm -Uvh /tmp/jdk-version-linux-x64.rpm
```

7. Verify the JDK package was updated.

With JDK 8, the updated package does not replace the currently installed package, so you will see two version of the JDK package installed.

```
rpm -qa | grep jdk
jdk1.8.0_66-1.8.0_66-fcs.x86_64
jdk1.8.0_172-1.8.0_172-fcs.x86_64
```

8. Remove the older JDK package from the server.

If the older version was update 66, then the command would be as follows:

```
rpm -e --nodeps jdk.1.8.0_66-1.8.0_66-fcs.x86_64
```

9. Verify only the updated JDK is available on the server.

```
rpm -qa |grep jdk
jdk1.8.0_172-1.8.0_172-fcs.x86_64
```

10. Removed the staged update file.

```
rm -f /tmp/jdk-version-linux-x64.rpm
```

To complete the JDK update, continue with the steps in [Reconfigure Management Server \(MS\)](#).

#### Related Topics

- [All Java SE Downloads on MOS \(My Oracle Support Doc ID 1439822.1\)](#)
- [HOWTO: Update JDK on Exadata Database Nodes \(My Oracle Support Doc ID 2069987.1\)](#)
- [HOWTO: Update JDK on Exadata Storage Cell Nodes \(My Oracle Support Doc ID 2075464.1\)](#)

### 8.11.3 Reconfigure Management Server (MS)

After you update the Java JDK package, you must reconfigure the MS processes to use the updated JDK version.

These steps are only applicable on physical deployments and management domains (dom0) on virtual deployments because user domains (domUs) do not have MS installed.

1. Log in to the database server or storage server as `root`.
2. Verify the MS process is still stopped.

- For database servers:

```
dbmcli -e alter dbserver shutdown services ms
```

- For storage servers:

```
cellcli -e alter cell shutdown services ms
```

3. For Oracle Exadata System Software release 12.2.x, version 12.2.1.1.4 or later, or Oracle Exadata System Software release 18c, version 18.1.2 or later, perform the following steps (not needed for other releases):

- a. Go to the UNIX scripts directory.

- For database servers:

```
cd /opt/oracle/dbserver/dbms/deploy/scripts/unix
```

- For storage servers:

```
cd /opt/oracle/cell/cellserv/deploy/scripts/unix
```

- b. Redeploy MS.

- For database servers:

```
sh setup_dynamicDeploy DB
```

- For storage servers:

```
sh setup_dynamicDeploy
```

#### 4. Restart the MS.

- For database servers:

```
dbmcli -e alter dbserver startup services ms
```

- For storage servers:

```
cellcli -e alter cell startup services ms
```

## 8.12 Setting up SSH Equivalence

Before updating software on your Oracle Exadata Database Machine, you must configure SSH equivalence.

You can run the Exadata update utilities for Oracle Exadata Database Server, Oracle Exadata Storage Server, and the RDMA Network Fabric switch as either `root` or as a non-`root` user from any server running Oracle Linux. The utility can perform precheck, update, and rollback actions on any Exadata server as long as SSH equivalence is set up for the `root` user for the target Exadata server.

1. Prepare a file named `cell_group` or `dbs_group` that has one storage server or database server host name or IP address per line for each storage server or database server you want to update.
2. Check for existing SSH equivalence.

The following command should require no password prompts and no interaction. It should return the list of host names in the `cell_group` file.

```
[oracle@nonExadataHost]# ./dcli -g cell_group -l root 'hostname -i'
```

3. Set up SSH equivalence if not already done so from the launch server.

Do not do this step if you already have `root` SSH equivalence.

Generate SSH keys using the following command:

```
[oracle@nonExadataHost]# ssh-keygen [-t rsa]
```

You can use the `-t` option to specify the key type, such as RSA or DSA. If you do not include the `-t` option, then RSA is configured by default.

Accept the defaults so that the SSH keys are created for the `root` user.

4. Push the SSH keys to set up SSH equivalence.

Enter the `root` password when prompted.

```
[oracle@nonExadataHost]# dcli -g cell_group -l root -k
```

 **Note:**

Customers in secure environments may have chosen to disable SSH access to Oracle Exadata Storage Server. During normal operations, Oracle Exadata Storage Server does not require SSH access. However, administrative utilities such as the update utility require SSH access. See the subsection "Unlocking a Cell Temporarily" in the topic Disabling SSH on Storage Servers for information on unlocking storage servers.

## 8.13 Troubleshooting Software Updates on Exadata Database Machine

Review these topics if you encounter errors or problems when updating the software on Exadata Database Machine

- [Troubleshooting Exadata Database Server Updates](#)  
You can use the log files generated by the update utility to troubleshoot updates.
- [Monitoring, Validating, and Troubleshooting Exadata Storage Server Updates](#)
- [Troubleshooting a Failing Prerequisite Check Due to Dependency Problems on Exadata Database Server](#)  
When you run a prerequisite check or an update on Exadata database servers that have custom rpms installed, conflicts may cause the prerequisite check or update to fail.
- [Troubleshooting a Multilib Problem on Exadata Database Server](#)

### 8.13.1 Troubleshooting Exadata Database Server Updates

You can use the log files generated by the update utility to troubleshoot updates.

The update utility orchestrates updating the Exadata database servers. Updating database nodes with the `patchmgr` tool is less verbose because it prints only minimal information to the screen. If additional information is required, you can view the `patchmgr` logs and the `dbnodeupdate.sh` logs that `patchmgr` copies over from other servers, if available. The log file (`dbnodeupdate.log`) and the `diag` file (`dbnodeupdate.<runid>.diag`) will eventually exist on two locations:

- On each updated database server, in the `/var/log/cellos` directory
- Consolidated on the node running the update utility.

On the node running the update utility, if the `-log_dir` flag was set to "auto", the log files will be stored in the `log/<directory based on contents of nodes in list file>` directory, relative from the directory where the update utility is started from. For example, if the update utility is located in `/u01/dbserver.patch`, then the log directory may be `/u01/dbserver.patch/dm01db01_dm01db02_e8f1f753`.

Important files found in the log directory are:

- `patchmgr.log` contains the consolidated screen output from running the remote update commands on the different database servers.
- `<hostname>_dbnodeupdate.<runid>.diag` is the diag file for the specific run on a database server.
- `<hostname>_dbnodeupdate.log` contains `dbnodeupdate.log` output appended from `/var/log/cellos` from the remote database server.

When a prerequisite check, backup, update, or rollback fails, error messages on screen should provide information on which step failed on which node. Consult the log files mentioned above if more information is required. Search the log file for the start of a new run (search for “zzz”).

Check if the time matches your run. If it matches, note the runid for further reference. Then search for `ERROR`.

If an update action fails before the actual YUM update, you can retry the update after resolving the error. If the update aborted half way, it is recommended that you roll back, resolve the error, and retry.

In rare cases, `patchmgr` may be unable to determine the status of an update, whether the update was successful or not. In such cases, it displays a message that the update failed. However, it is possible that the update still completed successfully. To determine the actual status of the update:

- Check the image status of the (database) node. You can do this by running the `imageinfo` command. The `Image status` line displays the status.
- Check the version of the Exadata software. This can also be determined from the `imageinfo` command.

If the image status is success, and the Exadata version is the new expected version, then the update was successful and you can ignore the `update failed` message.

Then:

- Run `dbnodeupdate.sh -c` manually on the particular node to perform the completion steps of the update.
- Remove the completed node from the (database) node file.
- Rerun `patchmgr` to perform the update on the remaining nodes.

Other things to check if the update fails include:

- The correct syntax for using `patchmgr` to update database nodes can be found in the `patchmgr` online help.
- SSH equivalence must be configured before using `patchmgr`.
- Download the latest `dbserver.patch.zip` from My Oracle Support note 1553103.1.
- Open a service request with Oracle Support Services to analyze why the `patchmgr` orchestration failed.

### Related Topics

- [Exadata Patchmgr Update Utility](#)  
Patchmgr is the utility used to update software for Exadata infrastructure components.

- [dbnodeupdate.sh and dbserver.patch.zip: Updating Exadata Database Server Software using the DBNodeUpdate Utility and patchmgr \(My Oracle Support Doc ID 1553103.1\)](#)

## 8.13.2 Monitoring, Validating, and Troubleshooting Exadata Storage Server Updates

The recommended way to monitor Exadata storage servers being updated is through e-mail alerts. These alerts can be set up by specifying email addresses in the `-smtp_from` and `-smtp_to` flags in the update utility.

If needed for troubleshooting, you can monitor update activity using `less -rf patchmgr.stdout` from another terminal session or window to see raw log details from the update utility.

You can also monitor activity of the storage server by logging in to the serial console or web-based ILOM console of individual storage servers being updated 5 minutes after the update utility has started. Waiting 5 minutes allows the update utility time to reset the ILOM. Resetting the ILOM disconnects you from the ILOM web console and serial console. You can reconnect once the ILOM has been reset. By waiting 5 minutes, you avoid having to reconnect. You lose the connection during any ILOM update, and need to reconnect. The ILOM does not show any update actions. When needed it is helpful to monitor the activities of the normal cell boot, reboot, and other activities to ensure that the process is proceeding correctly.

Verify the update status after the `patchmgr` utility completes as follows:

Check image status and history using the `imageinfo` and `imagehistory` commands on each cell. A successful update to a system with high capacity or high performance drives shows output similar to the following.

```
Kernel version: 2.6.39-400.281.1.el6uek.x86_64 #1 SMP Fri Jun 17 20:10:16
PDT 2016 x86_64
Cell version: OSS_12.1.2.3.2_LINUX.X64_160721
Cell rpm version: cell-12.1.2.3.2_LINUX.X64_160721-1.x86_64
```

```
Active image version: 12.1.2.3.2.160721
Active image kernel version: 2.6.39-400.281.1.el6uek
Active image activated: 2016-07-21 13:04:34 -0500
Active image status: success
Active system partition on device: /dev/md5
Active software partition on device: /dev/md7
```

```
Cell boot usb partition: /dev/sda1
Cell boot usb version: 12.1.2.3.2.160721
```

```
Inactive image version: 12.1.2.3.1.160411
Inactive image activated: 2016-04-11 19:58:28 -0700
Inactive image status: success
Inactive system partition on device: /dev/md6
Inactive software partition on device: /dev/md8
```

```
Inactive marker for the rollback: /boot/I_am_hd_boot.inactive
Inactive grub config for the rollback: /boot/grub/grub.conf.inactive
```

```
Inactive kernel version for the rollback: 2.6.39-400.128.21.el5uek
Rollback to the inactive partitions: Possible
```

**A successful update to a system with extreme flash drives shows a similar output but has a different path for USB:**

```
Kernel version: 2.6.39-400.281.1.el6uek.x86_64 #1 SMP Fri Jun 17 20:10:16 PDT
2016 x86_64
Cell version: OSS_12.1.2.3.2_LINUX.X64_160721
Cell rpm version: cell-12.1.2.3.2_LINUX.X64_160721-1.x86_64
```

```
Active image version: 12.1.2.3.2.160721
Active image kernel version: 2.6.39-400.281.1.el6uek
Active image activated: 2016-07-21 13:04:34 -0500
Active image status: success
Active system partition on device: /dev/md5
Active software partition on device: /dev/md7
```

```
Cell boot usb partition: /dev/sda1
Cell boot usb version: 12.1.2.3.2.160721
```

```
Inactive image version: 12.1.2.3.1.160411
Inactive image activated: 2016-04-11 19:58:28 -0700
Inactive image status: success
Inactive system partition on device: /dev/md6
Inactive software partition on device: /dev/md8
```

```
Inactive marker for the rollback: /boot/I_am_hd_boot.inactive
Inactive grub config for the rollback: /boot/grub/grub.conf.inactive
Inactive kernel version for the rollback: 2.6.39-400.128.21.el5uek
Rollback to the inactive partitions: Possible
```

### Related Topics

- [Running the Update for Exadata Storage Servers](#)

## 8.13.3 Troubleshooting a Failing Prerequisite Check Due to Dependency Problems on Exadata Database Server

When you run a prerequisite check or an update on Exadata database servers that have custom rpms installed, conflicts may cause the prerequisite check or update to fail.

When doing a prerequisite check or an update, the update utility makes two dependency checks: checks against “minimum dependencies” and checks against “exact dependencies”.

To triage a dependency failure, look in the log file. This can be either `dbnodeupdate.log` on the target Exadata database server or `<hostname>_dbnodeupdate.log` in the log directory from where the update utility (patchmgr) was launched.

To locate the start of each run in the log file, search for a line that starts with “zzz”. For example:

```
zzz - /u01/patches/YUM/dbnodeupdate.sh called with arguments -u -l
/u01/patches/YUM/p23564643_121232_Linux-x86-64.zip -v -N at 2016-08-23
23:31:54
```

The date stamp should match the time of the run you are researching. Each run is identified with a unique `runid` which can also be found at the start of each run in the same log file:

```
[1472009516][2016-08-23 23:31:59 -0400][INFO][u01/patches/YUM/
dbnodeupdate.sh][InitLogfile][] # dbnodeupdate.sh script rel.
: 5.160809 started at (runid :230816233155)
```

You can check for custom rpms, if any, by looking at the `diag` file of the specific run. The `diag` file is identified by the `runid` and can be found in `/var/log/cellos` on the target Exadata database server or in the directory driving the update utility. The filename for this example would be `dbnodeupdate.230816233155.diag`. In that file, look for the section with heading `RunDetectCustomRpmsSh` and `rpm -qa --qf "%{n}-%{v}-%{r}-%{arch}\n`. You can find details on the (additional) packages installed (if any).

In the log file, search down from the start (“zzz”) for “Exact dependencies” (case sensitive) for the results of the “minimum” and “exact” dependency checks. In this case it says:

```
Exact dependencies : Will fail on a next update
Minimum dependencies : Will fail on a next update
```

As long as the minimum dependencies check passes, updates or prerequisite checks won't fail. When both minimum and exact dependencies fail, you need to find out what caused the error.

In order to find the dependency that is causing the error, search for “[ExecUpgrade][ ] Performing yum package dependency”. This is where the YUM run (typically a dry-run first) is executed. When there is a dependency problem, you should see a YUM message starting with “Error:”. For example:

```
--> Finished Dependency Resolution
Error: Package: krb5-devel-1.10.3-33.el6.x86_64 (@ol6_latest)
Requires: krb5-libs = 1.10.3-33.el6
Removing: krb5-libs-1.10.3-33.el6.x86_64 (installed)
 krb5-libs = 1.10.3-33.el6
Updated By: krb5-libs-1.10.3-42z1.el6_7.x86_64
 (exadata_generated_160616114412)
 krb5-libs = 1.10.3-42z1.el6_7
You could try using --skip-broken to work around the problem
```

In the example error message above, the `krb5-devel-1.10.3-33.el6.x86_64` (`krb5-devel`) package is installed from a non-Exadata channel (`ol6_latest`). This `krb5-devel` package depends on `krb5-libs`. However in this Exadata update the `krb5-libs` package is not available. YUM fails the dependency check because updating `krb5-libs` is not possible without also updating `krb5-devel`. Because a new version of `krb5-devel` is not included in the Exadata update the package should be either “pre-updated” or removed. Pre-updating



means updating the individual package manually before running the Exadata update. This can be done using the command `rpm -Uvh <package-name>`.

Removing the custom package is recommended and should be done using the following rpm command:

```
[root@dm01]# rpm -e krb5-devel
```

After removing the rpm that is causing the error, restart the update or prerequisite check.

#### Related Topics

- [Running Prerequisite Checks](#)

## 8.13.4 Troubleshooting a Multilib Problem on Exadata Database Server

If you have custom packages with different architectures installed on database servers, you may see a similar problem as that described in [Troubleshooting a Failing Prerequisite Check Due to Dependency Problems on Exadata Database Server](#). This typically happens when i686 packages are installed on an Exadata database server.



#### Note:

While non-64-bit (x86\_64) packages are supported, it is recommended to stay away from 32-bit software. When you need specific functionality from a third party, it is recommended to ask for a 64-bit version.

Typically Exadata-branded x86\_64 bit rpms are updated in an Exadata update. However, when you have installed similar packages of non-x86\_64 bit architecture, the update utility cannot update the 64-bit packages. You would see the following error in the log file:

```
--> Finished Dependency Resolution
Error: Multilib version problems found. This often means that the root
 cause is something else and multilib version checking is just
 pointing out that there is a problem. Eg.:

 1. You have an upgrade for libuuid which is missing some
 dependency that another package requires. Yum is trying to
 solve this by installing an older version of libuuid of the
 different architecture. If you exclude the bad architecture
 yum will tell you what the root cause is (which package
 requires what). You can try redoing the upgrade with
 --exclude libuuid.otherarch ... this should give you an error
 message showing the root cause of the problem.

 2. You have multiple architectures of libuuid installed, but
 yum can only see an upgrade for one of those architectures.
 If you don't want/need both architectures anymore then you
 can remove the one with the missing update and everything
 will work.

 3. You have duplicate versions of libuuid installed already.
```

You can use "yum check" to get yum show these errors.

...you can also use `--setopt=protected_multilib=false` to remove this checking, however this is almost never the correct thing to do as something else is very likely to go wrong (often causing much more problems).

```
Protected multilib versions: libuuid-2.17.2-12.24.0.1.el6.x86_64 !=
libuuid-2.17.2-12.18.0.1.el6.i686
```

The solution for multilib problems is to remove the i686 or i386 package by running `"rpm -e <package_name.i686>"`.

# 9

## Using the DBMCLI Utility

The Oracle Exadata Database Machine Command-Line Interface (DBMCLI) utility is used to manage database servers and provides many of the features that are provided with SQL\*Plus, including the use of script files.

- [Overview of the DBMCLI Utility](#)  
The DBMCLI utility is the command-line administration tool for configuring database servers, and managing objects in the server environment.
- [About DBMCLI Administration Commands](#)  
DBMCLI administrative commands do not act directly on objects.
- [About DBMCLI Object Commands](#)  
DBMCLI has object commands, object types, and object attributes.
- [About DBMCLI Object Types](#)
- [DBMCLI Object Attributes](#)
- [Monitoring Database Server with Metrics](#)  
You can monitor a database server by viewing its metrics.
- [DBMCLI Command Reference](#)  
DBMCLI has both administrative and object commands.

### 9.1 Overview of the DBMCLI Utility

The DBMCLI utility is the command-line administration tool for configuring database servers, and managing objects in the server environment.

DBMCLI runs on each server to enable you to configure an individual database server. You use DBMCLI to start and stop the server, to manage server configuration information, and to enable or disable servers. The command-line utility is already installed when Exadata Database Machine is shipped.

- [Starting DBMCLI](#)
- [Understanding Command Syntax and Options for DBMCLI](#)  
This topic describes the syntax and command options for DBMCLI.
- [Reserved Words](#)
- [DBMCLI Command-Line Editing](#)
- [DBMCLI Input and Output Options](#)
- [Comments in DBMCLI Scripts](#)
- [Line Continuation in DBMCLI Commands](#)

## 9.1.1 Starting DBMCLI

You can start DBMCLI from the operating system command line on the database server that you want to manage or remotely from a network-attached client using Secure Shell (SSH). The command-line syntax is as follows:

```
dbmcli [port_number] [-n] [-m] [-xml] [-v | -vv | -vvv] [-x] [-e
command]
```

The *port\_number* argument specifies the HTTP port number of the database server. If the *port\_number* argument is omitted, then the DBMCLI utility uses the value assigned to the `HTTP_PORT` variable in the `cellinit.ora` file on the database server. The file is located in the `/opt/oracle/dbserver/dbms/deploy/config/` directory. If the `HTTP_PORT` variable in the `cellinit.ora` file is not set, then the port number defaults to 8888.

The following table lists the options that can be used with the DBMCLI command:

Option	Description
-n	Runs the DBMCLI utility in non-interactive mode. This option suppresses the command prompt and disables the command-line editing features.
-m	Runs DBMCLI monitor (read-only) mode.
-xml	Causes output to be displayed in XML format for the Oracle Enterprise Manager.
-v, -vv, and -vvv	Sets the log level. The <code>-v</code> option is for fine, <code>-vv</code> is for finer, and <code>-vvv</code> is for the finest level.
-x	Suppresses the banner.
-e <i>command</i>	Runs the specified DBMCLI command. DBMCLI exits after running the command. For example:  <pre>\$ dbmcli -e list dbserver detail</pre> <pre>\$ dbmcli -e "list dbserver attributes name where name like '*.dbm01'"</pre>

DBMCLI uses the database server operating system authentication. DBMCLI does not have a login parameter or a connect command. The directory from which DBMCLI is invoked is the default directory for unqualified file access in DBMCLI `SPOOL` and `START` commands.

## 9.1.2 Understanding Command Syntax and Options for DBMCLI

This topic describes the syntax and command options for DBMCLI.

DBMCLI syntax is as follows:

```
{admin-command | object-command object} [options] ;
```

In the preceding syntax, the following arguments are used:

- *admin-command* is an administrative action.
- *object-command* is an action performed on an object.
- *object* is an object or target on which a command performs an action.
- *options* extend the use of a command combination to include additional parameters for the command.

When using the DBMCLI utility, the following rules apply:

- Commands, objects, and options are not case-sensitive except where explicitly stated, such as in string patterns used in filtering strings with the `LIKE` operator.
- Use single quotation marks or double quotation marks around the name of an object that includes spaces or punctuation. The use of quotation marks should match. For example, "this is incorrect" is incorrect because the first mark is double quotation marks, and the second is a single quotation mark.
- The current, local cell is the cell to which all DBMCLI commands apply.
- A semicolon (;) is optional at the end of a DBMCLI command.
- A hyphen (-) is used at the end of a line to continue a long command onto the next line. If you are using hyphens in names or to denote negative values, it must be immediately followed by an alphanumeric value.

### 9.1.3 Reserved Words

The following are DBMCLI reserved words:

```
ALERTDEFINITION
ALERTHISTORY
ALL
ALTER
BMC
CONFIGUREBMC
CREATE
DESCRIBE
DETAIL
DROP
FORCE
LED
LIST
LUN
MAIL
MEMORY
METRICDEFINITION
METRICCURRENT
METRICHISTORY
MS
NULL
OFF
ON
PHYSICALDISK
```

```
RESTART
RS
SHUTDOWN
SNMP
STARTUP
THRESHOLD
VALIDATE
```

If these keywords are used as values in commands, then they must be enclosed in quotation marks.

## 9.1.4 DBMCLI Command-Line Editing

The DBMCLI utility supports command-line history and editing, similar to BSD editline and GNU readline functionality. Most of the command editing features of DBMCLI are similar to modern shells, such as `bash` and `tcsh`.

## 9.1.5 DBMCLI Input and Output Options

The database command-line utility reads commands from standard input and writes output to standard output. You can use the host operating system options for redirecting input and output to compose and process command scripts. For example, you can perform the following redirection:

```
$ dbmcli < command-script-in > results-out
```

In the preceding example, the output from DBMCLI commands in the `command-script-in` file are written to the `results-out` file.

## 9.1.6 Comments in DBMCLI Scripts

You can add single-line comments to DBMCLI scripts using several formats. You can begin the comment line with `REMARK`, `REM` or `--` (two hyphens).

For example, the following are valid syntax for comments:

```
REMARK This is a comment
REM This is a comment
-- This is a comment
```

## 9.1.7 Line Continuation in DBMCLI Commands

To continue a long command on to the next line, insert a hyphen (-) at the end of the line. Then, press Enter, and continue typing the command.

For example:

```
DBMCLI> LIST dserver WHERE name LIKE '*.dbm04' -
attributes name, status, comment
```

# 9.2 About DBMCLI Administration Commands

DBMCLI administrative commands do not act directly on objects.

DBMCLI uses the following administration commands:

- [HELP](#)
- [SET](#)
- [SPOOL](#)
- [START](#)



**Note:**

The `dbmadmin` user should be used to run all services on the database server. The `dbmmonitor` user is for monitoring purposes. The `dbmmonitor` user can run the following commands:

- DESCRIBE
- EXIT
- HELP
- LIST
- REMARK
- SET
- START

## 9.3 About DBMCLI Object Commands

DBMCLI has object commands, object types, and object attributes.

The following DBMCLI commands operate on Oracle Database objects:

- [ALTER](#)
- [CREATE](#)
- [DESCRIBE](#)
- [DROP](#)
- [LIST](#)

**Related Topics**

- [DBMCLI Command Reference](#)

## 9.4 About DBMCLI Object Types

The following Oracle Exadata System Software object types can be used with DBMCLI object commands:

- `ALERTDEFINITION` — An alert definition provides a definition for every alert that can be produced on the database server. Alerts are defined on metrics and other sources of alerts.
- `ALERTHISTORY` — An alert history provides a list of alerts that have occurred on the database server.

- `DBSERVER` — The current or local database server.
- `DBSERVER` — A diagpack represents a compressed file under `$LOG_HOME` and contains log files and trace files.
- `IBPORT` — The InfiniBand Network Fabric ports for the database server.

 **Note:**

This command does not work on Exadata Database Machine servers with RoCE Network Fabric.

- `LUN` — Logical unit number (LUN) is the address for a RAID set of physical disk devices. LUNs are automatically discovered when the server is started. They are assigned to the corresponding disk when the disk is first created or when disks are discovered after the system is restarted. LUNs that are not yet assigned to a disk have a `NULL` value.
- `METRICCURRENT` — A current metric describes a set of observations on the current value of an individual metric.
- `METRICDEFINITION` — A metric definition describes the configuration of a metric,
- `METRICHISTORY` — A metric history describes a collection of past individual observations of all metric values.
- `PHYSICALDISK` — A disk is called a physical disk on the server.
- `PRIVILEGE` — A right or permission assigned to a role.
- `ROLE` — A named group of related privileges.
- `SOFTWAREHISTORY` — A software history is a list of final states for past software updates
- `SOFTWAREUPDATE` — An object that contains the software location and time parameters for scheduling software updates.
- `THRESHOLD` — A threshold describes the rules for generating stateful alerts based on a specific metric. The rules include boundary (threshold) values and how long the metric values can violate these boundaries before an alert is generated.
- `USER` — An account that can manage the compute nodes using DBMCLI.

Not all possible command-object combinations are valid. For valid command-object combinations, review the syntax for the specific object command.

## 9.5 DBMCLI Object Attributes

Each DBMCLI object has a set of attributes that are assigned when the object is created or altered. Attribute filters and lists are used to specify which attributes and objects are displayed in the output of the `LIST` command.

All attributes can be displayed, but only some can be modified directly by the user. To display a list of attributes and determine which ones can be modified, use the `DESCRIBE` command.



- [Restrictions on Values of Common Attributes](#)  
Review the following restrictions for the values of attributes used by multiple DBMCLI objects.
- [Attribute Lists in LIST Command](#)
- [Attribute Filters in LIST and ALTER Commands](#)  
You can use the *attribute\_filters* clause to specify the objects to display in `LIST` commands. Some `ALTER` commands also support the *attribute\_filters* clause.

#### Related Topics

- [LIST](#)
- [DESCRIBE](#)

## 9.5.1 Restrictions on Values of Common Attributes

Review the following restrictions for the values of attributes used by multiple DBMCLI objects.

- The value of the `name` attribute must be less than 256 characters and composed only of the following ASCII characters:
  - Lowercase alphabetic characters (a to z)
  - Uppercase alphabetic characters (A to Z)
  - Numbers (0 to 9)
  - Underscore (`_`)
  - Hyphen (`-`)

#### Note:

On Oracle Exadata System Software release 19.2.0, or earlier, you must enclose the string in double quotes. For example: "hyphenated-string"

- The value of the `comment` attribute must be less than 256 characters.

See the syntax of each DBMCLI command for any additional restrictions on attribute values.

## 9.5.2 Attribute Lists in LIST Command

You can specify which attributes to display for the `LIST` command with the following optional clause:

```
ATTRIBUTES { ALL | attribute1 [, attribute2] ... }
```

`ALL` displays all possible object attributes for the `LIST` object combination.

## 9.5.3 Attribute Filters in LIST and ALTER Commands

You can use the *attribute\_filters* clause to specify the objects to display in `LIST` commands. Some `ALTER` commands also support the *attribute\_filters* clause.

This syntax of the *attribute\_filters* clause is:

```
WHERE attribute_filter1 [AND attribute_filter2 ...]
```

Each *attribute\_filterN* has the following syntax:

```
attribute [NOT | !] operator comparison_value
```

The *attribute* placeholder represents the name of the attribute to use for filtering. The supported types of *operator* are listed in the following table. These operators can be combined with *NOT* or *!*.

**Table 9-1 Supported Operators in Attribute Filters**

Operator	Description
=	Tests for equality between string, status, or numeric attributes. For example: <code>status NOT = normal</code>
>	Tests for values greater than the numeric attributes. For example: <code>size &gt; 139920M</code>
<	Tests for values less than the numeric attributes. For example: <code>freeSpace !&lt; 100M</code>
LIKE	Tests for a regular expression match with a string attribute using case-sensitive matching. For example: <code>LIKE 'GD_IO_RQ.*'</code>

When used with the supported operators, *comparison\_value* is one of the following data types:

- Numeric
- Literal: Value such as `active` or `normal`
- Datetime: Time value supported only for `ALERTHISTORY`
- String: Value delimited by single quotation marks ( `' '` ) or double quotation marks ( `" "` )
- `NULL`: Unassigned strings or empty lists

## 9.6 Monitoring Database Server with Metrics

You can monitor a database server by viewing its metrics.

Metrics are of the following type:

- Cumulative: Statistics since the metric was created.
- Instantaneous: Values at the time the metric is collected.
- Rate: Computed value for metric by averaging statistics over a period of time.

- [About Metrics](#)  
Exadata metrics are recorded observations of important properties or values relating to the Exadata system software.
- [Displaying Metrics](#)
- [Database Server Metrics](#)  
Database server metrics provide information about the server, such as CPU utilization, memory or network interconnections.
- [Real-Time Insight](#)  
You can use the real-time insight feature to enable real-time monitoring of your Exadata systems.
- [Using Real-Time Insight](#)  
You can use the real-time insight feature to enable real-time monitoring of your Exadata systems using an external metric collection platform.

## 9.6.1 About Metrics

Exadata metrics are recorded observations of important properties or values relating to the Exadata system software.

Exadata metrics contain detailed statistics for most Exadata components.

Metrics are of the following types:

- Cumulative metrics are statistics that accumulate over time since the metric was created or the server was restarted.
- Instantaneous metrics contain the current value at the time of the metric observation.
- Rate metrics are computed statistics where the value is observed over time.

By default, metric collections occur at 1-minute intervals. However, commencing with Oracle Exadata System Software 22.1.0, you can optionally configure fine-grained metrics. To enable fine-grained metrics, you must specify a collection interval between 1 and 60 seconds. You can also choose the metrics to include in the fine-grained collection.

Fine-grained metric collection is the foundation for real-time metric streaming. Consequently, fine-grained metrics are only recorded in memory to support streaming to an external collection. Regular (non-fine-grained) metrics are initially recorded in memory and later written to a disk-based repository. By default, seven days of metrics history is maintained for regular metrics.

## 9.6.2 Displaying Metrics

You can use the DBMCLI `LIST` command to display and monitor metrics for database objects. You can display metric definitions, current metrics, and metric history using the `LIST` command.

- [Displaying Specific Attributes for Metrics](#)  
Use the `ATTRIBUTES` keyword to list the specific attributes you want to view in the output.
- [Displaying Metric Definitions](#)  
Use the `LIST METRICDEFINITION` command to display the metric definitions for a database server.

- [Displaying Current Metrics](#)  
Use the `LIST METRICCURRENT` command to display the current metric values for a database server.
- [Displaying Metric History](#)  
Use the `LIST METRICHISTORY` command to display the metric history for a database server.

### 9.6.2.1 Displaying Specific Attributes for Metrics

Use the `ATTRIBUTES` keyword to list the specific attributes you want to view in the output.

This example shows the `LIST METRICHISTORY` command with the `name`, `metrictype`, and `alertState` attributes specified. The output is limited to a one hour time period. A portion of the output is shown in following the command.

```
LIST METRICHISTORY ATTRIBUTES name, metrictype, alertState -
WHERE collectionTime > '2019-07-12T13:00:00-05:00' -
AND collectionTime < '2019-07-12T14:00:00-05:00'
 DS_TEMP Instantaneous normal
 DS_FANS Instantaneous normal
 DS_BBU_CHARGE Instantaneous normal
 DS_BBU_TEMP Instantaneous normal
 DS_CPUT Instantaneous normal
 DS_CPUT_MS Instantaneous normal
 DS_FSUT Instantaneous normal
 DS_FSUT Instantaneous normal
 DS_FSUT Instantaneous normal
 DS_MEMUT Instantaneous normal
 DS_MEMUT_MS Instantaneous normal
 DS_RUNQ Instantaneous normal
 DS_SWAP_IN_BY_SEC Instantaneous normal
 DS_SWAP_OUT_BY_SEC Instantaneous normal
 DS_SWAP_USAGE Instantaneous normal
 DS_VIRTMEM_MS Instantaneous normal
 N_HCA_MB_RCV_SEC Rate normal
 N_HCA_MB_TRANS_SEC Rate normal
 N_IB_MB_RCV_SEC Rate normal
 N_IB_MB_RCV_SEC Rate normal
 N_IB_MB_RCV_SEC Rate normal
 ...
```

### 9.6.2.2 Displaying Metric Definitions

Use the `LIST METRICDEFINITION` command to display the metric definitions for a database server.

A metric definition listing shows the configuration of a metric. [Example 9-1](#) shows how to display attributes for the `METRICDEFINITION` object.

#### Example 9-1 Displaying Metric Definitions

```
DBMCLI> LIST METRICDEFINITION DS_CPUT DETAIL
```

```
DBMCLI> LIST METRICDEFINITION WHERE name LIKE 'DS_SWAP_*' -
 ATTRIBUTES name, metricType, description
```

#### Related Topics

- [DESCRIBE METRICDEFINITION](#)
- [LIST METRICDEFINITION](#)

### 9.6.2.3 Displaying Current Metrics

Use the `LIST METRICCURRENT` command to display the current metric values for a database server.

A current metric listing shows a set of observations on the current value of an individual metric. [Example 9-2](#) shows how to display attributes for the `METRICCURRENT` object.

#### Example 9-2 Displaying Current Metric Values

```
DBMCLI> LIST METRICCURRENT DS_FANS DETAIL

DBMCLI> LIST METRICCURRENT WHERE objectType = 'DBSERVER' AND
-
 metricValue != 0 ATTRIBUTES name,
metricObjectName, -
 metricValue, collectionTime
```

#### Related Topics

- [DESCRIBE METRICCURRENT](#)
- [LIST METRICCURRENT](#)

### 9.6.2.4 Displaying Metric History

Use the `LIST METRICHISTORY` command to display the metric history for a database server.

A metric history listing shows a collection of past individual observations of all metric values. [Example 9-3](#) shows how to display attributes for the `METRICHISTORY` object.

The retention period for metric history files is specified by the `metricHistoryDays` attribute. You can modify this setting with the DBMCLI `ALTER DBSERVER` command.

#### Example 9-3 Displaying Metric History Values

```
DBMCLI> LIST METRICHISTORY DS_TEMP WHERE alertState='critical' DETAIL

DBMCLI> LIST METRICHISTORY WHERE objectType = 'DBSERVER' AND metricValue !=
0 -
 AND collectionTime > '2014-08-12T09:10:51-07:00' -ATTRIBUTES
name, -
 metricObjectName, metricValue, collectionTime
```

### Related Topics

- [LIST METRICHISTORY](#)
- [ALTER DBSERVER](#)  
The `ALTER DBSERVER` command changes the attributes of the database server.

## 9.6.3 Database Server Metrics

Database server metrics provide information about the server, such as CPU utilization, memory or network interconnections.

To display database server metrics, use an object type attribute equal to `DBSERVER` with the `METRICCURRENT`, `METRICDEFINITION`, and `METRICHISTORY` objects. [Example 9-4](#) shows how to display database server metrics.

The following list describes the metric names and descriptions for database server metrics. The database server for the metric is specified by the `metricObjectName` attribute of the `METRICCURRENT` and `METRICHISTORY` objects.

- `DS_BBU_CHARGE`: Disk controller battery charge.
- `DS_BBU_TEMP`: Temperature of disk controller battery.
- `DS_CPUT`: The server CPU utilization which is the instantaneous percentage of time over the previous minute that the system CPUs were not idle (from `/proc/stat`).
- `DS_CPUT_MS`: The percentage of CPU time used by MS.
- `DS_FANS`: The instantaneous number of working fans on the server.
- `DS_FSUT`: The percentage of total space utilized on the file system that is currently in use. This metric shows the space utilization in the various files systems on the server.
- `DS_MEMUT`: The percentage of total physical memory used on the server.
- `DS_MEMUT_MS`: The percentage of physical memory used by MS.
- `DS_RUNQ`: The instantaneous average number (over the preceding minute) of processes in the Linux run queue marked running or uninterruptible (from `/proc/loadavg`).
- `DS_SWAP_IN_BY_SEC`: The number of swap pages read in KB per second.
- `DS_SWAP_OUT_BY_SEC`: The number of swap pages written in KB per second.
- `DS_SWAP_USAGE`: The percentage of swap space used.
- `DS_TEMP`: The instantaneous temperature (Celsius) of the server, provided by the Baseboard Management Controller (BMC).
- `DS_VIRTMEM_MS`: The amount of virtual memory used by MS in megabytes.
- `N_NIC_KB_RCV_SEC`: The number of kilobytes received by the Ethernet interfaces per second.
- `N_NIC_KB_TRANS_SEC`: The number of kilobytes transmitted by the Ethernet interfaces per second
- InfiniBand Network Fabric metrics, not applicable to Exadata Database Machine servers with RoCE Network Fabric.:

- `N_HCA_MB_RCV_SEC`: The number of megabytes received by the InfiniBand interfaces per second.
- `N_HCA_MB_TRANS_SEC`: The number of megabytes transmitted by the InfiniBand interfaces per second.
- `N_IB_MB_RCV_SEC`: The number of megabytes received by an InfiniBand port per second.
- `N_IB_MB_TRANS_SEC`: The number of megabytes transmitted by an InfiniBand port per second.
- `N_IB_UTIL_RCV`: The percentage of an InfiniBand port capacity utilized for receiving data.
- `N_IB_UTIL_TRANS`: The percentage of an InfiniBand port capacity utilized for transmitting data.

#### Example 9-4 Displaying Database Server Metrics

```
CellCLI> LIST METRICCURRENT DS_CPU_T DETAIL
```

```
name: DS_CPU_T
alertState: normal
collectionTime: 2014-12-17T15:54:25-08:00
metricObjectName: abcd2x3
metricType: Instantaneous
metricValue: 6.7 %
objectType: DBSERVER
```

## 9.6.4 Real-Time Insight

You can use the real-time insight feature to enable real-time monitoring of your Exadata systems.

Commencing with Oracle Exadata System Software 22.1.0, real-time insight provides infrastructure to:

- Categorize specific metrics as fine-grained, and enable the collection of fine-grained metrics as often as every second.
- Stream fine-grained metric observations to user-defined locations in real-time, using either push (upload) or pull (download) transmission models.

#### Related Topics

- [Using Real-Time Insight](#)  
You can use the real-time insight feature to enable real-time monitoring of your Exadata systems using an external metric collection platform.

## 9.6.5 Using Real-Time Insight

You can use the real-time insight feature to enable real-time monitoring of your Exadata systems using an external metric collection platform.

- [Using Fine-Grained Metrics](#)
- [Uploading Metric Observations to a Collection Platform](#)

- [Downloading Metric Observations from an Exadata Server](#)
- [Tagging Metric Observations](#)
- [Understanding the Metric Stream Format](#)

### 9.6.5.1 Using Fine-Grained Metrics

Traditionally, Exadata metric collections occur at 1-minute intervals. However, real-time monitoring requires more timely metrics. Commencing with Oracle Exadata System Software 22.1.0, you can configure fine-grained metric collection.

Fine-grained metric collection is the foundation for real-time insight. Fine-grained metric collection works in conjunction with an external metric collection platform, where the fine-grained metric observations are collected and processed for visualization. Fine-grained metric observations reside only in server memory and are not persisted on the server. Consequently, no fine-grained metric history is maintained on each server.

- [Controlling Fine-Grained Metric Collection](#)
- [Customizing Fine-Grained Metric Collection](#)

#### 9.6.5.1.1 Controlling Fine-Grained Metric Collection

The `metricFGCollIntvlInSec` attribute controls fine-grained metric collection.

- To enable fine-grained metric collection, you must set the collection interval to a value between 1 and 60 seconds.

For example:

```
DBMCLI> ALTER DBSERVER metricFGCollIntvlInSec=1
```

The collection interval for fine-grained metric collection is related to the automatic upload frequency specified in the `metricStreamIntvlInSec` attribute. When automatic upload is enabled, `metricStreamIntvlInSec` must be between 5 and 30 times `metricFGCollIntvlInSec`, and any change to `metricFGCollIntvlInSec` must honor this relationship. For example, if `metricStreamIntvlInSec` is set to 60, then `metricFGCollIntvlInSec` must be between 2 and 12.

- To disable fine-grained metric collection on a server, set `metricFGCollIntvlInSec=0`.

For example:

```
DBMCLI> ALTER DBSERVER metricFGCollIntvlInSec=0
```

You cannot disable fine-grained metric collection when automatic metric uploads are enabled using the `metricStreamIntvlInSec` attribute.

#### 9.6.5.1.2 Customizing Fine-Grained Metric Collection

By default, a set of key performance metrics is automatically enabled for fine-grained collection. But, you can customize fine-grained metric collection by enabling or disabling specific metrics.



- To enable a metric for fine-grained collection, use the `ALTER METRICDEFINITION` command and specify `finegrained=enabled`. For example:

```
DBMCLI> ALTER METRICDEFINITION N_NIC_KB_TRANS_SEC finegrained=enabled
```

```
DBMCLI> ALTER METRICDEFINITION N_MB_SENT,N_MB_RECEIVED finegrained=enabled
```

```
DBMCLI> ALTER METRICDEFINITION finegrained=enabled WHERE name LIKE
'N_NIC.*'
```

- To disable a metric for fine-grained collection, use the `ALTER METRICDEFINITION` command and specify `finegrained=disabled`. For example:

```
DBMCLI> ALTER METRICDEFINITION N_MB_SENT finegrained=disabled
```

- At any time, to view the metrics enabled for fine-grained collection, use the following command:

```
DBMCLI> LIST METRICDEFINITION WHERE finegrained=enabled
```

- At any time, to view all of the metric definition details, including each metric description and whether the metric is enabled for fine-grained collection, use the following command:

```
DBMCLI> LIST METRICDEFINITION DETAIL
```

## 9.6.5.2 Uploading Metric Observations to a Collection Platform

You can enable an Exadata server to automatically upload (push) fine-grained metric observations to an external metric collection platform.

- [Controlling Automatic Metric Uploads](#)
- [Configuring the Endpoints for Automatic Metric Upload](#)

### 9.6.5.2.1 Controlling Automatic Metric Uploads

The `metricStreamIntvlInSec` attribute sets the upload interval (in seconds) for automatic uploads to the metric streaming endpoints specified by the `metricStreamEndPoint` attribute.

- To enable automatic upload, the `metricStreamIntvlInSec` value must be between 5 and 30 times the collection interval specified in the `metricFGCollIntvlInSec` attribute. For example, if `metricFGCollIntvlInSec` is set to 5, then `metricStreamIntvlInSec` must be between 25 and 150.

For example:

```
DBMCLI> ALTER DBSERVER metricStreamIntvlInSec=25
```

You cannot enable automatic upload when fine-grained metric collection is disabled (`metricFGCollIntvlInSec=0`).

- To disable automatic metric uploads on a server, set `metricStreamIntvlInSec=0`.

For example:

```
DBMCLI> ALTER DBSERVER metricStreamIntvlInSec=0
```

### 9.6.5.2.2 Configuring the Endpoints for Automatic Metric Upload

You can automatically upload (push) the metric stream to one or more collection endpoints by setting the `metricStreamEndPoint` attribute as follows:

```
metricStreamEndPoint[+]=((host="endpoint-URL"[,type="stream-format"]
[,token="authentication-token"][, {httpProxy|httpsProxy}="proxy-
server"])
 [, (host="endpoint-URL"[,type="stream-format"]
[,token="authentication-token"][, {httpProxy|httpsProxy}="proxy-
server"])]...)
```

In the `metricStreamEndPoint` definition:

- `host`: Specifies the URL for the collection endpoint. The URL can use HTTP or HTTPS.
- `type`: Optionally specifies the format of the stream. Supported values are:
  - `json`: Provides the stream in a JSON format
  - `plaintext`: Provides the stream in a plain text format
 The default value is `json`.
- `token`: Optionally specifies the authentication token for the collection endpoint. Consult the metric collection platform for details about generating the token.
- `httpProxy` or `httpsProxy`: Optionally specifies a proxy server to facilitate network connectivity to the collection endpoint. A proxy server is required if a firewall resides between the Exadata system and the collection endpoint.

You can use the optional `+=` operator to add collection endpoints to an existing `metricStreamEndPoint` definition. Otherwise, the `=` operator overwrites the previous attribute value.

#### Example 9-5 Setting Up a JSON Stream

This example shows how to set up a JSON stream. In the example command, the `host` and `token` values come from the collection platform and we assume that network connectivity is through the specified proxy.

```
DBMCLI> ALTER DBSERVER metricStreamEndPoint=((host="https://
ingest.stream.example.com/v2/
datapoint",type="json",token="wcfA_*****Z58QpKg",httpProxy="www-
proxy.example.com:80"))
```

#### Example 9-6 Adding a Plain Text Endpoint

This example shows how to add a plain text endpoint to the existing `metricStreamEndPoint`. In the example command, the `host` and `token` values come

from the collection platform. The example also assumes that the collection platform is within the corporate network requiring no network proxy.

```
DBMCLI> ALTER DBSERVER metricStreamEndPoint+=((host="http://
idbsrv.example.com:8086/api/v2/write?
org=Exadata&bucket=Metrics&precision=ns",type="plaintext",token="6unif*****
***rOXwtfkG0gWGENyePd6uN6OLR_deTZL4IuG9VTfDWwvpB-
QvJcCcFs_NVjmpsyANz0Q8psA=="))
```

### 9.6.5.3 Downloading Metric Observations from an Exadata Server

You can download (pull) the metrics stream from an Exadata server by using the provided REST endpoint.

On each storage server, the REST endpoint URL is:

```
https://server-name/metricstream/list?stream=true
```

On each database server, the endpoint uses port 7879. Consequently, the URL is:

```
https://server-name:7879/metricstream/list?stream=true
```

To maximize efficiency, match the download frequency to the collection interval specified in the `metricFGCollIntvlInSec` attribute on the storage server or database server.

To facilitate access to the metric stream, you should use a dedicated user account, which only has access to the stream. You can use the following command sequence to appropriately configure a user account in DBMCLI, which you can then use for authentication to the REST endpoint. In the command sequence, substitute your own user and role names.

```
CREATE ROLE metric_collector_role
GRANT PRIVILEGE LIST ON METRICSTREAM ALL ATTRIBUTES WITH ALL OPTIONS TO ROLE
metric_collector_role
CREATE USER metric_collector PASSWORD=<password>
GRANT ROLE metric_collector_role TO USER metric_collector
```

### 9.6.5.4 Tagging Metric Observations

In each Exadata database server and storage server you can define a set of metric tags, which are included in every metric observation generated by the server. These tags can help you to organize and group observations generated by numerous Exadata servers.

You can configure metric tags by setting the `metricStreamTags` attribute to a valid JSON string containing tag and value pairs as follows:

```
metricStreamTags='{ "tag1": "value1" [, "tag2": "value2"] ... }'
```

For example:

```
DBMCLI> ALTER DBSERVER
metricStreamTags='{ "application": "personnel", "department": "HR" }'
```

### 9.6.5.5 Understanding the Metric Stream Format

Real-time metric observations contain a core set of attributes. However, the format of the metric stream depends on the mode of access.

If you are automatically uploading the metric stream to a metric collection platform, you can specify the metric stream format in the metric endpoint configuration. The available formats are JSON or plain text.

Following is an example of the JSON format:

```
json: {
 "gauge": [{
 "metric": "OS_NET_RX_BY_SEC",
 "value": "0.0012989044189453125",
 "timestamp": 1652473286000,
 "unit": "MB/sec",
 "dimensions": {
 "server": "celadm09.example.com",
 "objectName": "eth0",
 "nodeType": "STORAGE",
 "fleet": "example-fleet",
 "pod": "dbm01",
 "cluster": "c01"
 }
 }, {
 "metric": "SIO_IO_RD_FC_HD_SEC",
 "value": "0.0",
 "timestamp": 1652473286000,
 "unit": "MB/sec",
 "dimensions": {
 "server": "celadm09.example.com",
 "objectName": "SMARTIO",
 "nodeType": "STORAGE",
 "fleet": "example-fleet",
 "pod": "dbm01",
 "cluster": "c01"
 }
 }
]
}
```

The plain text format contains essentially the same information as the JSON stream. However, with the plain text format, each metric observation is presented on a separate line. Following is an example of the plain text format:

```
metrics,name=OS_NET_RX_BY_SEC,objectName=eth0,server=celadm09.example.c
om,unit=MB/sec,nodeType=STORAGE,fleet=example-
fleet,pod=dbm01,cluster=c01 value=9.441184615324398E-4
1652473456000000000
metrics,name=OS_NET_RX_BY_SEC,objectName=eth0,server=celadm09.example.c
om,unit=MB/sec,nodeType=STORAGE,fleet=example-
```

```
fleet,pod=dbm01,cluster=c01 value=0.002647613311980988 1652473457000000000
```

If you are downloading the metric stream by using the provided REST endpoint, the data is presented in a format similar to the plan text upload format where each metric observation is presented on a separate line. Following is an example of the download format:

```
DS_CPU{objectName="dbadm05",unit="%",server="dbadm05.example.com",nodeType="KVMHOST",fleet="example-fleet",pod="dbm01",cluster="c01"} 23.10906363831155 1652485449597
DS_MEMUT{objectName="dbadm05",unit="%",server="dbadm05.example.com",nodeType="KVMHOST",fleet="example-fleet",pod="dbm01",cluster="c01"} 99 1652485449597
DS_MEMUT_MS{objectName="dbadm05",unit="%",server="dbadm05.example.com",nodeType="KVMHOST",fleet="example-fleet",pod="dbm01",cluster="c01"} 0.12396045794483294 1652485449597
```

The following list describes the attributes contained in the metric stream:

- The metric name is identified as follows:
  - In the JSON upload format, the metric name follows the `metric` tag.
  - In the plain text upload format, the metric name is the value following `name=`.
  - In the download format, the metric name is the first element on each line, preceding the left braces (`{}`).
- The metric value is located as follows:
  - In the JSON upload format, the metric value follows the `value` tag.
  - In the plain text upload format, the metric value follows `value=`.
  - In the download format, the metric value is the second last element on each line, following the right braces (`}`).
- The time of the metric observation is located as follows:
  - In the JSON upload format, the timestamp follows the `timestamp` tag. The timestamp is expressed as the number of milliseconds ( $1 \times 10^{-3}$  sec) since January 1, 1970, 00:00:00 GMT.
  - In the plain text upload format, the timestamp is the last element on each line. The timestamp is expressed as the number of nanoseconds ( $1 \times 10^{-9}$  sec) since January 1, 1970, 00:00:00 GMT.
  - In the download format, the timestamp is the last element on each line. The timestamp is expressed as the number of milliseconds ( $1 \times 10^{-3}$  sec) since January 1, 1970, 00:00:00 GMT.
- The `unit` value describes the unit of measurement for the metric observation.
- The `server` value contains the name of the Exadata server that generated the metric observation.
- The `objectName` value contains the name of the Exadata object associated with the metric.
- The `nodeType` value contains the type of the Exadata server that generated the metric observation.

- The `fleet`, `pod`, and `cluster` attributes are examples of user-defined metric tags, which you can use to organize and group observations generated by numerous Exadata servers. You can tag metrics by setting the `metricStreamTags` DBSERVER attribute.

#### Related Topics

- [Tagging Metric Observations](#)

## 9.7 DBMCLI Command Reference

DBMCLI has both administrative and object commands.

The following commands are available with the DBMCLI utility:

- ALTER
- CREATE
- DESCRIBE
- DROP
- GRANT
- HELP
- LIST
- REVOKE
- SET
- SPOOL
- START

### 9.7.1 ALTER

#### Purpose

The `ALTER` command performs an action on or changes attributes of a single database server object or multiple objects. The `ALTER` command can be used to change an attribute or to take an action upon the object.

#### Syntax

```
ALTER { object_type object_name [, object_name]... operation
 | attribute_name = attribute_value
 [, attribute_name = attribute_value]...
 }
```

#### Usage Notes

The following arguments can be used with the command:

- *object\_type* can be the following:
  - ALERTHISTORY
  - DBSERVER

- IBPORT (not available in user domain (domU) environments)
- THRESHOLD (not available in user domain (domU) environments)
- USER (not available in user domain (domU) environments)
- *object\_name* is the name or identifier of an alert history, database server, LUN, or threshold.
- *operation* is a valid action for the specified *object\_type*.

When multiple objects are the target of an `ALTER` command, there is the possibility of partial success. If an error occurs, then the command is interrupted, and the remaining objects are not changed.

- [ALTER ALERTHISTORY](#)
- [ALTER DBSERVER](#)  
The `ALTER DBSERVER` command changes the attributes of the database server.
- [ALTER IBPORT](#)
- [ALTER METRICDEFINITION](#)
- [ALTER SOFTWAREUPDATE](#)  
Software updates can be scheduled for database servers by setting the `SoftwareUpdate` attributes starting with Oracle Exadata System Software release 19.3.0 or later.
- [ALTER THRESHOLD](#)
- [ALTER USER](#)

#### Related Topics

- [About DBMCLI Object Types](#)

### 9.7.1.1 ALTER ALERTHISTORY

#### Purpose

The `ALTER ALERTHISTORY` command changes the attributes of all or specified alert histories.

#### Syntax

```
ALTER ALERTHISTORY { ALL | alertid1 [,alertid2 ...]}
 examinedBy=user_name
```

#### Usage Notes

The following arguments can be used with the command:

- *alertidn*: The identifier of the alerts to be changed.
- *user\_name*: The name of the user who acknowledged the alert.

#### Examples

[Example 9-7](#) shows the `ALTER` command used with the `ALERTHISTORY` object to update the `examinedBy` attribute. The `examinedBy` attribute is the only `ALERTHISTORY` attribute that can be modified.

**Example 9-7 Altering ALERTHISTORY Attributes**

```
DBMCLI> ALTER ALERTHISTORY 1671443714 -
 examinedBy="jdoe"

DBMCLI> ALTER ALERTHISTORY ALL examinedBy="jdoe"
```

## 9.7.1.2 ALTER DBSERVER

The ALTER DBSERVER command changes the attributes of the database server.

**Syntax**

```
ALTER DBSERVER {
 | SHUTDOWN SERVICES { RS | MS | ALL }
 | RESTART SERVICES { RS | MS | ALL }
 | RESTART BMC
 | STARTUP SERVICES { RS | MS | ALL }
 | LED { ON | OFF }
 | VALIDATE { MAIL | SNMP | CONFIGURATION }
 | VALIDATE SYSLOGCONF facility.priority
 | CONFIGUREBMC
 | { snmpuser=((user_clause) [, (user_clause) [, ..]]) |
snmpuser.name=(user_clause) }
 | attribute_name = attribute_value
 [, attribute_name = attribute_value]. . .
}
```

**Usage Notes**

The following table lists the arguments and options for the ALTER DBSERVER command:

Argument	Options	Description
SHUTDOWN SERVICES	MS (Management Server) RS (Restart Server) ALL	The MS service is shut down. The RS service is shut down. The ALL option shuts down the MS and RS services.
RESTART SERVICES	MS RS ALL	The MS service is shut down, and then started. The RS service is shut down, and then started. The ALL option shuts down, then starts up, the MS and RS services.
RESTART BMC	n/a	Restarts the Baseboard Management Controller (BMC).
STARTUP SERVICES	MS RS ALL	The MS service is started. The RS service is started. The ALL option starts up the MS and RS services.



Argument	Options	Description
LED	ON OFF	LED ON and LED OFF operations turn on and off the chassis LED.  You can manually set the LED to light to indicate that a database server requires maintenance. The LED is also set to light automatically if a component fails.
VALIDATE	MAIL SNMP CONFIGURATION	The VALIDATE MAIL operation sends a test message using the email attributes configured for the database server.  The VALIDATE SNMP operation sends a test message using the SNMP attributes configured for the database server. The VALIDATE SNMP TYPE=ASR operation validates ASR on Exadata Storage Server.  The VALIDATE CONFIGURATION operation validates the server configuration.
VALIDATE SYSLOGCONF	<i>facility.priority</i>	The VALIDATE SYSLOGCONF <i>facility.priority</i> sends a test message for the specified facility and priority.
CONFIGUREBMC	n/a	Configures the BMC for hardware alerts to the local database server so that Management Server can pick up the alerts.

Argument	Options	Description
snmpuser or snmpuser.name	<i>user_clause</i>	<p>Defines the user who receives SNMP alerts. You can either supply a list of user clauses for all SNMP users, or individual user clauses using the attribute <code>snmpuser.name</code>. The <i>user_clause</i> is of the form:</p> <pre>([name=user1,] authProtocol=auth_type ' authPassword=*, \ privProtocol=priv_type ' privPassword=*) ,  \ (name=user2, authProtocol=auth_type ' authPassword=*, \ privProtocol=priv_type , privPassword=*, )</pre> <p>If updating a single user, then do not include the phrase <code>name=user1</code>, in the <i>user_clause</i>.</p>
attributes	N/A	<p>The attributes that can be changed using the <code>ALTER DBSERVER</code> command are shown as modifiable in <a href="#">Example 9-43</a>.</p>



**Note:**

In user domain (domU) environments, the following commands are not permitted:

- ALTER DBSERVER CONFIGUREBMC
- ALTER DBSERVER RESTART BMC
- ALTER DBSERVER LED { ON | OFF }

The following are additional usage notes for the `ALTER DBSERVER` command:

- It may be necessary to restart, shut down, or start up a database server for the following reasons:

- Software upgrades
- Service outages that include any condition under which a database server is not responding to service requests
- To set up the database server to send notifications about alerts, you can configure the following database server attributes:
  - mailServer
  - smtpPort
  - smtpUseSSL
  - smtpFrom
  - smtpFromAddr
  - smtpToAddr
  - snmpSubscriber
  - snmpUser
  - snmpEngineID
  - notificationMethod
  - notificationPolicy
  - emailSubscriber

The `smtpToAddr` attribute can be set to a list of comma-delimited email addresses that are the recipients of the alert notification. The list must be enclosed in quotation marks.

The `snmpSubscriber` attribute can be set to a list of SNMP targets to which the SNMP alert notification is sent. These targets are specified as follows:

```
snmpSubscriber=(
 (host=host[,port=port][,community=community][,type=user_type]
 [,fromIP="ip"]
 [,asrmPort="ASRManager_port"])
 [, (host=host[,port=port][,community=community][,type=user_type]
 [,fromIP="ip"]
 [,asrmPort="ASRManager_port"])] ...)
```

 **Note:**

- The `ALTER DBSERVER ... snmpSubscriber` command replaces the existing value set for `snmpSubscriber`. To add to the list of SNMP targets ensure that you include the existing values in the command. The following message is displayed after running the command:

```
snmpSubscriber {0} has been replaced with {1}.
```

- The `snmpSubscriber` attribute is used to configure Oracle Auto Service Request (ASR).
- The SNMP alerts conform to MIBs (management information base) which are included with each installation of Oracle Exadata System Software. The MIB files on the database server are available at `/opt/oracle/dbserver/dbms/deploy/config/`.

Type ASR and v3ASR alerts conform to SUN-HW-TRAP-MIB, which is available in the `SUB-HW-TRAP-MIB.mib` file.

The other alert types conform to the `cell_alert` MIB, which is available in the `cell_alert.mib` file.

The SNMP alerts and MIB conform to SNMP version 1 (SNMPv1). However, the Auto Service Request alerts conform to SNMP version 2c (SNMPv2c).

The default value for `port` is 162. The default value for `community` is `public`.

The `notificationMethod` attribute value can be `mail`, `snmp`, `none`, or a combination of `mail` and `snmp`, such as `notificationMethod='mail,snmp'`. The default value is `mail`.

The `snmpSubscriber` types are `ASR`, `v3`, and `v3ASR`. For types `v3` and `v3ASR`, a `snmpUser` must be defined, and the user name is provided instead of `community`.

For the `v3ASR` type, the user must be defined with `authProtocol=SHA`, and `privProtocol=AES`. These are the only protocols supported by ASR Manager. Setting the `snmpSubscriber` as type `v3ASR` also sets the ILOM properties and rules for traps sent by ILOM.

If `type` is not specified, the default is version 1, `cell_alert` traps. There is no string to specify this type. To use this type, just omit the `type` field.

The `fromIP` field enables you to specify an IP address from which the trap is sent. If this field is not specified, it defaults to the IP address associated with `eth0`. Use this field if the default IP address is not registered with Oracle ASR Manager. Oracle ASR Manager only processes SNMP traps that are sent from IP addresses that it recognizes.

The `fromIP` field is allowed only for `snmpSubscriber` with `type` set to either `ASR` or `v3ASR`.

For example:

```
DBMCLI> alter dbserver
snmpSubscriber=((host=asrhost,port=162,community=public,fromIP="1.1.1.1",t
ype=ASR))
```

The following example returns an error because the type is not ASR or v3ASR.

```
DBMCLI> alter dbserver
snmpSubscriber=((host=localhost,port=162,community=public,fromIP="1.1.1.1"
))
DBM-00068: The fromIP field is only supported for ASR SNMP subscribers.
```

The `asrmPort` field enables you to specify the port number on an Oracle ASR Manager machine that MS uses to communicate with Oracle ASR Manager. This port must be the same as the HTTP port of Oracle ASR Manager's HTTP Receiver. You can check this by running `asr show_http_receiver` on the Oracle ASR Manager machine.

The `asrmPort` field is allowed only for `snmpSubscriber` with type set to either ASR or v3ASR. The default value for this port is 16161.

The `snmpUser` attribute defines the user who receives SNMP alerts. This command can only be run in interactive mode. There are two methods for configuring this attribute.

```
snmpuser=((user_clause) [, (user_clause) [, ...]])
snmpuser.name=(user_clause)
```

- If you specify `snmpuser`, then you must provide a `user_clause` for every configured user. If you omit a user, then that user will no longer receive SNMP alerts. The `((user_clause) [, (user_clause) [, ...]])` string that you provide overwrites the previous string used for the `snmpuser` attribute.
- If you specify `snmpuser.name`, then you must provide a `user_clause` for only the specified user. This allows you to add, delete, or modify each user individually, without having to supply the entire `snmpuser` attribute string each time.
- If you use `snmpuser=''`, then all SNMP users are removed. If you use `snmpuser.name=''`, then only the specified user is removed. You cannot remove an SNMP user while it is still referenced by a V3 `SnmpSubscriber`.

Each method uses a `user_clause`, which has the following format:

```
(([name=user1,] authProtocol=auth_type, authPassword=*, \
privProtocol=priv_type, privPassword=*) , \
(name=user2, authProtocol=auth_type, authPassword=*, \
privProtocol=priv_type, privPassword=*,), ...)
```

If updating a single user, then do not include the phrase `name=user1`, in the `user_clause`, because you have already supplied the name as part of `snmpuser.name`.

- `name` is the user name.

- Only `*` is allowed for the password values in the command. Passwords are not stored or displayed. Secure hash keys are computed and used for trap authentication and encryption.

- `authProtocol` is the authentication protocol. Options are `MD5` or `SHA`.

The `authProtocol` must be specified for the `snmpUser` attribute.

The system prompts for the authentication password. The authentication password must have 8 to 12 alphanumeric characters.

- `privProtocol` is encryption protocol. Options are `none`, `AES`, or `DES`. The default is `none` when the `privProtocol` attribute is not specified.

The system prompts for an encryption password if the encryption protocol is specified. The password is exactly 8 alphanumeric characters, and they are case sensitive.

The `smtpUseSSL` attribute enables Secure Socket Layer (SSL) encryption on the email notifications when the attribute is set to `true`.

The `notificationPolicy` attribute value can be `none` or a combination of `critical`, `warning`, or `clear`, such as `notificationPolicy='warning,clear.'`

- The `critical` value refers to hardware-generated alerts or alerts generated by Automatic Diagnostic Repository (ADR) or Baseboard Management Controller (BMC). The `critical` value also refers to a metric alert when the value exceeds the critical threshold specified in the metric definition.
- The `warning` value refers to a metric alert when the value exceeds the warning threshold specified in the metric definition.
- The `clear` value refers to a metric alert when the value is below the threshold boundary after having previously exceeded a warning or critical threshold.
- The `maintenance` value refers to all hardware-related errors. The hardware errors are reported as "Maintenance" in email message subject lines.
- For each subscriber, the host must be specified as either a domain name or an IP address. Enclose the host name or IP address in quotation marks if it contains non-alphanumeric characters. Port and community values are optional. The default port value is `162`. The default community value is `public`. The type value is optional. The default value for type is `NULL`. The types `ASR`, `V3`, and `v3ASR` are the only supported non-NULL value.
- After startup of the Management Server (MS), the `snmpSubscriber` list entries with type `ASR` or `v3ASR` are added to the ILOM for the database server. This ensures that when an ILOM is replaced, the entries are set for the new ILOM. If the entries are removed from the ILOM, then they must be manually added to the ILOM using the `ALTER DBSERVER ... snmpUser=` command.
- The `snmpSubscriber` with `type=asr` or `type=v3ASR` should only be configured to point to Oracle ASR Manager.
- To validate that email messages are successfully sent for database server alerts or events, use the `ALTER` command with the `VALIDATE MAIL` option. The validation process sends a test email message to the configured recipient. If that test email message is not received, then an email configuration setting is not valid.
- The `emailFormat` attribute can be `html` or `text`. By default, email notifications are sent in HTML format. Change the value to `text` to receive plain text email notifications.

- The `ALTER DBSERVER snmpEngineID` command is used by the SNMP managers to subscribe to alerts from the database servers. The `snmpEngineID` parameter can be up to 20 characters. It should be unique for each target within a data center. The default is the database server name. This default is used if the `snmpEngineID` attribute is not set before the SNMP users are defined.

The engine identifier should not be changed after SNMP users are defined. Any change to an engine identifier causes the user keys to be re-computed, and user passwords must be re-entered.

- If the database server name is changed, then you must choose a unique database server name.
- If an `ipaddressN` attribute is modified, then the network configuration file `/etc/oracle/cell/network-config/cellinit.ora` is modified.
- The `ALTER DBSERVER snmpSubscriber` command configures the Oracle ASR subscriber, and sends traps.
- The `ALTER DBSERVER emailSubscriber` command sets a list of comma-delimited email addresses that are the recipients of alert notifications for specific alert types. The following is an example of the syntax:

```
ALTER DBSERVER emailSubscriber = ((email="email_address1",
\
 alertType="alert_type")
[, (email="email_address2",alertType="alert_type"), ...])
```

The email address must be a valid email address. The `email` parameter is mandatory. The `alertType` parameter specifies the type of alert, and is optional. The alert types are `HARDWARE`, `SOFTWARE`, `METRIC` or `ADR`. If the alert type is not specified, then the subscription is for all alert types.

An empty input string removes the current set of subscribers.

The notification policy must be set before alert notifications can be received. The policy applies to all email subscribers. The notification policy for these alerts are the same as for `snmpSubscriber` alerts.

- The `syslogconf` attribute extends syslog rules for a database server. The attribute can be used to designate that syslog messages be forwarded to a specified management server. On the management server, the forwarded messages are directed to a file, console, or management application, depending on the syslog configuration on the management server. The following shows the syntax for the attribute:

```
syslogconf = ('selector @node' [, 'selector @node']...)
```

In the preceding syntax, `selector` is the message type, and `node` is the specified server. Both variables follow `syslog.conf` standard syntax rules.

- The `facility` option for the `syslogconf` attribute must be one of the following: `auth`, `authpriv`, `cron`, `daemon`, `ftp`, `kern`, `lpr`, `mail`, `mark`, `news`, `security`, `syslog`, `user`, `uucp`, `local0`, `local1`, `local2`, `local3`, `local4`, `local5`, `local6`, `local7`, `none`, or `*`.
- The `priority` option for the `syslogconf` attribute must be one of the following: `alert`, `crit`, `debug`, `emerg`, `err`, `error`, `info`, `notice`, `panic`, `warn`, `warning`, `none`, or `*` (asterisk).

- The `ALTER DBSERVER VALIDATE syslogconf selector` command sends a test log message. The test message is directed as specified by rules in the `/etc/syslog.conf` file. If the `syslogconf` assignment extends the `syslog` rules, then a test message is forwarded to the specified management servers.
- Starting with Oracle Exadata System Software release 19.1.0, you can use the `syslogFormat` attribute to change the standard format for `syslog` to any format by setting the value to the desired format string. Setting the `syslogFormat` attribute to an empty string removes the format change, reverting the `syslog` format to the default format. If the format string contains a control character, it must be preceded by a backslash when entering the command.

See [Example 9-24](#) for examples.

- Starting with Oracle Exadata System Software release 19.3.0, you can use the `syslogFormat` attribute to enable sending `syslog` in an encrypted format. For the complete configuration steps, refer to [Encrypting System Log Information](#).
- You can turn off the diagnostic pack attachment to emails by running the following command:

```
ALTER DBSERVER diagPackEmailAttach=FALSE
```

- Starting with Oracle Exadata System Software release 19.1.0, the `httpsAccess` attribute can be used to specify a list of IP addresses or IP subnet masks that control who can access the RESTful service via HTTPs. The value you specify for `httpsAccess` overwrites any previous value. You can use the following values for `httpsAccess`:
  - ALL — to allow access to all hosts (Default)
  - NONE — to disable the HTTPs port completely
  - IP1, IP2, ..., IPn — to only allow access to hosts with IP addresses IP1, IP2, ..., IPn where IPn is a valid IP address in IPv4, IPv4 subnet, IPv6 or IPv4-embedded IPv6 format. You can specify a maximum of 512 IP addresses for the access control list.

Additionally, instead of a single IP address, you can use the `/` character to specify a range of IP addresses using a subnet mask. For example the range `'192.168.10.0/24'` corresponds to hosts having IP addresses from 192.168.10.1 to 192.168.10.255. If you specify an IP address range, you need to enclose the IP address string in quotes.

- To set up CA-certified security certificates on the cell for use with ExaCLI, use the following attributes:

 **Note:**

The following attributes can be used only if you are running the `ALTER CELL` command from ExaCLI.

- `securityPubKey` - Specifies the URL to the public key file.
- `securityPrivKey` - Specifies the URL to the private key file.



- `securityPrivKeyPW` - Specifies the password to use if the private key file is encrypted.

For example:

```
ExaCLI> alter cell securityPubKey="http://www.example.com/security/
newkey.crt", -
 securityPrivKey="http://www.example.com/security/
newkey.key", -
 securityPrivKeyPW=*
password=*****
```

After you upload the CA-certified security certificate, you must restart MS before the new security certificate is visible.

```
CellCLI> alter cell restart services ms
```

#### See Also:

[Using a CA-Certified Security Certificate](#)

- Starting with Oracle Exadata System Software release 21.2.0, the `ilomSyslogClients` attribute specifies the remote destination to forward syslog messages from the Integrated Lights Out Manager (ILOM) service processor (SP).

The `ilomSyslogClients` attribute accepts a comma-separated list of up to two loghost servers. For each loghost server, you must specify a valid hostname or IP address.

For example:

```
DBMCLI> ALTER DBSERVER ilomSyslogClients="192.0.2.101,192.0.2.201"
```

#### Note:

The specified `ilomSyslogClients` must listen on port 514 to receive the ILOM syslog messages.

## Examples

[Example 9-8](#) shows how to set the `asrmPort` field for an `snmpSubscriber`.

[Example 9-9](#) shows how to enable and disable the auto diagpack upload feature.

[Example 9-10](#) shows how to set up email notifications for the database server.

[Example 9-11](#) shows how to validate the email setup on a database server.

[Example 9-12](#) shows how to change the format of email messages.

[Example 9-13](#) shows how to validate the SNMP setup on a database server.

[Example 9-14](#) shows how to specify the type of email alerts. In the example, one subscriber gets hardware and software alerts, and the other subscriber gets ADR alerts.

[Example 9-15](#) shows how to modify the SNMP user.

[Example 9-16](#) shows how to modify a single SNMP user.

[Example 9-17](#) shows how to start up and shut down database server services.

[Example 9-18](#) shows how to set the LED on the database server.

[Example 9-19](#) shows the pending core count for capacity-on-demand. The server must be restarted in order for the new core count to be active.

[Example 9-20](#) shows how to add a rule using the `syslogconf` attribute.

[Example 9-21](#) shows how to add and validate a rule with test message.

[Example 9-22](#) shows how to remove the `syslog.conf` rule.

[Example 9-23](#) shows how to restrict HTTPS Access to the Exadata RESTful service to a specific range of IP addresses.

#### **Example 9-8 Setting the `asrmPort` for an `snmpSubscriber`**

```
DBMCLI> ALTER DBSERVER
snmpSubscriber=((host=host1,port=162,community=public,type=asr,asrmPort
=16161))
```

#### **Example 9-9 Enabling/Disabling Auto Diagpack Upload**

You can enable or disable this feature by setting the `diagPackUploadEnabled` attribute on the `dbserver` object.

Set the attribute to `false` to disable this feature, `true` to enable it. The default is `true`.

```
DBMCLI> ALTER DBSERVER diagPackUploadEnabled=FALSE
```

#### **Example 9-10 Configuring Email Notifications for a Database Server**

```
DBMCLI> ALTER DBSERVER mailServer='mail_relay.example.com',
-
smtpFromAddr='john.doe@example.com', -
smtpFrom='John Doe', -
smtpToAddr='jane.smith@example.com', -
snmpSubscriber=((host=host1),(host=host2)), -
notificationPolicy='clear', -
notificationMethod='mail,snmp'
```

#### **Example 9-11 Validating Email on a Database Server**

```
DBMCLI> ALTER DBSERVER VALIDATE MAIL
```

**Example 9-12 Changing the Format of Email Messages**

```
DBMCLI> ALTER DBSERVER emailFormat='text'
DBMCLI> ALTER DBSERVER emailFormat='html'
```

**Example 9-13 Validating SNMP on a Database Server**

```
DBMCLI> ALTER DBSERVER VALIDATE SNMP
```

**Example 9-14 Specifying the Type of Email Alert**

```
ALTER DBSERVER emailSubscriber=
 ((email="email1@example.com",alertType="HARDWARE,SOFTWARE"), \
 (email="email2@example.com",alertType="ADR")) \
```

**Example 9-15 Modifying the SNMP User**

This example shows the initial configuration of a single SNMP user, where the administrator is prompted to enter the passwords.

```
CellCLI> ALTER CELL snmpuser = ((name=ASR, authprotocol=md5,
authpassword=*, \
 privprotocol=AES, privpassword=*))
snmpUser ASR authpassword: password
Confirm snmpUser ASR authpassword: password
snmpUser ASR privpassword: password
Confirm snmpUser ASR privpassword: password
```

**Example 9-16 Modifying a Single SNMP User**

The following code examples show adding an SNMP user, changing that user's password, and then removing that user.

```
adding users individually
CellCLI> ALTER CELL snmpuser.user2=(authprotocol=SHA,authpassword=*)

snmpUser user2 authpassword: password
Confirm snmpUser user2 authpassword: password

snmpUser ((name=user1, authProtocol=SHA, privProtocol=AES)) has been
replaced with
((name=user1, authProtocol=SHA, privProtocol=AES),(name=user2,
authProtocol=SHA)).
Cell cel01 successfully altered

changing a password of an existing user
CellCLI> ALTER CELL snmpuser.user2 = (authprotocol=SHA,authpassword=password)

Cell cel01 successfully altered

delete a user individually
CellCLI> ALTER CELL snmpuser.user2=''
```

```
snmpUser ((name=user1, authProtocol=SHA, privProtocol=AES),
(name=user2, authProtocol=SHA)) has
been replaced with ((name=user1, authProtocol=SHA, privProtocol=AES)).
Cell cel01 successfully altered
```

### Example 9-17 Starting Up and Shutting Down Database Server Services

```
DBMCLI> ALTER DBSERVER SHUTDOWN SERVICES MS

DBMCLI> ALTER DBSERVER RESTART SERVICES MS
```

### Example 9-18 Setting the Database Server LED Off and On

```
DBMCLI> ALTER DBSERVER LED OFF
DBMCLI> ALTER DBSERVER LED ON
```

### Example 9-19 Setting Pending Core Count

This example shows how to increase the number of active physical cores on a database server before restarting the server. Use the `LIST DBSERVER` command to verify the correct value is set.

```
DBMCLI> ALTER DBSERVER pendingCoreCount=20
```

### Example 9-20 Using the syslogconf Attribute

```
DBMCLI> ALTER DBSERVER syslogconf=('*.err;authpriv.none @loghost', -
'*.emerg @loghost')
```

### Example 9-21 Adding and Validating a Rule

```
DBMCLI> ALTER DBSERVER syslogconf=('kern.crit @loghost')
DBMCLI> ALTER DBSERVER VALIDATE syslogconf 'kern.crit'
```

### Example 9-22 Removing All syslog.conf Rules

```
DBMCLI> ALTER DBSERVER syslogconf=''
```

### Example 9-23 Restricting HTTPS Access to the Exadata RESTful Service

The following command allows HTTPS port access to hosts having IP addresses in the range from 192.168.10.1 to 192.168.10.255.

```
DBMCLI> ALTER DBSERVER httpsAccess="192.168.10.0/24"
```

### Example 9-24 Setting the Syslog Format to a Custom String Then Reverting to the Default Format

```
DBMCLI> ALTER DBSERVER syslogformat="%TIMESTAMP:::date-rfc3339%
%HOSTNAME%%syslogtag%
```

```
%syslogseverity-text%: %msg:::sp-if-no-1st-sp%%msg:::drop-last-1f%\\n"
DBMCLI> ALTER DBSERVER syslogformat="%TIMESTAMP% %HOSTNAME% %msg%\\n"
DBMCLI> ALTER DBSERVER syslogformat=""
```

- [Real-Time Insight Attributes](#)

#### Related Topics

- [CREATE DIAGPACK](#)
- [Restrictions on Values of Common Attributes](#)  
Review the following restrictions for the values of attributes used by multiple DBMCLI objects.

### 9.7.1.2.1 Real-Time Insight Attributes

Commencing with Oracle Exadata System Software 22.1.0, you can use the real-time insight feature to enable real-time monitoring of your Exadata systems.

#### Fine-Grained Metric Collection

The `metricFGCollIntvlInSec` attribute controls fine-grained metric collection.

- To enable fine-grained metric collection, you must set the collection interval to a value between 1 and 60 seconds.

For example:

```
DBMCLI> ALTER DBSERVER metricFGCollIntvlInSec=1
```

The collection interval for fine-grained metric collection is related to the automatic upload frequency specified in the `metricStreamIntvlInSec` attribute. When automatic upload is enabled, `metricStreamIntvlInSec` must be between 5 and 30 times `metricFGCollIntvlInSec`, and any change to `metricFGCollIntvlInSec` must honor this relationship. For example, if `metricStreamIntvlInSec` is set to 60, then `metricFGCollIntvlInSec` must be between 2 and 12.

- To disable fine-grained metric collection on a server, set `metricFGCollIntvlInSec=0`.

For example:

```
DBMCLI> ALTER DBSERVER metricFGCollIntvlInSec=0
```

You cannot disable fine-grained metric collection when automatic metric uploads are enabled using the `metricStreamIntvlInSec` attribute.

#### Automatic Metric Upload

The `metricStreamIntvlInSec` attribute sets the upload interval (in seconds) for automatic uploads to the metric streaming endpoints specified by the `metricStreamEndPoint` attribute.

- To enable automatic upload, the `metricStreamIntvlInSec` value must be between 5 and 30 times the collection interval specified in the `metricFGCollIntvlInSec` attribute. For example, if `metricFGCollIntvlInSec` is set to 5, then `metricStreamIntvlInSec` must be between 25 and 150.

For example:

```
DBMCLI> ALTER DBSERVER metricStreamIntvlInSec=25
```

You cannot enable automatic upload when fine-grained metric collection is disabled (`metricFGCollIntvlInSec=0`).

- To disable automatic metric uploads on a server, set `metricStreamIntvlInSec=0`.

For example:

```
DBMCLI> ALTER DBSERVER metricStreamIntvlInSec=0
```

## Metric Upload Endpoints

The `metricStreamEndPoint` attribute specifies one or more collection endpoints that automatically receive the metric stream. You can set `metricStreamEndPoint` as follows:

```
metricStreamEndPoint[+]=((host="endpoint-URL" [,type="stream-format"]
[,token="authentication-token"] [, {httpProxy|httpsProxy}="proxy-
server"])
 [, (host="endpoint-URL" [,type="stream-format"]
[,token="authentication-token"] [, {httpProxy|httpsProxy}="proxy-
server"])]...)
```

In the `metricStreamEndPoint` definition:

- `host`: Specifies the URL for the collection endpoint. The URL can use HTTP or HTTPS.
- `type`: Optionally specifies the format of the stream. Supported values are:
  - `json`: Provides the stream in a JSON format
  - `plaintext`: Provides the stream in a plain text format

The default value is `json`.

- `token`: Optionally specifies the authentication token for the collection endpoint. Consult the metric collection platform for details about generating the token.
- `httpProxy` or `httpsProxy`: Optionally specifies a proxy server to facilitate network connectivity to the collection endpoint. A proxy server is required if a firewall resides between the Exadata system and the collection endpoint.

You can use the optional `+=` operator to add collection endpoints to an existing `metricStreamEndPoint` definition. Otherwise, the `=` operator overwrites the previous attribute value.

## Metric Tags

The `metricStreamTags` attribute defines a set of metric tags, which are included in every metric observation generated by the server. These tags can help you to organize and group observations generated by numerous Exadata servers.

You can set the `metricStreamTags` attribute to a valid JSON string containing tag and value pairs as follows:

```
metricStreamTags='{ "tag1": "value1" [, "tag2": "value2"] ... }'
```

For example:

```
DBMCLI> ALTER DBSERVER
metricStreamTags='{ "application": "personnel", "department": "HR" }'
```

### 9.7.1.3 ALTER IBPORT

#### Purpose

The `ALTER IBPORT` command performs an action on all InfiniBand Network Fabric ports, or specific ports.



#### Note:

This command does not work on Exadata Database Machine servers with RoCE Network Fabric.

#### Syntax

```
ALTER IBPORT {ALL | ibport_name [, ibport_name] ...} RESET COUNTERS
```

#### Usage Notes

The `RESET COUNTERS` option resets all counters on the InfiniBand Network Fabric port.

#### Example 9-25 Altering IBPORT Attributes

This example shows the how to reset the `IBPORT` counters for all ports, or for a specific port.

```
DBMCLI> ALTER IBPORT ALL RESET COUNTERS

InfiniBand Port HCA-1:1 successfully altered.
InfiniBand Port HCA-1:2 successfully altered.

DBMCLI> ALTER IBPORT "HCA-1:1" RESET COUNTERS

InfiniBand Port HCA-1:1 successfully altered.
```

### 9.7.1.4 ALTER METRICDEFINITION

#### Purpose

The `ALTER METRICDEFINITION` command enables and disables metrics for fine-grained collection.

## Syntax

```
ALTER METRICDEFINITION [metric_name [, metric_name]...]
finegrained={enabled|disabled} [attribute_filters]
```

## Usage Notes

- In the command:
  - *metric\_name*: Identifies a specific metric definition to alter.
  - *attribute\_filters*: Identifies the metric definitions to alter by filtering according to attribute values.
- By default, a set of key performance metrics is automatically enabled for fine-grained collection. But, you can customize fine-grained metric collection by enabling or disabling specific metrics.
- To view the metrics enabled for fine-grained collection, use the `LIST METRICDEFINITION` command and examine the `finegrained` attribute value.

### Example 9-26 Alter a Specific Metric Definition

This example shows the command to enable fine grained metric collection for the metric that shows the Ethernet network interface transfer rate (`N_NIC_KB_TRANS_SEC`).

```
DBMCLI> ALTER METRICDEFINITION N_NIC_KB_TRANS_SEC finegrained=enabled
```

### Example 9-27 Alter a List of Metric Definitions

This example shows the command to disable fine grained metric collection for a listed set of metrics.

```
DBMCLI> ALTER METRICDEFINITION N_MB_SENT,N_MB_RECEIVED
finegrained=disabled
```

### Example 9-28 Alter Metric Definitions Using a Filter

This example shows the command to enable fine grained metric collection for all metrics specified by the attribute filter.

```
DBMCLI> ALTER METRICDEFINITION finegrained=enabled WHERE name LIKE
'N_NIC.*'
```

## Related Topics

- [Real-Time Insight](#)  
You can use the real-time insight feature to enable real-time monitoring of your Exadata systems.
- [Attribute Filters in LIST and ALTER Commands](#)  
You can use the *attribute\_filters* clause to specify the objects to display in `LIST` commands. Some `ALTER` commands also support the *attribute\_filters* clause.



## 9.7.1.5 ALTER SOFTWAREUPDATE

Software updates can be scheduled for database servers by setting the `SoftwareUpdate` attributes starting with Oracle Exadata System Software release 19.3.0 or later.

### Purpose

The `ALTER SOFTWAREUPDATE` command enables you to alter modifiable Software Update attributes, to validate the prerequisites for the software update, or to start the upgrade immediately.

You can also run the `ALTER SOFTWAREUPDATE` command using `dcli` or `exacli`.

### Note:

Currently, database node rolling upgrade automation is not supported, so you must plan the upgrades to avoid service outage downtime.

### Syntax

```
ALTER SOFTWAREUPDATE {VALIDATE PREREQ [FORCE] | UPGRADE [FORCE] | attribute
= attribute value
[,attribute = attribute value...]}
```

### Options for the ALTER SOFTWAREUPDATE Command

- `VALIDATE PREREQ [FORCE]`

Run software update check prerequisite steps now. This will download the software update prerequisite code for the update specified by the `store` attribute. These checks are run automatically as part of update. Use this option only if you want to run prerequisite checks explicitly. Any error found will be displayed. A stateful alert will be raised if any error is found by the `VALIDATE PREREQ` command.

If you include the `FORCE` option, then the `dbnodeupdate` command is run with the `-modify_on_prereq` option.

- `UPGRADE [FORCE]`

Run the software update (including the prerequisite steps) now, using the software location specified by the Software Update `store` attribute. Use this command if you want to perform the update now rather than wait for the time specified by the Software Update `time` attribute.

If `FORCE` is specified, then the upgrade continues despite any prerequisite check errors.

- `attribute = attribute value`

Modify the specified Software Update attributes to the values provided.

### Attributes for SOFTWAREUPDATE

The following attributes for the `ALTER SOFTWAREUPDATE` command are modifiable:

- `frequency`: Database server updates can automatically be done periodically by setting the `frequency` attribute to the desired frequency. You can specify one of the following values: {`none` | `daily` | `weekly` | `biweekly` }. If the value specified for `frequency` is '' or `none`, then the scheduled update is only done once.
- `name`: The name of the patch to use in the update, which includes the software version string such as 19.3.1.0.0.190918. The name can be an ARU-style file name containing the ISO file (for example, `p28802069_193000` or `p28802067_193000`), or it can be the label-based name (for example, `exadata_ovs_19.3.0.0.0.190913` or `exadata_o17_19.3.0.0.0.190913`). After downloading these ISO zip files, the database server update software is automatically downloaded from the same store location.

If `name` is not set (it has a value of `unknown`), then the latest version of the `exadata_oxx_...zip` patch will be selected from the store when `UPGRADE` or `VALIDATE PREREQ` is specified.

If `name` is not set (it has a value of `unknown`), then when you use the `ALTER SOFTWAREUPDATE UPGRADE` or `ALTER SOFTWAREUPDATE VALIDATE PREREQ` command, the software store is checked and the `name` attribute is set automatically (if not already set) to the latest available version in the software store. If there are multiple software versions at the store site, then this attribute can be used to specify which version should be used.

- `store`: A URL for the location of the software store.

If you specify a network store location using the `HTTP` or `HTTPS` protocol:

- The URL must be accessible using the management network or RDMA Network Fabric.
- MS automatically finds and downloads the required software update files.
- The local staging location for the downloaded software update files depends on the system configuration:
  - \* When Oracle Exadata Database Machine is configured without virtual machines (VMs) (otherwise known as a bare-metal configuration), the local staging location for the downloaded software update files is `/u01/swupdate`.
  - \* When Oracle Exadata Database Machine is configured with virtual machines (VMs), the local staging location for the downloaded software update files is:
    - \* `/EXAVMIMAGES/swupdate` inside the VM server (KVM host or Oracle VM management domain).
    - \* `/u01/swupdate` inside each VM (KVM guest or Oracle VM user domain).

Alternatively, you can use the `FILE` protocol to specify a URL for a local store.

When using a local store:

- MS does not need to download the patch zip files or perform the associated space checks.
- MS does not manage the local software store content. You must download the required patch files before patching and remove them afterward to free up space.

- The file URL must use one of the following formats:
  - \* `file:///localpath`
  - \* `file:/localpath`
- `time`: A future date and time at which the software update should be performed. The time can be specified as a local informal date and time, for example "1 AM, next Tuesday". If the date and time is valid then the output from setting this attribute shows the interpreted time in standard format with timezone offset, such as `2017-08-22T01:00:00-08:00`.  
If you set this attribute to the empty string "", it cancels the scheduled software update.
- `timeLimitInMinutes`: An update may wait for other database servers to complete their updates in order to preserve disk group redundancy. By default, there is no limit on the amount of time which can be spent waiting to update. This attribute may be set to a positive, maximum integer which represents the number of minutes a database server will spend waiting to update. If the update does not start within the time specified by the limit, then the update is canceled and an update alert is reported.

### Usage Notes

- Currently, only ISO file repository downloads are supported.
- Software download and the prerequisite check will begin up to a week before the scheduled update time.
- The update progress can be monitored by displaying the non-modifiable Software Update `status` attribute.
- Software updates do not occur if the upgrade software is already installed.
- You can use `dcli` or `exacli` to schedule and install updates using the `ALTER SOFTWAREUPDATE` command.
- The Software Update feature supports using HTTPS transport for software downloads. When using HTTPS, TLS certificate checks are performed by default. If the remote server's certificate cannot be validated then the following error is reported:

```
CELL-00076: An error occurred during download of software update:
source https://hostname:port is not available.
CELL-00092: The store's TLS certificate cannot be authenticated with
known CA certificates.
```

This can happen if the remote server uses a self-signed certificate or if the remote server uses a certificate signed by a certificate authority (CA) that is not included in the database server's CA store. You can use the following procedure to add a CA certificate to the database servers CA store. This is a security setup step which requires shell access as `root` on the database server.

1. Get a CA certificate that can verify the remote server. The certificate should be stored in PEM or DER file format.
2. Copy the file to the database server at this directory: `/etc/pki/ca-trust/source/anchors/`
3. Run following commands:

```
update-ca-trust enable
update-ca-trust extract
```

Use `man` to get more information about the `update-ca-trust` command.

### Examples

#### Example 9-29 Modifying the Software Update time Attribute

Modify the scheduled time of the next software update to 12:30 a.m. on December 27, 2019.

```
ALTER SOFTWAREUPDATE time="2019-12-27T00:30"
```

#### Example 9-30 Setting the Software Update store Attribute

This example shows how to set the store attribute to a location that uses HTTPS protocol.

```
ALTER SOFTWAREUPDATE store="https://my-exa-store/compute"
```

#### Example 9-31 Starting a Software Update Immediately

This example shows how to immediately start a software update using the attribute values already specified, after running a prerequisites check.

```
ALTER SOFTWAREUPDATE VALIDATE PREREQ
```

```
ALTER SOFTWAREUPDATE UPGRADE FORCE
```

## 9.7.1.6 ALTER THRESHOLD

### Purpose

The `ALTER THRESHOLD` command updates the attribute values of all thresholds or the specified thresholds.

### Syntax

```
ALTER THRESHOLD { ALL | threshold_name [, threshold_name ...] }
 attribute_name = attribute_value
 [, attribute_name = attribute_value]...
```

### Usage Notes

The attributes that can be changed with the `ALTER` command are shown as modifiable in [Example 9-53](#).

### Examples

[Example 9-32](#) shows how to alter threshold attributes.

#### Example 9-32 Altering Threshold Attributes

```
DBMCLI> ALTER THRESHOLD temp warning=30
```

```
DBMCLI> ALTER THRESHOLD ALL occurrences=3
```

**See Also:**

"CREATE" for additional information about setting threshold attributes

## 9.7.1.7 ALTER USER

### Purpose

The `ALTER USER` command changes the attributes of a user.

### Syntax

```
ALTER USER username attribute_name = attribute_value \
[, attribute_name = attribute_value, ...]
```

### Usage Notes

- The *username* cannot be `root`, `dbmadmin` or `dbmmonitor`. Those are reserved.
- The user name should be unique.
- The system prompts for a password for the new user. The password must have 12 to 40 alphanumeric characters or special characters `!@#%&^&*()` with at least one digit, one lowercase letter, and one uppercase letter. Starting with Oracle Exadata System Software release 18.1.0.0.0, the password can be 8 to 40 characters in length and can also utilize the special characters `-` and `_`.
- The new password cannot be the same as the current password for the user.

### Example 9-33 Using the ALTER USER Command

```
DBMCLI> ALTER USER scott password=TOPsecret2345
```

## 9.7.2 CREATE

### Purpose

The `CREATE` command creates a new object and assigns initial attributes to the object.

### Syntax

```
CREATE object_type [name] [attributename=attribute_value [,
attributename=attribute_value]...]
```

### Usage Notes

- *object\_type* can be as follows:
  - `DBSERVER`
  - `DIAGPACK`
  - `ROLE`

- THRESHOLD
- USER
- *name* is the name of the new object.
- *attribute\_value* is the name of a modifiable attribute for the *object\_type*.
- *attribute\_value* is the initial setting for the attribute.
- When multiple objects are valid as the target of a `CREATE` command, there is the possibility of partial success. If an error occurs, then the command is interrupted, and the remaining objects are not created.
- [CREATE DBSERVER](#)
- [CREATE DIAGPACK](#)
- [CREATE ROLE](#)
- [CREATE THRESHOLD](#)
- [CREATE USER](#)

#### Related Topics

- [About DBMCLI Object Types](#)

## 9.7.2.1 CREATE DBSERVER

### Purpose

The `CREATE DBSERVER` command creates the `DBSERVER` object type.

### Syntax

```
CREATE DBSERVER [name]
 [interconnect1=intValue1 [, interconnect2=intValue2 ...]]
 [{ , ipAddress1=ipValue1 [, ipAddress2=ipValue2 ...] }]
 [, attributeName = attributeValue ...]
```

### Usage Notes

The attributes that can be set are shown as `modifiable` in `DESCRIBE DBSERVER`.

- This command can be used to assign the ASR value to the `snmpSubscriber` attribute.
- If a name is not provided, then the default is to set the database server name to the network host name of the server with hyphens in the network name replaced with underscores. You can display the network name with the `uname -n` command. If you change the database server name, then you must choose a unique database server name.

If you provide a name, ensure that the name meets the guidelines.

- One to four interconnects can be specified. The `interconnect1` attribute must be specified if the `interconnect2` attribute is specified. The `interconnect1` and `interconnect2` attributes must be specified if `interconnect3` is specified, and so on.

- If `interconnectn` and `ipaddressn` are specified, then an `ipaddressn` entry is automatically written to the `/etc/oracle/cell/network-config/cellinit.ora` file.
- Interconnects use the RDMA Network Fabric (`ibN` or `reN`).
- Starting with Oracle Exadata System Software release 19.1.0, the `httpsAccess` attribute can be used to specify a list of IP addresses or IP subnet masks that control who can access the RESTful service via HTTPS. The value you specify for `httpsAccess` overwrites any previous value. You can use the following values for `httpsAccess`:
  - `ALL` — to allow access to all hosts (Default)
  - `NONE` — to disable the HTTPS port completely
  - `IP1, IP2, ..., IPn` — to only allow access to hosts with IP addresses `IP1, IP2, ..., IPn` where `IPn` is a valid IP address in IPv4, IPv4 subnet, IPv6 or IPv4-embedded IPv6 format. You can specify a maximum of 512 IP addresses for the access control list.

Additionally, instead of a single IP address, you can use the `/` character to specify a range of IP addresses using a subnet mask. For example the range `'192.168.10.0/24'` corresponds to hosts having IP addresses from 192.168.10.1 to 192.168.10.255. If you specify an IP address range, you need to enclose the IP address string in quotes.

## Examples

[Example 9-34](#) shows the `CREATE DBSERVER` command with the `interconnectn` attribute.

[Example 9-35](#) shows how to create a `DBSERVER` object that restricts HTTPS access to the RESTful server to a set of hosts.

### Example 9-34 Creating a DB Server and setting the `interconnectn` attribute for the `DBSERVER` object

For racks that use InfiniBand Network Fabric:

```
DBMCLI> CREATE DBSERVER interconnect1=ib0, interconnect2=ib1
```

For racks that use RoCE Network Fabric:

```
DBMCLI> CREATE DBSERVER interconnect1=re0, interconnect2=re1
```

### Example 9-35 Creating a DBServer with Restricted HTTPS Access

This example shows how to create a `DBSERVER` object that allows HTTPS port access only from hosts having IP addresses in the range 192.168.10.1 to 192.168.10.255.

For racks that use InfiniBand Network Fabric:

```
DBMCLI> CREATE DBSERVER interconnect1=ib0, httpsAccess="192.168.10.0/24"
```

For racks that use RoCE Network Fabric:

```
DBMCLI> CREATE DBSERVER interconnect1=re0, httpsAccess="192.168.10.0/24"
```

## Related Topics

- [ALTER DBSERVER](#)  
The `ALTER DBSERVER` command changes the attributes of the database server.
- [Restrictions on Values of Common Attributes](#)  
Review the following restrictions for the values of attributes used by multiple DBMCLI objects.
- [DESCRIBE DBSERVER](#)

## 9.7.2.2 CREATE DIAGPACK

### Purpose

The `CREATE DIAGPACK` command creates a diagnostic package, which contains logs and traces that you can use to troubleshoot problems in your system. You can also send the package to Oracle Support, as needed.

### Syntax

```
CREATE DIAGPACK packStartTime=time, [durationInHrs=duration]
```

or

```
CREATE DIAGPACK alertName=alertName
```

### Usage Notes

When an alert occurs, a diagnostic package is created automatically. This package contains logs and traces related to the alert.

The `CREATE DIAGPACK` command enables you to generate diagnostic packages manually.

- The `packStartTime` parameter specifies when to start collecting the logs and traces. The format of `packStartTime` is:

```
yyyy_MM_ddTHH_mm_ss
```

For example: 2015\_07\_07T09\_00\_00

You can also specify the keyword `now` for `packStartTime`. The `packStartTime` cannot be in the future and cannot be older than 7 days. The value of `packStartTime` is used as part of the name of the diagnostic package.

- The `durationInHrs` parameter specifies the number of hours of logs and traces to include in the diagnostic package. Valid values are from 1 (default) to 6.  
Every diagnostic package includes logs 1 hour before and 1 hour after the `packStartTime`. For example, if you specify a time of 12\_00\_00, then logs will be collected from 11\_00\_00 to 13\_00\_00, unless the end time is in the future.
- The `alertName` parameter specifies the alert name for which to create the diagnostic package. You can run the `LIST ALERTHISTORY` command to view the alert names.



### Name of Diagnostic Packages

The name of the diagnostic package is formed as follows:

```
hostname + packStartTime + unique package ID
```

For example: `scaqab04adm08_2016_05_17T11_58_54_5_1.tar.bz2`

For alerts, the name of the diagnostic package is formed as:

```
hostname + timestamp of when the package was created + alert ID
```

For example: `scaqab04adm08_2016_05_17T11_58_54_5_1.tar.bz`

### Location of Diagnostic Packages

The location of the diagnostic packages is `/opt/oracle/dbserver/dbms/deploy/log`.

### Status of Diagnostic Packages

You can run the LIST DIAGPACK command to get a list of diagnostic packages in your system, and their status.

### Privileges Needed to Create, List, and Download Diagnostic Packages

Users need to be granted roles before they can download the diagnostic packages. Use DBMCLI to grant the following privileges to a role:

- Privilege to create diagnostic packages:

```
grant privilege CREATE ON DIAGPACK to ROLE role
```

- Privilege to list diagnostic packages and check their status:

```
grant privilege LIST ON DIAGPACK to ROLE role
```

- Privilege to download diagnostic packages:

```
grant privilege DOWNLOAD ON DIAGPACK to ROLE role
```

You can then grant the role to users. For example, if you named your role `diagpack_role`, the following command grants the role to the user `fred`.

```
DBMCLI> GRANT ROLE diagpack_role TO USER fred
```

During deployment, Oracle Exadata Deployment Assistant (OEDA) creates an Exadata storage software user called `CELLDIAG`. You can use this user to connect to a cell remotely using ExaCLI or REST API. This user has privileges to create, list, and download diagnostic packages.

## Downloading Diagnostic Packages

You can download diagnostic packages using any of the following methods. Note that you need the `DOWNLOAD ON DIAGPACK` privilege before you can download diagnostic packages.

- Using the REST API
  - To download the diagnostic package by name, use the following URL, where *hostname* specifies the host name of the server and *diagpackname* specifies the name of the diagnostic package. The URL must point to the secure socket port, which is 7879 by default:

```
https://hostname:7879/diagpack/download?name=diagpackname
```

If the user is not already logged in, the URL will prompt for a user name and password.

Diagnostic packages can also be accessed at

```
https://hostname:7879/diagpack
```

For example:

```
https://dbnode01.example.com:7879/diagpack
```

The page then prompts the user to log in:

```
User: fred
Password: *****
```

Based on the user's privileges, various sections of this page could be hidden. The form to create a new diagpack will not be shown if the user does not have the `CREATE ON DIAGPACK` privilege. Similarly, the list of alerts and their diagnostic packages will not be shown if the user does not have the `LIST ON DIAGPACK` privilege.

To download the diagnostic package by alert name, use the following URL, where *hostname* specifies the host name of the server and *alertName* specifies the alert name of the diagnostic package:

```
https://hostname:7879/diagpack/download?alert=alertName
```

This is the same alert name that is used in `AlertHistory`. It looks like 1, 2, 3 for stateless alerts, and 1\_1, 2\_1, 3\_1, 3\_2 for stateful alerts.

- Using the `download` ExaCLI command

ExaCLI enables you to run CellCLI commands on storage nodes remotely from compute nodes. To run the `download` command, run the following commands on a compute node:

1. Start up ExaCLI and connect to the cell containing the diagnostic pack. Use a command similar to the following where *hostname* specifies the host name of the cell:

```
exaccli -l celladministrator -c hostname
Password=*****
```

2. Run the `download` command. Use a command similar to the following where *name* specifies the name of the diagnostic package to download, and *destinationFolder* specifies the directory where you want to save the downloaded diagnostic package:

```
exaccli> download diagpack name destinationFolder
```

- Getting the diagnostic package from the alert emails

The alert emails include diagnostic packages for all alerts except INFO, CLEAR, and WARNING. Diagnostic packages are generated for critical alerts only.

### Turning Off the Diagnostic Pack Attachment in Emails

To turn off the diagnostic pack attachment in emails, use the command `alter dbserver diagPackEmailAttach=FALSE`. The diagnostic packs are still generated and stored on the system. To download the diagnostic packs, see the section "[Downloading Diagnostic Packages](#)."

### Examples

#### Example 9-36 Using "now" for packStartTime

This example creates a diagnostic package using `NOW` as the start time and the default duration of one hour.

The output is a single compressed file in the directory `/opt/oracle/dbserver/dbms/deploy/log`.

```
DBMCLI> create diagpack packStartTime="now"
Processing: dm01db01_diag_2015_07_08T17_53_58_1
Use 'list diagPack' to check its status.
```

#### Example 9-37 Specifying a duration

This example creates 3 diagnostic packages under `/opt/oracle/dbserver/dbms/deploy/log`:

- The first package has a start time of `2015_07_07T09_00_00`.
- The second package has a start time of `2015_07_07T10_00_00`.
- The third package has a start time of `2015_07_07T11_00_00`.

```
DBMCLI> create diagpack packStartTime="2015_07_07T09_00_00", durationInHrs=3
Processing: dm01db01_diag_2015_07_07T09_00_00_1
dm01db01_diag_2015_07_07T10_00_00_1 (In queue...)
dm01db01_diag_2015_07_07T11_00_00_1 (In queue...)
Use 'list diagPack' to check its status.
```

### Related Topics

- About Automatic Diagnostic Repository
- [LIST ALERTHISTORY](#)
- [LIST DIAGPACK](#)
- Using the ExaCLI Utility

## 9.7.2.3 CREATE ROLE

### Purpose

The `CREATE ROLE` command creates a role for a user accessing a database server.

### Syntax

```
CREATE ROLE role_name1 [, role_name2, ...]
```

### Usage Notes

The role name should be unique.

### Examples

[Example 9-38](#) shows how to create a role.

#### Example 9-38 Creating a Role

```
DBMCLI> CREATE ROLE dbmonitor
```

## 9.7.2.4 CREATE THRESHOLD

### Purpose

The `CREATE THRESHOLD` command creates a threshold object that specifies the conditions for generation of a metric alert.

### Syntax

```
CREATE THRESHOLD name attributename=value [, attributename=value]...
```

### Usage Notes

The attributes that can be specified are `comparison`, `critical`, `occurrences`, `observation`, and `warning`.

- The *name* argument is required. The name is comprised of a metric name and an object name with the format `metricName.objectName`, such as `db_io_rq_sm_sec.db123` or `ct_io_wt_rq.interactive`. Use the `LIST METRICCURRENT metric` command to display the available object name for metric. The object name is optional.
- When a object name is not specified, then the threshold is applied to all metric objects for the given metric.
- The `comparison` attribute is required with a condition value. The value must be '`<`', '`<=`', '`=`', '`>=`', or '`>`'.

- The `occurrences` attribute specifies the number of consecutive measurements over the threshold value that trigger a state change.

When specifying occurrences and observations, you need the specified number of consecutive occurrences of sample averages over the number of observations to cause an alert. For example, if the following five observations (`observations=5`) happen on a database server, then the average sample would be 10 because the number of consecutive occurrences (`occurrences=2`) had values of 5 and 15.

```
Observation 1: 0
Observation 2: 30
Observation 3: 0
Observation 4: 5
Observation 5: 15
```

- The `observation` attribute is the number of measurements over which measured values are averaged.
- A state change to the value set in `warning` or `critical` causes a stateful alert to be generated.

### Example 9-39 Creating a Threshold

This example shows how to create a threshold.

```
DBMCLI> CREATE THRESHOLD temp comparison='>', critical=30
```

#### See Also:

- *Oracle Exadata System Software User's Guide* for information about metrics
- "[LIST ALERTDEFINITION](#)" for information about metrics and thresholds

## 9.7.2.5 CREATE USER

### Purpose

The `CREATE USER` command creates a user.

### Syntax

```
CREATE USER name PASSWORD = *
```

### Usage Notes

- The user name should be unique.
- The `username` cannot be `root`, `dbmadmin` or `dbmmonitor`. Those are reserved.
- The system prompts for a password for the new user. The password must have 12 to 40 alphanumeric characters or special characters `!@#$$%^&*()` with at least one digit, one lowercase letter, and one uppercase letter. Starting with Oracle Exadata System Software release 18.1.0.0.0, the password can be 8 to 40 characters in length and can also utilize the special characters `-` and `_`.

- The new password cannot be the same as the current password for the user.

**Example 9-40 Creating a User**

```
DBMCLI> CREATE USER jdoe PASSWORD = *
password:
Confirm password: password
User jdoe successfully created.
```

## 9.7.3 DESCRIBE

**Purpose**

The `DESCRIBE` command displays a list of attributes for the object type that is provided as an argument. The tag `modifiable` indicates whether an attribute can be modified using the `ALTER` command.

**Syntax**

```
DESCRIBE object_type
```

**Usage Notes**

- The *object\_type* is one of the supported object types.
- The list of attributes can be used as arguments in the `LIST` command.
- `DESCRIBE` does not display all of the attributes for the objects.
- [DESCRIBE ALERTDEFINITION](#)
- [DESCRIBE ALERTHISTORY](#)
- [DESCRIBE DBSERVER](#)
- [DESCRIBE IBPORT](#)
- [DESCRIBE LUN](#)
- [DESCRIBE METRICCURRENT](#)
- [DESCRIBE METRICDEFINITION](#)
- [DESCRIBE METRICHISTORY](#)
- [DESCRIBE PHYSICALDISK](#)
- [DESCRIBE ROLE](#)
- [DESCRIBE SOFTWAREHISTORY](#)
- [DESCRIBE SOFTWAREUPDATE](#)
- [DESCRIBE THRESHOLD](#)
- [DESCRIBE USER](#)

**Related Topics**

- [About DBMCLI Object Types](#)

## 9.7.3.1 DESCRIBE ALERTDEFINITION

### Purpose

The `DESCRIBE ALERTDEFINITION` command displays a list of attributes for the `ALERTDEFINITION` object type.

### Syntax

```
DESCRIBE ALERTDEFINITION
```

### Usage Notes

[Table 9-2](#) lists the attributes for the `DESCRIBE ALERTDEFINITION` command.

**Table 9-2 DESCRIBE ALERTDEFINITION Attributes**

Attribute	Description
alertShortName	Abbreviated name for the alert. If the alert is based on a metric, then the short name is the same as the corresponding metric <code>name</code> attribute.
alertSource	Source of the alert, such as BMC or ADR.
alertType	Type of the alert. Values are stateful or stateless. <ul style="list-style-type: none"> <li>Stateful alerts are automatically cleared on transition to <code>normal</code>.</li> <li>Stateless alerts are never cleared. You can change the alert by setting the <code>examinedBy</code> attribute.</li> </ul>
description	Description for the alert.
metricName	Metric name if the alert is based on a metric.
name	Identifier for the alert.

### Examples

[Example 9-41](#) shows the `DESCRIBE` command with the `ALERTDEFINITION` object.

#### Example 9-41 Describing the ALERTDEFINITION Object

```
DBMCLI> DESCRIBE ALERTDEFINITION
```

```

name
alertShortName
alertSource
alertType
description
metricName
```

## 9.7.3.2 DESCRIBE ALERTHISTORY

### Purpose

The `DESCRIBE ALERTHISTORY` command displays a list of attributes for the `ALERTHISTORY` object type.

## Syntax

```
DESCRIBE ALERTHISTORY
```

## Usage Notes

The following table lists the attributes for the `DESCRIBE ALERTHISTORY` command.

**Table 9-3 DESCRIBE ALERTHISTORY Attributes**

Attribute	Description
alertAction	Recommended action to perform for this alert.
alertMessage	Brief explanation of the alert.
alertSequenceID	Unique sequence ID for the alert. When an alert changes its state, such as <code>warning</code> to <code>critical</code> or <code>critical</code> to <code>clear</code> , another occurrence of the alert is created with the same sequence number and a time stamp of the transition.
alertShortName	Abbreviated name for the alert. If the alert is based on a metric, then the short name is the same as the corresponding metric <code>name</code> attribute.
alertType	Type of the alert. Values are <code>stateful</code> or <code>stateless</code> . <ul style="list-style-type: none"> <li>Stateful alerts are automatically cleared on transition to <code>normal</code>.</li> <li>Stateless alerts are never cleared. You can change the alert by setting the <code>examinedBy</code> attribute.</li> </ul>
beginTime	Time stamp when an alert changes its state.
endTime	Time stamp for the end of the period when an alert changes its state.
examinedBy	Administrator who reviewed the alert.
failedMail	Intended e-mail recipient when a notification failed.
failedSNMP	Intended SNMP subscriber when a notification failed.
metricObjectName	Object, such as database server disk for which a metric threshold has caused an alert.
metricValue	Value of the metric that caused the alert.
name	Unique identifier for the alert.
notificationState	Number indicating progress in notifying subscribers to alert messages: <ul style="list-style-type: none"> <li>0: never tried</li> <li>1: sent successfully</li> <li>2: retrying, up to 5 times</li> <li>3: 5 failed retries</li> </ul>
sequenceBeginTime	Time stamp when an alert sequence ID is first created.
severity	Severity level. Values are <code>clear</code> , <code>info</code> , <code>warning</code> , or <code>critical</code> .

## Examples

[Example 9-42](#) shows the `DESCRIBE` command with the `ALERTHISTORY` object.



**Example 9-42 Describing the ALERTHISTORY Object**

```

DBMCLI> DESCRIBE ALERTHISTORY

 name
 alertAction
 alertMessage
 alertSequenceID
 alertShortName
 alertType
 beginTime
 endTime
 examinedBy modifiable
 failedMail
 failedSNMP
 metricObjectName
 metricValue
 notificationState
 sequenceBeginTime
 severity

```

**Related Topics**

- [DESCRIBE METRICDEFINITION](#)

### 9.7.3.3 DESCRIBE DBSERVER

**Purpose**

The `DESCRIBE DBSERVER` command displays a list of attributes for the `DBSERVER` object type.

**Syntax**

```
DESCRIBE DBSERVER
```

**Usage Notes**

The following table lists the attributes for the `DESCRIBE DBSERVER` command.

**Table 9-4 DESCRIBE DBSERVER Attributes**

Attribute	Description
accountLockInDays	Number of days after a password expires before a user account is locked.
bbuStatus	Status of hard disk controller battery-backed unit (BBU).
comment	User-supplied text string.
corecount	Number of CPU cores in the database server. If expressed using the format <i>n1/n2</i> , then <i>n1</i> is the number of currently enabled CPU cores and <i>n2</i> is the total number of CPU cores in the database server.

Table 9-4 (Cont.) DESCRIBE DBSERVER Attributes

Attribute	Description
cpuCount	Number of CPUs on the database server as reported by the operating system in <code>/proc/cpuinfo</code> and other utilities. This value factors in hyperthreading, so will always be twice the value of <code>corecount</code> . If expressed using the format <code>n1/n2</code> , then <code>n1</code> is twice the number of currently enabled CPU cores and <code>n2</code> is twice the total number of CPU cores in the database server.
diagHistoryDays	Number of days ADR files are retained. The default is 7 days.
diagPackEmailAttach	Whether a diagpack is included as an attachment in the alert email or not. The default is <code>true</code> .
diagPackUploadEnabled	Whether the auto diagpack upload feature is enabled or not. The default is <code>true</code> .
emailFormat	File format for email messages. The value can be HTML or text.
fanCount	Count of working fans and total fans, displayed as <code>working/total</code> .
fanStatus	Status of the fan. The value can be <code>normal</code> , <code>warning</code> , or <code>critical</code> .
httpsAccess	Control list of IP addresses for HTTPs port access to the Exadata RESTful Service
id	Global unique identifier (GUID) supplied by the hardware vendor.
interconnect1-4	Interconnect 1 to 4 for the database server. For example: <code>bondeth0</code> .
interconnectCount	Number of network interconnects.
ipaddress1-4	IP address 1 to 4 for the database server.
kernelVersion	Version of the host kernel software.
location	Physical location of the cell hardware supplied by the user.
locatorLEDStatus	Status of cell LOCATE LED. The value can be <code>on</code> or <code>off</code> .
mailServer	Fully qualified domain name of the email relay server used to send alert notifications. This attribute only requires specification in cases where DNS returns an unreachable or invalid mail exchange (MX) record for the email server specified in <code>smtpToAddr</code> .
makeModel	Make and model of the cell hardware supplied by the vendor.

**Table 9-4 (Cont.) DESCRIBE DBSERVER Attributes**

Attribute	Description
metricCollection	Indicator for whether Management Server performs metric collection. Values are TRUE or FALSE. If set to FALSE, then all collection and alert mining is stopped. The default setting is TRUE.
metricHistoryDays	Number of days metric history files are retained. The default is 7 days.
msVersion	Version of the management server.
name	Name of the database server.
notificationMethod	Notification method for alerts. The value should be mail, snmp, none, or both mail and snmp.
notificationPolicy	Indicator for severity alerts to be sent to subscribers. The value for notificationPolicy should be none or any combination of critical, warning, and clear.
powerCount	Count of power supplies, displayed as working/total.
powerStatus	Status of the power. The value can be normal, warning, or critical.
pwdExpInDays	The number of days before a user's password expires.
pwdExpWarnInDays	The number of days before a user's password expires that a warning message is issued during login attempts.
releaseImageStatus	Indicator for knowing whether imaging is successful or not.
releaseTrackingBug	Patch number for the cell software, such as 11223344.
releaseVersion	Release number for the cell software, such as 11.2.2.3.0.
remotePwdChangeAllowed	Whether or not a user password can be changed remotely through REST services.
rescuePlan	A list of commands that you can run after a server rescue to restore settings, such as thresholds and notifications, to the last known values.
smtpFrom	User name that appears in the From: header of the alert notifications.
smtpFromAddr	Email address that appears in the From: header of the alert notifications. This email address is not authenticated with the email server.
smtpPort	Email server port used to send alert notifications.
smtpSubscriber	List of hosts that subscribe to the SNMP alert notifications.

**Table 9-4 (Cont.) DESCRIBE DBSERVER Attributes**

Attribute	Description
smtpToAddr	Address to which email is sent. It can be a comma-delimited list in quotation marks to allow multiple subscribers to alerts.
smtpUseSSL	Specification to use Secure Socket Layer (SSL) encryption for alert notifications.
status	Status of the database server.
syslogConf	Rules for the system logging utility syslogd.
temperatureReading	Current temperature (Celsius) of the cell obtained from the BMC.
temperatureStatus	Status of the temperature. The value can be normal, warning, or critical.
traceLevel	Level for which trace messages are written. This can be a valid Java logging level (SEVERE, WARNING, INFO, CONFIG, FINE, FINER, FINEST) or a valid Oracle Diagnostic Logging (ODL) logging level (INCIDENT_ERROR:1, ERROR:1, WARNING:1, NOTIFICATION:1, NOTIFICATION:16, TRACE:1, TRACE:16, TRACE:32). The default is FINE.
upTime	Time (days, hours:minutes) since the system was restarted.

**Examples****Example 9-43 Describing the DBSERVER Object**

```
DBMCLI> DESCRIBE DBSERVER
```

```

name modifiable
bbuStatus
comment modifiable
coreCount
cpuCount
diagHistoryDays modifiable
diagPackEmailAttach modifiable
diagPackUploadEnabled modifiable
emailFormat modifiable
emailSubscriber modifiable
fanCount
fanStatus
httpsAccess modifiable
iaasIdleInUse modifiable
iaasMode modifiable
iaasReason modifiable
id
interconnectCount
interconnect1 modifiable
interconnect2 modifiable

```

interconnect3	modifiable
interconnect4	modifiable
interconnect5	modifiable
interconnect6	modifiable
interconnect7	modifiable
interconnect8	modifiable
ipaddress1	
ipaddress2	
ipaddress3	
ipaddress4	
ipaddress5	
ipaddress6	
ipaddress7	
ipaddress8	
kernelVersion	
locatorLEDStatus	
location	modifiable
mailServer	modifiable
makeModel	
metricCollection	modifiable
metricHistoryDays	modifiable
msVersion	
notificationMethod	modifiable
notificationPolicy	modifiable
pendingCoreCount	modifiable
powerCount	
powerStatus	
releaseImageStatus	
releaseVersion	
releaseTrackingBug	
securityCert	modifiable
smtpFrom	modifiable
smtpFromAddr	modifiable
smtpPort	modifiable
smtpToAddr	modifiable
smtpUseSSL	modifiable
snmpSubscriber	modifiable
snmpUser	modifiable
status	
syslogConf	modifiable
temperatureReading	
temperatureStatus	
traceLevel	modifiable
upTime	
msStatus	
rsStatus	

### 9.7.3.4 DESCRIBE IBPORT

#### Purpose

The `DESCRIBE IBPORT` command displays a list of attributes for the IBPORT object type.

 **Note:**

This command does not work on Exadata Database Machine servers with RoCE Network Fabric.

**Syntax**

```
DESCRIBE IBPORT
```

**Usage Notes****Table 9-5 DESCRIBE IBPORT Attributes**

Attribute	Description
activeSlave	Indicator whether the port is currently the active port for the bonded IP.
dataRate	The data rate of the InfiniBand Network Fabric port.
hcaFWVersion	The version of the host channel adapter firmware.
id	The Global unique identifier (GUID) of the InfiniBand Network Fabric port.
lid	The local identifier of the InfiniBand Network Fabric port. It is unique within the subnet, and the 16-bit identifiers are used within a network by switches for routing.
linkDowned	The number of times the port training state machine has failed the link error recovery process, and halted the link.
linkIntegrityErrs	The number of link integrity errors.
linkRecovers	The number of times the port training state machine has successfully completed the link error recovery process.
name	The name of the InfiniBand Network Fabric port.
physLinkState	The physical link state.
portNumber	The port number of the InfiniBand Network Fabric port.
rcvConstraintErrs	The number of received constraint errors experienced by the InfiniBand Network Fabric port.
rcvData	The number of 32-bit data words received by the InfiniBand Network Fabric port.
rcvErrs	The number of packets received at the InfiniBand Network Fabric port containing an error.
rcvRemotePhysErrs	The number of physical errors experienced at the InfiniBand Network Fabric port.
status	The link status.
symbolErrs	The number of minor link errors experienced at the InfiniBand Network Fabric port.
vl15Dropped	The number of incoming VL15 packets dropped at the InfiniBand Network Fabric port due to resource limitations, such as lack of buffers.

**Table 9-5 (Cont.) DESCRIBE IBPORT Attributes**

Attribute	Description
xmtConstraintErrs	The number of transmitted constraint errors experienced at the InfiniBand Network Fabric port.
xmtData	The number of 32-bit data words transmitted on the InfiniBand Network Fabric port.
xmtDiscards	The number of outbound packets discarded by the InfiniBand Network Fabric port because the port was down or congested.

**Example 9-44 Describing the IBPORT Object**

This example shows the possible output for the `DESCRIBE IBPORT` command.

```
DBMCLI> DESCRIBE IBPORT

 name
 activeSlave
 dataRate
 hcaFWVersion
 id
 lid
 linkDowned
 linkIntegrityErrs
 linkRecovers
 physLinkState
 portNumber
 rcvConstraintErrs
 rcvData
 rcvErrs
 rcvRemotePhysErrs
 status
 symbolErrs
 vl15Dropped
 xmtConstraintErrs
 xmtData
 xmtDiscards
```

## 9.7.3.5 DESCRIBE LUN

**Purpose**

The `DESCRIBE LUN` command displays a list of attributes for the `LUN` object type.

**Syntax**

```
DESCRIBE LUN
```

## Usage Notes

**Table 9-6 DESCRIBE LUN Attributes**

Attribute	Description
deviceName	Operating system device name for the LUN. For example, /dev/c1d5
diskType	The type of disk.
errorCount	Number of errors on this LUN.
id	Identifier assigned by the system.
lunSize	Raw size of the LUN before being converted to a database server disk.
lunUID	Unique identifier assigned by the system.
name	Unique name assigned to the LUN. This might be different (or extended from) the LUN ID if the ID is not unique.
overProvisioning	Indicator of the percentage of over-provisioned blocks in flash storage that are still available for a particular LUN. This attribute is only used for flash disks.
raidLevel	Value of the RAID level that is used on the LUN. For example: RAID 0.
status	Status of the LUN, which can be normal, warning, or critical.

**Example 9-45 Describing the LUN Object**

This example shows the DESCRIBE command with the LUN object.

```
DBMCLI> DESCRIBE LUN

 name
 deviceName
 diskType
 errorCount
 id
 lunSize
 lunUID
 overProvisioning
 raidLevel
 status
```

### 9.7.3.6 DESCRIBE METRICCURRENT

**Purpose**

The DESCRIBE METRICCURRENT command displays a list of attributes for the METRICCURRENT object type.



## Syntax

```
DESCRIBE METRICCURRENT
```

## Usage Notes

**Table 9-7 DESCRIBE METRICCURRENT Attributes**

Attribute	Description
alertState	Indicator of the alert state. Values are normal, warning, or critical.
collectionTime	Time stamp when the metric value was collected.
metricObjectName	Name of the object, such as database server being measured.
metricType	Specification for how the statistic was created or defined.
metricValue	Value of the metric when it was collected.
name	Unique name of the current metric.
objectType	Options are DBSERVER, DBSERVER_FILESYSTEM, and IBPORT.

### Example 9-46 Describing the METRICCURRENT Object

This example shows the DESCRIBE command with the METRICCURRENT object.

```
DBMCLI> DESCRIBE METRICCURRENT
```

```

name
alertState
collectionTime
metricObjectName
metricType
metricValue
objectType
```

## Related Topics

- [DESCRIBE METRICDEFINITION](#)

## 9.7.3.7 DESCRIBE METRICDEFINITION

### Purpose

The DESCRIBE METRICDEFINITION command displays a list of attributes for the METRICDEFINITION object type.

### Syntax

```
DESCRIBE METRICDEFINITION
```

### Usage Notes

[Table 9-8](#) lists the attributes for the DESCRIBE METRICDEFINITION command.

**Table 9-8 DESCRIBE METRICDEFINITION Attributes**

Attribute	Description
description	Description of the metric.
metricType	Indicator of how the statistic was created or defined. The options are as follows: <ul style="list-style-type: none"> <li>• <code>cumulative</code>: Cumulative statistics since the metric was created.</li> <li>• <code>instantaneous</code>: Value at the time that the metric is collected.</li> <li>• <code>rate</code>: Rates computed by averaging statistics over observation periods.</li> <li>• <code>transition</code>: Transition metrics are collected at the time their value has changed and typically capture important transitions in hardware status.</li> </ul>
name	Unique name of the metric definition. The value of the <code>name</code> attribute is a composite of abbreviations. The attribute value starts with an abbreviation of the object type on which the metric is defined: <ul style="list-style-type: none"> <li>• <code>DS_</code> (database server)</li> <li>• <code>N_</code> (network)</li> </ul>
objectType	Options are <code>DBSERVER</code> , <code>DBSERVER_FILESYSTEM</code> , and <code>IBPORT</code> .
persistencePolicy	Amount of time metric values are stored.
unit	Unit for the metric explicitly, and is related to the metric collected: <ul style="list-style-type: none"> <li>• Number</li> <li>• % (percentage)</li> <li>• F (Fahrenheit)</li> <li>• C (Celsius)</li> </ul>

**Example 9-47 Describing the METRICDEFINITION Object**

This example shows the `DESCRIBE` command with the `METRICDEFINITION` object.

```
DBMCLI> DESCRIBE METRICDEFINITION
```

```

name
description
metricType
objectType
persistencePolicy
unit
```

### 9.7.3.8 DESCRIBE METRICHISTORY

**Purpose**

The `DESCRIBE METRICHISTORY` command displays a list of attributes for the `METRICHISTORY` object type.

## Syntax

```
DESCRIBE METRICHISTORY
```

## Usage Notes

[Table 9-9](#) lists the attributes for the `DESCRIBE METRICHISTORY` command.

**Table 9-9 DESCRIBE METRICHISTORY Attributes**

Attribute	Description
alertState	Indicator of the alert state. Values are normal, warning, or critical.
collectionTime	Time stamp when the metric value was collected.
memory	The metrics in Management Server memory. This attribute can be used instead of the <code>collectionTime</code> attribute when the collection time is less than an hour.
metricObjectName	Name of the object, such as database server disk being measured.
metricType	Specification for how the statistic was created or defined.
metricValue	Value of the metric when it was collected.
metricValueAvg	Average value of the metric.
metricValueMax	Maximum value of the metric.
metricValueMin	Minimum value of the metric.
name	Unique name of the current metric.
objectType	Options are <code>DBSERVER</code> , <code>DBSERVER_FILESYSTEM</code> , and <code>IBPORT</code>
over	Specification for amount of time in minutes for the aggregation. This attribute works in conjunction with the <code>max</code> , <code>min</code> and <code>avg</code> keywords.

### Example 9-48 Describing the METRICHISTORY Object

This example shows the `DESCRIBE` command with the `METRICHISTORY` object.

```
DBMCLI> DESCRIBE METRICHISTORY
```

```

name
alertState
collectionTime
memory
metricObjectName
metricType
metricValue
metricValueAvg
metricValueMax
metricValueMin
objectType
over
```

## 9.7.3.9 DESCRIBE PHYSICALDISK

### Purpose

The `DESCRIBE PHYSICALDISK` command displays a list of attributes for the `PHYSICALDISK` object type.

### Syntax

```
DESCRIBE PHYSICALDISK
```

### Usage Notes

The following table lists the attributes for the `DESCRIBE PHYSICALDISK` command.

**Table 9-10 DESCRIBE PHYSICALDISK Attributes**

Attribute	Description
<code>ctrlHwVersion</code>	Controller version. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
<code>diskType</code>	Type of the disk, whether it is a <code>HARDDISK</code> or a <code>FLASHDISK</code> .
<code>enclosureDeviceId</code>	Identifier for the hard disk enclosure. This attribute is only applicable to Oracle Exadata System Software on Oracle Exadata Storage Server.
<code>errCmdTimeoutCount</code>	Count of the number of command timeout SCSI errors on the disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
<code>errHardReadCount</code>	Count of the number of hard read errors on the disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
<code>errHardWriteCount</code>	Count of the number of hard write errors on the disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
<code>errMediaCount</code>	Count of the number of media errors on the hard disk. This attribute is only applicable to Oracle Exadata System Software on Oracle Exadata Storage Server.
<code>errOtherCount</code>	Count of unclassified errors on the hard disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
<code>errSeekCount</code>	Count of the number of SCSI seek errors on the disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
<code>hotPlugCount</code>	Count of the times the disk has been inserted into the rack. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
<code>lastFailureReason</code>	Type of the most-recent error on the disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
<code>luns</code>	List of LUNs converted from this disk.

**Table 9-10 (Cont.) DESCRIBE PHYSICALDISK Attributes**

Attribute	Description
makeModel	Model description provided by the system.
name	Unique name of the physical disk.
physicalFirmware	System-assigned name of the firmware for the disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
physicalInsertTime	Time that the disk was inserted.
physicalInterface	Interface type used by the hard disk. For example, SAS
physicalPort	Port on the controller for this disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
physicalRPM	Disk revolutions per minute. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
physicalSerial	System-assigned unique ID.
physicalSize	Size of the disk in bytes.
physicalUseType	Intended use of the disk. For example, Data Drive.
sectorRemapCount	Count of all sector remap operations for the physical disk. This attribute is only applicable to Oracle Exadata System Software on HP Oracle Database Machine.
slotNumber	Physical location of disk.
status	Status of the physical disk. Values can be as follows: <ul style="list-style-type: none"> <li>failed: disk has failed. In earlier releases, this status was called critical.</li> <li>normal: disk is functioning normally</li> <li>not present: disk has been removed</li> <li>peer failure: flash disk failure only</li> <li>poor performance: disk is performing poorly</li> <li>predictive failure: disk is expected to fail</li> <li>write-through caching: flash disk caching only.</li> </ul>

**Example 9-49 Describing the PHYSICALDISK Object on a Hard Drive in a Database Server**

This example shows the DESCRIBE command with the PHYSICALDISK object on Oracle Exadata Storage Server.

```
DBMCLI> DESCRIBE PHYSICALDISK

name
ctrlFirmware
diskType
enclosureDeviceId
errMediaCount
errOtherCount
luns
```

```
makeModel
physicalFirmware
physicalInsertTime
physicalInterface
physicalSerial
physicalSize
slotNumber
status
```

### 9.7.3.10 DESCRIBE ROLE

#### Purpose

The `DESCRIBE ROLE` command displays a list of attributes for the `ROLE` object type.

#### Syntax

```
DESCRIBE ROLE
```

#### Usage Notes

**Table 9-11** DESCRIBE ROLE Attributes

Attribute	Description
name	Unique name of the user assigned the role.
privileges	Privileges granted to the role.

#### Examples

##### Example 9-50 Describing the ROLE Object

This example shows the `DESCRIBE` command with the `ROLE` object.

```
DBMCLI> DESCRIBE ROLE
name
privileges
```

### 9.7.3.11 DESCRIBE SOFTWAREHISTORY

#### Purpose

The `DESCRIBE SOFTWAREHISTORY` command displays a list of attributes for the `ALERHISTORY` object type.

#### Syntax

```
DESCRIBE SOFTWAREHISTORY
```

#### Usage Notes

The attributes for the `DESCRIBE SOFTWAREHISTORY` command can include the following:

- `name`: The name of the software update
- `status`: The status of the software update

**Example 9-51 Describing the SOFTWAREHISTORY Object**

```
CellCLI> DESCRIBE SOFTWAREHISTORY
 name
 status
```

## 9.7.3.12 DESCRIBE SOFTWAREUPDATE

**Purpose**

The `DESCRIBE SOFTWAREUPDATE` command displays a list of attributes for the `SOFTWAREUPDATE` object type.

**Syntax**

```
DESCRIBE SOFTWAREUPDATE
```

**Usage Notes**

The attributes for the `DESCRIBE SOFTWAREUPDATE` command can include the following:

- `frequency`: The time period in which this software update is performed automatically. The value can be `none`, `daily`, `weekly`, or `biweekly`. The value `none` is available in Oracle Exadata System Software release 19.1.0 or later.
- `name`: The name of the patch to use in the update, or `unknown`. If the name defaults to `unknown`, then when the software update is performed, the most recent patch is chosen for the upgrade.
- `status`: The status of this software update.
- `store`: The URL for the location of the software update file
- `time`: The specified date and time at which the software update should be performed
- `timeLimitInMinutes`: The number of minutes a cell will spend waiting to update the software before canceling and issuing an alert.

**Example 9-52 Describing the SOFTWAREUPDATE Object**

```
CellCLI> DESCRIBE SOFTWAREUPDATE
 name modifiable
 status
 store modifiable
 time modifiable
 timeLimitInMinutes modifiable
```

## 9.7.3.13 DESCRIBE THRESHOLD

**Purpose**

The `DESCRIBE THRESHOLD` command displays a list of attributes for the `THRESHOLD` object type.

## Syntax

```
DESCRIBE THRESHOLD
```

## Usage Notes

[Table 9-12](#) lists the attributes for the `DESCRIBE THRESHOLD` command.

**Table 9-12 DESCRIBE THRESHOLD Attributes**

Attribute	Description
comparison	Operator for comparing the metric value to the threshold value (>, >=, =, <, <=) to determine whether the value violates the threshold.
critical	Limit beyond which the metric value is considered to be in the critical state for generating alerts.
name	Unique name of the threshold.
observation	Number of measurements over which the rate metric is averaged before being compared with the threshold value.
occurrences	Number of consecutive violations of the threshold limit by the metric value before the appropriate alert is issued.
warning	Limit beyond which the metric value is considered to be in the warning state for generating alerts.

### Example 9-53 Describing the THRESHOLD Object

This example shows the `DESCRIBE` command with the `THRESHOLD` object.

```
DBMCLI> DESCRIBE THRESHOLD

 name
 comparison modifiable
 critical modifiable
 observation modifiable
 occurrences modifiable
 warning modifiable
```

## 9.7.3.14 DESCRIBE USER

### Purpose

The `DESCRIBE USER` command displays a list of attributes for the `USER` object type.

### Syntax

```
DESCRIBE USER
```

### Usage Notes

The following table lists the attributes for the `DESCRIBE USER` command.



**Table 9-13 DESCRIBE USER Attributes**

Attribute	Description
name	Unique name of the user.
roles	Roles assigned to the user.

**Example 9-54 Describing the USER Object**

This example shows the `DESCRIBE` command with the `USER` object.

```
DBMCLI> DESCRIBE USER
```

```
 name
 roles
```

## 9.7.4 DROP

**Purpose**

The `DROP` command removes the named objects from the database server or resets a database server.

**Syntax**

```
DROP object_type [object_name [, object_name]...] [options]
```

**Usage Notes**

- *object\_type* can be one of the following:
  - ALERTHISTORY
  - DBSERVER
  - ROLE
  - THRESHOLD
  - USER
- *object\_name* is the name of a database server disk, role, threshold, or user.
- When multiple objects are the target of a `DROP` command, there is the possibility of partial success. If an error occurs, then the command is interrupted, and the remaining objects are not dropped.
- [DROP ALERTHISTORY](#)
- [DROP DBSERVER](#)
- [DROP ROLE](#)
- [DROP SOFTWAREHISTORY](#)
- [DROP THRESHOLD](#)
- [DROP USER](#)

**Related Topics**

- [About DBMCLI Object Types](#)

## 9.7.4.1 DROP ALERTHISTORY

**Purpose**

The `DROP ALERTHISTORY` command removes alerts from the alert history of a database server.

**Syntax**

```
DROP ALERTHISTORY {ALL | alert1 {, alert2}, ...}
```

**Usage Notes**

- In the command, *alertN* is the name of the alert to be dropped from the history.
- When dropping stateful alerts, you must drop all members of the alert sequence at the same time. If you do not drop all members, then an error is issued by the system.

**Example 9-55 Dropping a Database Server Alert History**

```
DBMCLI> DROP ALERTHISTORY 1, 2_1, 2_2
```

## 9.7.4.2 DROP DBSERVER

**Purpose**

The `DROP DBSERVER` command resets MS on the database server to its original state.

**Syntax**

```
DROP DBSERVER
```

**Usage Notes**

- Any thresholds that were set for the database server are also dropped for the database server.
- The `DBSERVER` attributes are set to their default values.

**Example 9-56 Dropping a Database Server**

```
DBMCLI> DROP DBSERVER
```

## 9.7.4.3 DROP ROLE

**Purpose**

The `DROP ROLE` command removes user roles from the cell.

**Syntax**

```
DROP ROLE { ALL | role_name1 [, role_name2, ...] } [FORCE]
```

**Usage Notes**

The `FORCE` option drops the role even if the role has been granted to a user.

**Example 9-57 Dropping a Role**

```
DBMCLI> DROP ROLE r1,r2
```

## 9.7.4.4 DROP SOFTWAREHISTORY

**Purpose**

The `DROP SOFTWAREHISTORY` command removes all history or individual update history.

**Syntax**

```
DROP SOFTWAREHISTORY { ALL | 'update_name [, update_name...]' }
```

**Example 9-58 Dropping the History of Scheduled Software Updates**

```
CellCLI> DROP SOFTWAREHISTORY '12.2.1.2.0.170509,12.2.1.2.0.17052'
```

```
CellCLI> DROP SOFTWAREHISTORY ALL
```

## 9.7.4.5 DROP THRESHOLD

**Purpose**

The `DROP THRESHOLD` command removes all or the specified thresholds from the database server.

**Syntax**

```
DROP THRESHOLD { ALL | threshold_name [, threshold_name ...] }
```

**Example 9-59 Dropping Thresholds**

```
DBMCLI> DROP THRESHOLD temp
```

## 9.7.4.6 DROP USER

**Purpose**

The `DROP USER` command removes a user from a database server.

**Syntax**

```
DROP USER { ALL | user1 [, user2]... }
```

**Usage Notes**

- *user(n)* is the name of a user to drop.

**Example 9-60 Dropping a User**

```
DBMCLI>DROP USER jdoe
```

## 9.7.5 GRANT

**Purpose**

The GRANT command sets attributes for privileges and roles.

**Syntax**

```
GRANT object_type [name] TO sub_object_type [sub_object_name]
```

**Usage Notes**

- *object\_type* can be as follows:
  - PRIVILEGE
  - ROLE
- The following values can be used for PRIVILEGE object type:
  - *name* is in the following format:
 

```
{ ALL ACTIONS | action } ON { ALL OBJECTS |
object } \
[{ ALL ATTRIBUTES | ATTRIBUTES attribute1 [,
attribute2, ...] }] \
[{ WITH ALL OPTIONS | WITH OPTIONS option1 [, option2, ...] }]
```
  - The *sub\_object\_type* must be ROLE.
  - The *sub\_object\_name* is the name of the role.
- The following can be used for the ROLE object type:
  - *name* is the role name.
  - The *sub\_object\_type* must be USER.
  - The *sub\_object\_name* is the name of the user.
- GRANT PRIVILEGE
- GRANT ROLE

## 9.7.5.1 GRANT PRIVILEGE

### Purpose

The `GRANT PRIVILEGE` command sets the access privileges for a role.

### Syntax

```
GRANT PRIVILEGE { ALL ACTIONS | action } ON { ALL OBJECTS | object } \
{ ALL ATTRIBUTES | ATTRIBUTES attribute1 [, attribute2, ...] } \
{ WITH ALL OPTIONS | WITH OPTIONS option1 [, option2, ...] } \
TO ROLE { ALL | role1 [, role2, ...] }
```

### Usage Notes

- *action* is the command. Examples: ALTER, CREATE, DESCRIBE, DROP, EXPORT, IMPORT, LIST.

#### Notes:

- The `GRANT` and `REVOKE` commands cannot be granted.
- `CREATE USER` and `DROP USER` cannot be granted.
- `CREATE ROLE` and `DROP ROLE` cannot be granted.
- *object* is object type for the action. It can be any DBMCLI object. Examples: DBSERVER, THRESHOLD, PHYSICALDISK, ALERTHISTORY, ROLE.
- *attribute* are the attributes for the object. To get a list of attributes for an object, run the `LIST object_type` command.
- *option* are the options for the object. Examples: DETAIL, LIMIT, ORDER BY, WHERE.
- *role* is the name of the role to grant privileges.
- The `ALL ACTIONS` argument grants privileges for all actions.
- The `ALL OBJECTS` argument grants privileges for all objects.
- The `ALL ATTRIBUTES` argument grants privileges for all attributes.
- The `WITH ALL OPTIONS` argument grants privileges for all options.
- Specifying attributes and `WITH OPTIONS` is optional. If they are not specified, then all attributes and options are granted with the privilege.

### Examples

#### Example 9-61 Granting Privileges to a Role

This example shows how to grant privileges to a role.

```
DBMCLI> GRANT PRIVILEGE list on alerthistory ATTRIBUTES
alertAction,alertMessage \
 WITH OPTIONS detail TO ROLE dbmonitor
```

**Example 9-62 Granting All Attributes and Options to a Role**

This example shows how to grant all attributes and options for a specified action and object to a role.

```
DBMCLI> GRANT PRIVILEGE { ALL ACTIONS | action } ON { ALL OBJECTS |
object } to ROLE role1
```

**Example 9-63 Granting All Options with Specified Action, Object and Attributes**

This example shows how to grant all options with a specified action, object and attributes to a role.

```
DBMCLI> GRANT PRIVILEGE { ALL ACTIONS | action } ON { ALL OBJECTS |
object } \
ATTRIBUTES attribute1 [, attribute2, ...] to ROLE role1
```

**Example 9-64 Granting All Attributes with Specified Action, Object and Options**

This example shows how to grant all attributes with a specified action, object, and options to a role.

```
DBMCLI> GRANT PRIVILEGE { ALL ACTIONS | action } ON { ALL OBJECTS |
object } \
WITH OPTIONS option1 [, option, ...] to ROLE role1
```

## 9.7.5.2 GRANT ROLE

**Purpose**

The `GRANT ROLE` command sets the role for a user.

**Syntax**

```
GRANT ROLE { ALL | role1 [, role2, ...] } TO USER { ALL | user1 [,
user2...] }
```

**Usage Notes**

- *role* is the name of the role.
- The `ALL` argument grants all roles to the user.
- The `TO USER ALL` argument grants the role to all users.

**Example 9-65 Granting a Role to a User**

This example shows how to grant a role to a user.

```
DBMCLI> GRANT ROLE dbmonitor TO USER agarcia
```

## 9.7.6 HELP

### Purpose

The `HELP` command displays syntax and usage descriptions for all DBMCLI commands.

### Syntax

```
HELP [help_topic]
```

If no topic argument is provided, `HELP` displays the name of all available topics. If a topic is specified, then detailed help text is displayed for that topic.

[Example 9-66](#) shows examples of the `HELP` command.

### Example 9-66 Display Help Text with the HELP Command

```
DBMCLI> HELP
DBMCLI> HELP ALTER
DBMCLI> HELP DROP ALERTHISTORY
```

## 9.7.7 LIST

### Purpose

The `LIST` command displays attributes for database server objects. Objects displayed are identified by name or by filters. The attributes displayed for each object are determined by the specified attribute list.

### Syntax

```
LIST object_type [name | attribute_filters] [attribute_list]
[DETAIL] [ORDER BY order_by_attribute_list] [LIMIT integer]
```

### Usage Notes

- *object\_type* can be the following:
  - ALERTDEFINITION
  - ALERTHISTORY
  - DBSERVER
  - IBPORT (not available in user domain (domU) environments)
  - LUN (not available in user domain (domU) environments)
  - METRICCURRENT
  - METRICDEFINITION
  - METRICHISTORY
  - PHYSICALDISK (not available in user domain (domU) environments)
  - ROLE
  - THRESHOLD

- USER
- Using `LIST` with only an *object\_type* (without the `DETAIL` option or an attribute list) displays the names of the existing objects of this type and a default list of attributes.
  - For an object type that has a `status` attribute, the object name and the status are displayed.
  - For the `METRICHISTORY` object type, the collection time, the object name, and value are displayed.
  - For the `PHYSICALDISK` and `LUN` object types, the ID attribute is displayed.
  - For the `ALERTHISTORY` object type, the time and alert message are displayed.
  - For the `KEY` object type, the key value is displayed.
- The attributes displayed for each object are determined by the specified attribute list. Attribute values that are strings with embedded blank spaces or tabs must be enclosed in quotation marks.
- Attribute filters determine the specific objects that are displayed. Because of the amount of metrics, you should use filters when using the `LIST METRICCURRENT` or `LIST METRICHISTORY` commands to narrow the output of the command.
- In the default format without the `DETAIL` option, each object is displayed on a separate line, with successive attribute values separated by tabs in the order of the specified list of attributes.
- In the `DETAIL` format, each attribute of a specific object is displayed on a separate line, with an attribute name followed by its value. If no attribute list is provided, then all attributes that have values are displayed. Blank lines separate each object in the display. `DETAIL` is similar to the `ATTRIBUTES ALL` option, only the format is different.
- Attributes that are not set are not listed with the `DETAIL` option. However, attributes that are set to an empty value are listed with the `DETAIL` option.
- `ORDER BY` orders the objects by attributes in ascending or descending order. The default is ascending.
- `LIMIT` sets a limit on the number of objects to display.
- [LIST ALERTDEFINITION](#)
- [LIST ALERTHISTORY](#)
- [LIST DBSERVER](#)
- [LIST DIAGPACK](#)
- [LIST IBPORT](#)
- [LIST LUN](#)
- [LIST METRICCURRENT](#)
- [LIST METRICDEFINITION](#)
- [LIST METRICHISTORY](#)
- [LIST METRICSTREAM](#)
- [LIST PHYSICALDISK](#)



- [LIST ROLE](#)
- [LIST SOFTWAREHISTORY](#)
- [LIST SOFTWAREUPDATE](#)
- [LIST THRESHOLD](#)
- [LIST USER](#)

#### Related Topics

- [Attribute Filters in LIST and ALTER Commands](#)  
You can use the *attribute\_filters* clause to specify the objects to display in `LIST` commands. Some `ALTER` commands also support the *attribute\_filters* clause.
- [About DBMCLI Object Types](#)
- [Attribute Lists in LIST Command](#)

### 9.7.7.1 LIST ALERTDEFINITION

#### Purpose

The `LIST ALERTDEFINITION` command displays all available sources of the alerts on the database server.

#### Syntax

```
LIST ALERTDEFINITION [name | attribute_filters] [attribute_list] [DETAIL]
```

#### Usage Notes

The list of attributes that can be displayed is shown in [Example 9-41](#).

#### Examples

[Example 9-67](#) shows the `LIST` command with the `ALERTDEFINITION` object.

#### Example 9-67 Listing ALERTDEFINITION Attributes

```
DBMCLI> LIST ALERTDEFINITION HardwareAlert DETAIL
name: HardwareAlert
alertShortName: Hardware
alertSource: Hardware
alertType: Stateless
description: "Hardware Alert"
metricName:
```

### 9.7.7.2 LIST ALERTHISTORY

#### Purpose

The `LIST ALERTHISTORY` command displays all alerts that occurred on the database server.

#### Syntax

```
LIST ALERTHISTORY [name | attribute_filters] [attribute_list] [DETAIL]
```

## Usage Notes

The list of attributes that can be displayed is shown in [Example 9-42](#).

A `WHERE` clause can include the `ageInMinutes` attribute to specify the list is limited to those alerts which have the specified age. For example, the following command would show the alerts created in the previous 15 minutes:

```
DBMCLI> LIST ALERTHISTORY WHERE ageInMinutes < 15
```

## Examples

[Example 9-68](#) shows the `LIST` command with the `ALERTHISTORY` object.

### Example 9-68 Listing ALERTHISTORY Attributes

```
DBMCLI> LIST ALERTHISTORY 1_1 DETAIL
name: 1_1
alertDescription: "Data hard disk of size 300GB in slot 3 entered
predictive failure status"
alertMessage: "Data hard disk entered predictive failure
status.
Status : WARNING - PREDICTIVE FAILURE
Manufacturer : HITACHI Model
Number : H103030SCSUN300G Size :
300GB
Serial Number : 1026GZKMDE
Firmware : A2A8 Slot Number : 3"
alertSequenceID: 1
alertShortName: Hardware
alertType: Stateful
beginTime: 2014-09-22T20:18:23-05:00
examinedBy:
metricObjectName: 252:3
notificationState: 0
sequenceBeginTime: 2014-09-22T20:18:23-05:00
severity: critical
alertAction: "Informational. The hard disk has entered
predictive failure status. A white DB server locator LED has been turned on to
help locate the affected cell, and an amber service action LED has been lit on
the
drive to help locate the affected drive. Detailed information on this problem can
be found at
https://support.oracle.com/CSP/main/article?cmd=show&type=NOT&id=1112995.1
Automatic Service Request has been notified with Unique Identifier:
c84839be-e8d8-4e5e-b315-e145ccd7cbc8."
```

## 9.7.7.3 LIST DBSERVER

### Purpose

The `LIST DBSERVER` command displays all alerts that occurred on the database server.

### Syntax

```
LIST DBSERVER [name | attribute_filters] [attribute_list] [DETAIL]
```

## Usage Notes

The list of attributes that can be displayed is shown in [Example 9-43](#).

## Examples

[Example 9-69](#) shows the output of the `LIST DBSERVER DETAIL` command.

[Example 9-70](#) shows how to display the value of the `rescuePlan` attribute.

[Example 9-71](#) shows how to retrieve the value of the `httpsAccess` attribute.

### Example 9-69 Listing DBSERVER Attributes

```

DDBMCLI> LIST DBSERVER DETAIL
 name: myexadb03
 bbuStatus: normal
 cpuCount: 24
 diagHistoryDays: 7
 fanCount: 16/16
 fanStatus: normal
 httpsAccess ALL
 id: 1029FMM085
 interconnectCount: 2
 ipAddress1: 192.168.10.25/22
 kernelVersion: 2.6.39-400.220.0.el6uek.x86_64
 locatorLEDStatus: on
 makeModel: Oracle Corporation SUN FIRE X4170 M2 SERVER
 metricHistoryDays: 7
 msVersion: OSS_MAIN_LINUX.X64_140921
 powerCount: 2/2
 powerStatus: normal
 releaseImageStatus: success
 releaseVersion: 12.1.2.1.0.140728
 releaseTrackingBug: 17885582
 snmpSubscriber:
 host=10.133.131.130,port=162,community=public,type=asr
 status: online
 temperatureReading: 20.0
 temperatureStatus: normal
 upTime: 55 days, 5:12
 msStatus: running
 rsStatus: running

DBMCLI> LIST DBSERVER ATTRIBUTES status,uptime
 online 161 days, 3:11

```

### Example 9-70 Displaying the rescuePlan Attribute

```

DBMCLI> LIST DBSERVER ATTRIBUTES rescuePlan

CREATE ROLE "listdbserverattrs"

GRANT PRIVILEGE list ON DBSERVER ATTRIBUTES bbuStatus, coreCount WITH ALL
OPTIONS TO ROLE "listdbserverattrs"

```

```
ALTER DBSERVER diagHistoryDays="7", metricHistoryDays="7",
bbuLearnSchedule="MONTH 1 DATE 17 HOUR 2 MINUTE 0",
alertSummaryStartTime="2016-09-26T08:00:00-07:00",
alertSummaryInterval=weekly, pendingCoreCount="128" force
```

### Example 9-71 Displaying the HTTPs Access Control List

This example shows how to view the HTTPs access control list for the Exadata RESTful service.

```
DBMCLI> LIST DBSERVER ATTRIBUTES httpsAccess
ALL
```

The value of `ALL` is the default value and allows access to all hosts.

## 9.7.7.4 LIST DIAGPACK

### Purpose

The `LIST DIAGPACK` command lists the diagnostic packages in your system, along with their status.

### Syntax

```
LIST DIAGPACK [DETAIL]
```

### Usage Notes

The location of the diagnostic packages is `/opt/oracle/dbserver/dbms/deploy/log`.

### Examples

#### Example 9-72 Output of the LIST DIAGPACK Command

This example shows the output of the `LIST DIAGPACK` command.

```
DBMCLI> LIST DIAGPACK
dbm04adm01_2016_06_07T12_28_23_17_1.tar.bz2
dbm04adm01_2016_04_11T12_51_05_16_1.tar.bz2
```

#### Example 9-73 Output of the LIST DIAGPACK command with the DETAIL option

This example shows the output of the `LIST DIAGPACK` command with the `DETAIL` option.

```
DBMCLI> LIST DIAGPACK DETAIL
name: dbm04adm01_2019_08_07T12_28_23_17_1.tar.bz2
alertDescription: "Hard disk of size 600GB in slot 5 failed"
alertName: 17_1
diagPackHostName: dbm04adm01
```

```

packStartTime: 2016-06-07T12:28:23-07:00

name: dbm04adm01_2019_08_11T12_51_05_16_1.tar.bz2
alertDescription: "InfiniBand Port HCA-4:1 may require attention.
State:Down, Physical State:Disabled."
alertName: 16_1
diagPackHostName: dbm04adm01
packStartTime: 2016-04-11T12:51:05-07:00

name: dbm04adm01_diag_2019_08_12T18_37_30_1
alertDescription: Processing...

```

### Related Topics

- [CREATE DIAGPACK](#)

## 9.7.7.5 LIST IBPORT

### Purpose

The `LIST IBPORT` command displays attributes for InfiniBand Network Fabric ports determined by the specified attributes and filters.



#### Note:

This command does not work on Exadata Database Machine servers with RoCE Network Fabric.

### Syntax

```
LIST IBPORT [name | attribute_filters] [attribute_list] [DETAIL]
```

### Usage Notes

The list of attributes that can be displayed is shown in [Example 9-44](#).

If the `activeSlave` attributes for both InfiniBand Network Fabric ports on a server are not listed, then active-active bonding is being used. If an `activeSlave` attribute is `TRUE`, then active-passive bonding is being used.

#### Example 9-74 Listing IBPORT Attributes

This example demonstrates the output you might see from the `LIST IBPORT` command.

```

DBMCLI> LIST IBPORT
 HCA-1:1 Active
 HCA-1:2 Active

DBMCLI> LIST IBPORT DETAIL
 name: HCA-1:1
 activeSlave: TRUE
 dataRate: "40 Gbps"
 hcaFWVersion: 2.7.0

```

```

id: 0x00212800013e8c67
lid: 20
linkDowned: 0
linkIntegrityErrs: 0
linkRecovers: 0
physLinkState: LinkUp
portNumber: 1
rcvConstraintErrs: 0
rcvData: 84653709
rcvErrs: 0
rcvRemotePhysErrs: 0
status: Active
symbolErrs: 0
vll5Dropped: 0
xmtConstraintErrs: 0
xmtData: 84572496
xmtDiscards: 0

name: HCA-1:2
activeSlave: FALSE
dataRate: "40 Gbps"
hcaFWVersion: 2.7.0
id: 0x00212800013e8c68
lid: 21
linkDowned: 0
linkIntegrityErrs: 0
linkRecovers: 0
physLinkState: LinkUp
portNumber: 2
rcvConstraintErrs: 0
rcvData: 79355427
rcvErrs: 0
rcvRemotePhysErrs: 0
status: Active
symbolErrs: 0
vll5Dropped: 0
xmtConstraintErrs: 0
xmtData: 79274016
xmtDiscards: 0

```

### 9.7.7.6 LIST LUN

#### Purpose

The LIST LUN command displays attributes for LUNs determined by the specified attributes and filters.

#### Syntax

```
LIST LUN [name | attribute_filters] [attribute_list] [DETAIL]
```

#### Usage Notes

The list of attributes that can be displayed is shown in [Example 9-45](#).

## Examples

[Example 9-75](#) shows the LIST command with the LUN object, and the corresponding output.

### Example 9-75 Listing LUN Attributes

```

DBMCLI> LIST LUN
 0_0 0_0 normal

DBMCLI> LIST LUN 0_0 DETAIL
 name: 0_0
 diskType: HardDisk
 id: 0_0
 lunSize: 556.9289999008179G
 lunUID: 0_0
 raidLevel: 5
 lunWriteCacheMode: "WriteBack, ReadAheadNone, Direct, \
 No Write Cache if Bad BBU"
 status: normal

```

## 9.7.7.7 LIST METRICCURRENT

### Purpose

The LIST METRICCURRENT command displays a list of collections of all metrics.

### Syntax

```
LIST METRICCURRENT [name | attribute_filters] [attribute_list] [DETAIL]
```

### Usage Notes

- The list of attributes that can be displayed is shown in [Example 9-46](#).
- To reduce the size of the output when you run the LIST METRICCURRENT command, use filters.

### Examples

[Example 9-76](#) shows the LIST command with filters to display information about the METRICCURRENT object, and the corresponding output.

### Example 9-76 Listing METRICCURRENT Attributes

```

DBMCLI> LIST METRICCURRENT attributes name, objecttype
 DS_BBU_CHARGE DBSERVER
 DS_BBU_TEMP DBSERVER
 DS_CPU DBSERVER
 DS_CPU_MS DBSERVER
 DS_FANS DBSERVER
 DS_FSUT DBSERVER_FILESYSTEM
 DS_FSUT DBSERVER_FILESYSTEM
 DS_FSUT DBSERVER_FILESYSTEM
 DS_MEMUT DBSERVER

```

DS_MEMUT_MS	DBSERVER
DS_RUNQ	DBSERVER
DS_SWAP_IN_BY_SEC	DBSERVER
DS_SWAP_OUT_BY_SEC	DBSERVER
DS_SWAP_USAGE	DBSERVER
DS_TEMP	DBSERVER
DS_VIRTMEM_MS	DBSERVER
N_HCA_MB_RCV_SEC	DBSERVER
N_HCA_MB_TRANS_SEC	DBSERVER
N_IB_MB_RCV_SEC	IBPORT
N_IB_MB_RCV_SEC	IBPORT
N_IB_MB_TRANS_SEC	IBPORT
N_IB_MB_TRANS_SEC	IBPORT
N_IB_UTIL_RCV	IBPORT
N_IB_UTIL_RCV	IBPORT
N_IB_UTIL_TRANS	IBPORT
N_IB_UTIL_TRANS	IBPORT
N_NIC_KB_RCV_SEC	DBSERVER
N_NIC_KB_TRANS_SEC	DBSERVER
N_NIC_NW	DBSERVER

DBMCLI> LIST METRICCURRENT where objectType='IBPORT'

N_IB_MB_RCV_SEC	HCA-1:1	0.026 MB/sec
N_IB_MB_RCV_SEC	HCA-1:2	0.009 MB/sec
N_IB_MB_TRANS_SEC	HCA-1:1	0.021 MB/sec
N_IB_MB_TRANS_SEC	HCA-1:2	0.004 MB/sec
N_IB_UTIL_RCV	HCA-1:1	0.0 %
N_IB_UTIL_RCV	HCA-1:2	0.0 %
N_IB_UTIL_TRANS	HCA-1:1	0.0 %
N_IB_UTIL_TRANS	HCA-1:2	0.0 %

DBMCLI> LIST METRICCURRENT where name = DS\_CPUT and metricobjectname= myexadb04 detail

name:	DS_CPUT
alertState:	normal
collectionTime:	2014-03-13T16:15:25-05:00
metricObjectName:	myexadb04
metricType:	Instantaneous
metricValue:	0.9 %
objectType:	DBSERVER

DBMCLI> LIST METRICCURRENT DS\_FSUT

DS_FSUT	/	50 %
DS_FSUT	/boot	18 %
DS_FSUT	/u01	82 %

## Related Topics

- [Attribute Filters in LIST and ALTER Commands](#)  
You can use the *attribute\_filters* clause to specify the objects to display in LIST commands. Some ALTER commands also support the *attribute\_filters* clause.



## 9.7.7.8 LIST METRICDEFINITION

### Purpose

The `LIST METRICDEFINITION` command displays a list of metric definitions on the database server.

### Syntax

```
LIST METRICDEFINITION [name | attribute_filters] [attribute_list] [DETAIL]
```

### Usage Notes

The list of attributes that can be displayed is shown in [Example 9-47](#).

### Examples

[Example 9-77](#) shows the `LIST` command with the `METRICDEFINITION` object, and the corresponding output.

#### Example 9-77 Listing METRICDEFINITION Attributes

```
DBMCLI> LIST METRICDEFINITION
 DS_BBU_CHARGE
 DS_BBU_TEMP
 DS_CPUT
 DS_CPUT_MS
 DS_FANS
 DS_FSUT
 DS_MEMUT
 DS_MEMUT_MS
 DS_RUNQ
 DS_SWAP_IN_BY_SEC
 DS_SWAP_OUT_BY_SEC
 DS_SWAP_USAGE
 DS_TEMP
 DS_VIRTMEM_MS
 N_HCA_MB_RCV_SEC
 N_HCA_MB_TRANS_SEC
 N_IB_MB_RCV_SEC
 N_IB_MB_TRANS_SEC
 N_IB_UTIL_RCV
 N_IB_UTIL_TRANS
 N_NIC_KB_RCV_SEC
 N_NIC_KB_TRANS_SEC
 N_NIC_NW
```

## 9.7.7.9 LIST METRICHISTORY

### Purpose

The `LIST METRICHISTORY` command displays a list of individual metrics.

### Syntax

```
LIST METRICHISTORY [name | attribute_filters] [attribute_list]
 {over_specification} [MEMORY] [DETAIL]
```

## Usage Notes

- The list of attributes that can be displayed is shown in [Example 9-48](#).
- The retention period for metric history entry files is specified by the `metricHistoryDays` database server attribute. You can modify this setting with the DBMCLI `ALTER DBSERVER` command.

- The *over\_specification* syntax is as follows:

```
OVER number [aggregation_type [aggregation_type]...]
```

In the preceding syntax, *number* is amount of time in minutes for the aggregation, and *aggregation\_type* can be `max`, `min`, or `avg`.

- A WHERE clause can include the `ageInMinutes` attribute to specify the list is limited to those metrics which have the specified age. For example, the following command would show the metrics created in the previous 15 minutes:

```
DBMCLI> LIST METRIC HISTORY WHERE ageInMinutes < 15
```

## Examples

[Example 9-78](#) shows the `LIST METRIC HISTORY` command with the `name` and `collectionTime` attributes.

[Example 9-79](#) shows the `LIST METRIC HISTORY` command with the `ds_cput` attribute.

### Example 9-78 Listing METRIC HISTORY Using the Name and CollectionTime Attributes

```
DBMCLI> LIST METRIC HISTORY WHERE name LIKE 'DS_.*' AND collectionTime >
'2014-03-27T17:48:16-05:00'
```

DS_FANS	myexadb03	16		2014-03-27T17:49:15-05:00
DS_TEMP	myexadb03	20.0 C		2014-03-27T17:49:15-05:00
DS_BBU_CHARGE	myexadb03	52.0 %		2014-03-27T17:49:16-05:00
DS_BBU_TEMP	myexadb03	45.0 C		2014-03-27T17:49:16-05:00
DS_CPUT	myexadb03	0.3 %		2014-03-27T17:49:16-05:00
DS_CPUT_MS	myexadb03	0.0 %		2014-03-27T17:49:16-05:00
DS_FSUT	/	73 %		2014-03-27T17:49:16-05:00
DS_FSUT	/boot	18 %		2014-03-27T17:49:16-05:00
DS_FSUT	/u01	43 %		2014-03-27T17:49:16-05:00
DS_MEMUT	myexadb03	42 %		2014-03-27T17:49:16-05:00
DS_MEMUT_MS	myexadb03	0.3 %		2014-03-27T17:49:16-05:00
DS_MEMUT_MS	myexadb03	0.3 %		2014-03-27T17:49:16-05:00
DS_RUNQ	myexadb03	0.2		2014-03-27T17:49:16-05:00
DS_SWAP_IN_BY_SEC	myexadb03	0.0 KB/sec		2014-03-27T17:49:16-05:00
DS_SWAP_OUT_BY_SEC	myexadb03	0.0 KB/sec		2014-03-27T17:49:16-05:00
DS_SWAP_USAGE	myexadb03	0 %		2014-03-27T17:49:16-05:00
DS_VIRTMEM_MS	myexadb03	981 MB		2014-03-27T17:49:16-05:00
DS_VIRTMEM_MS	myexadb03	981 MB		2014-03-27T17:49:16-05:00

### Example 9-79 Listing METRIC HISTORY for the ds\_cput Attribute

```
DBMCLI> LIST METRIC HISTORY ds_cput OVER 10 MIN MAX MEMORY
DS_CPUT myexadb03 1.0 % 2014-03-27T16:59:16-05:00 0.3 % 1.0
%
DS_CPUT myexadb03 0.7 % 2014-03-27T17:09:16-05:00 0.2 % 0.8
%
DS_CPUT myexadb03 0.7 % 2014-03-27T17:19:16-05:00 0.3 % 1.1
%
```

DS_CPU	myexadb03	0.7 %	2014-03-27T17:29:16-05:00	0.3 %	0.9 %
DS_CPU	myexadb03	0.7 %	2014-03-27T17:39:16-05:00	0.2 %	0.8 %
DS_CPU	myexadb03	0.8 %	2014-03-27T17:49:16-05:00	0.3 %	0.8 %

**Related Topics**

- [Attribute Filters in LIST and ALTER Commands](#)  
You can use the *attribute\_filters* clause to specify the objects to display in LIST commands. Some ALTER commands also support the *attribute\_filters* clause.
- [ALTER DBSERVER](#)  
The ALTER DBSERVER command changes the attributes of the database server.
- [DESCRIBE DBSERVER](#)

## 9.7.7.10 LIST METRICSTREAM

**Purpose**

The LIST METRICSTREAM command displays metrics in the real-time metric stream.

**Syntax**

```
LIST METRICSTREAM [name | attribute_filters] [attribute_list] [DETAIL]
```

**Usage Notes**

- The LIST METRICSTREAM command is functionally equivalent to the LIST METRICCURRENT command, except that LIST METRICSTREAM only displays metrics that are enabled for real-time metric streaming.

**Example 9-80 Listing METRICSTREAM Attributes**

This example shows the LIST METRICSTREAM command with a filter to display cell disk attributes.

```
DBMCLI> LIST METRICSTREAM WHERE name LIKE 'N_NIC.*'
```

**Related Topics**

- [Real-Time Insight](#)  
You can use the real-time insight feature to enable real-time monitoring of your Exadata systems.
- [Attribute Filters in LIST and ALTER Commands](#)  
You can use the *attribute\_filters* clause to specify the objects to display in LIST commands. Some ALTER commands also support the *attribute\_filters* clause.
- [LIST METRICCURRENT](#)

## 9.7.7.11 LIST PHYSICALDISK

**Purpose**

The LIST PHYSICALDISK command displays attributes for one or more physical disks determined by the specified attributes and filters.

### Syntax

```
LIST PHYSICALDISK [name | attribute_filters] [attribute_list] [DETAIL]
```

### Usage Notes

The list of attributes that can be displayed is shown in [Example 9-49](#).

### Examples

[Example 9-81](#) shows the `LIST` command with the `PHYSICALDISK` object, and the corresponding output.

#### Example 9-81 Listing Physical Disk Attributes

```
DBMCLI> LIST PHYSICALDISK
 252:0 G08VKE normal
 252:1 GYVAXE normal
 252:2 G07Z5E normal
 252:3 G0889E normal
```

## 9.7.7.12 LIST ROLE

### Purpose

The `LIST ROLE` command displays the specified attributes for a role.

### Syntax

```
LIST ROLE [name | filters] [attribute_list] [DETAIL]
```

### Usage Notes

- *name* is the name of the role.
- *filters* is an expression that determines which roles are displayed.
- *attribute\_list* is the attributes to display. The `ALL` option can be used to display all attributes.
- The `DETAIL` option formats the output as an attribute on each line, with an attribute descriptor preceding each value.

### Examples

[Example 9-82](#) shows the `LIST ROLE` command.

#### Example 9-82 Displaying a Role

```
DBMCLI> LIST ROLE DETAIL

DBMCLI> LIST ROLE where name like 'db_*
```

### 9.7.7.13 LIST SOFTWAREHISTORY

#### Purpose

The `LIST SOFTWAREHISTORY` command displays a list of final states for past software updates.

#### Syntax

```
LIST SOFTWAREHISTORY [attribute_filters] [attribute_list] [DETAIL]
```

#### Example 9-83 Displaying the History of a Scheduled Software Update

By default, only the update name and status are shown.

```
CellCLI> LIST SOFTWAREHISTORY
12.2.1.2.0.170509 Last update completed at: 2017-05-20T08:00:57-07:00
12.2.1.2.0.170520 Last update completed at: 2017-05-21T06:39:54-07:00
12.2.1.2.0.17052 Last update completed at: 2017-06-08T08:56:45-07:00
12.2.1.2.0.170603 Last update completed at: 2017-06-08T16:03:17-07:00
```

#### Example 9-84 Displaying the Detailed History of a Specific Software Update

This example shows the detailed software update history for a particular update which is referenced by its name.

```
CellCLI> LIST SOFTWAREHISTORY WHERE name='12.2.1.2.0.170808.1' DETAIL
name: 12.2.1.2.0.170808.1
status: Upgrade failed. See alerts at:
2017-08-10T10:56:15-07:00
```

### 9.7.7.14 LIST SOFTWAREUPDATE

#### Purpose

The `LIST SOFTWAREUPDATE` command displays the status of the most recently scheduled update.

#### Syntax

```
LIST SOFTWAREUPDATE [attribute_list] [DETAIL]
```

#### Usage Notes

The possible states for the software update are:

- Ready to update at: *update\_time*
- Downloading
- Checking prerequisites
- Prerequisites failed. See alerts.
- Last update completed at: *update\_time*

- Running
- Upgrade failed. See alerts.

### Example 9-85 Displaying the Status of a Scheduled Software Update

By default, only the update name and status are shown.

```
CellCLI> LIST SOFTWAREUPDATE
12.2.1.2.0.170603 Last update completed at: 2017-06-08 16:03:17
-0700
```

### Example 9-86 Displaying the Detailed Status of a Scheduled Software Update

By default, only the update name and status are shown.

```
CellCLI> LIST SOFTWAREUPDATE DETAIL
name: 12.2.1.2.0.170808.1
status: Upgrade failed. See alerts
store: https://mystore_url:4443
time: 2017-08-10T10:35:00-07:00
```

## 9.7.7.15 LIST THRESHOLD

### Purpose

The `LIST THRESHOLD` command displays attributes for one or more thresholds determined by the specified attributes and filters.

### Syntax

```
LIST THRESHOLD [name | attribute_filters] [attribute_list] [DETAIL]
```

### Usage Notes

The list of attributes that can be displayed is shown in [Example 9-53](#).

### Examples

[Example 9-87](#) shows the `LIST` command with the `THRESHOLD` object, and the corresponding output.

### Example 9-87 Listing Threshold Attributes

```
DBMCLI> LIST THRESHOLD
DS_CPU.T.myexadb04
DS_TEMP.myexadb04

DBMCLI> LIST THRESHOLD DS_CPU.T.myexadb04 detail
name: DS_CPU.T.myexadb04
comparison: >
warning: 90.0

DBMCLI> LIST THRESHOLD DS_TEMP.myexadb04 detail
name: DS_TEMP.myexadb04
comparison: >
critical: 30.0
```

## 9.7.7.16 LIST USER

### Purpose

The `LIST USER` command displays the specified attributes for a user.

### Syntax

```
LIST USER [name | filters] [attribute_list] [DETAIL]
```

### Usage Notes

- *name* is the user name.
- *filters* is an expression that determines which users are displayed.
- *attribute\_list* is the attributes to display. The `ALL` option can be used to display all attributes.
- The `DETAIL` option formats the output as an attribute on each line, with an attribute descriptor preceding each value.

### Examples

[Example 9-88](#) shows the `LIST USER` command.

#### Example 9-88 Using the LIST USER Command

```
DBMCLI> LIST USER DETAIL
```

```
DBMCLI> LIST USER where name like 'jdoe' DETAIL
name: jdoe
roles: role=db_monitor
Privileges: object=dbserver
 verb=list
 attributes=all attributes
 options= all options
```

## 9.7.8 REVOKE

### Purpose

The `REVOKE` command removes privileges and roles.

### Syntax

```
REVOKE object_type [name] FROM sub_object_type [sub_object_name]
```

### Usage Notes

- *object\_type* can be as follows:
  - PRIVILEGE
  - ROLE

- The following can be used for `PRIVILEGE` object type:
  - *name* is in the following format:
 

```
{ ALL ACTIONS | action } ON { ALL OBJECTS | object } { ALL
ATTRIBUTES | \
ATTRIBUTES attribute1 [, attribute2, ...] } { WITH ALL OPTIONS
| \
WITH OPTIONS option1 [, option2, ...] }
```
  - The *sub\_object\_type* must be `ROLE`.
  - The *sub\_object\_name* is a role name, a comma-delimited list of role names, or the keyword `ALL`.
- The following can be used for the `ROLE` object type:
  - *name* is a role name, a comma-delimited list of role names, or the keyword `ALL`.
  - The *sub\_object\_type* must be `USER`.
  - The *sub\_object\_name* is the name of the user, a comma-delimited list of user names, or the keyword `ALL`.
- [REVOKE PRIVILEGE](#)
- [REVOKE ROLE](#)

### 9.7.8.1 REVOKE PRIVILEGE

#### Purpose

The `REVOKE PRIVILEGE` command revokes privileges from a role.

#### Syntax

```
REVOKE PRIVILEGE { ALL ACTIONS | action } ON { ALL OBJECTS | object }
{ ALL ATTRIBUTES | ATTRIBUTES attribute1 [, attribute2...] } { WITH
ALL OPTIONS | WITH OPTIONS option1 [, option2, ...] } FROM ROLE { ALL
| role1 [, role2, ...] }
```

#### Usage Notes

- *action* is the command.
- *object* is object type for the action.
- *attribute* are the attributes for the object.
- *option* are the options for the object.
- *role* is the name of the role from which to revoke privileges.
- The `ALL ACTIONS` argument revokes privileges for all actions.
- The `ALL OBJECTS` argument revokes privileges for all objects.
- The `ALL ATTRIBUTES` argument revokes privileges for all attributes.
- The `WITH ALL OPTIONS` argument revokes privileges for all options.



## Examples

[Example 9-89](#) shows the `REVOKE PRIVILEGE` command.

### Example 9-89 Revoking a Privilege

```
DBMCLI> REVOKE PRIVILEGE ALL ACTIONS ON ALL OBJECTS ALL ATTRIBUTES -
 WITH ALL OPTIONS FROM ROLE ALL

DBMCLI> REVOKE PRIVILEGE list ON dserver ATTRIBUTES name -
 WITH OPTIONS detail FROM ROLE db_monitor
```

## 9.7.8.2 REVOKE ROLE

### Purpose

The `REVOKE ROLE` command revokes the role for a user.

### Syntax

```
REVOKE ROLE { ALL | role1 [, role2, ...] } FROM USER { ALL | user1 [, user2...] }
```

### Usage Notes

- *role1* and *role2* are the names of roles.
- *user1* and *user2* are the names of users.
- The `ALL` argument revokes all roles from the user.
- The `FROM USER ALL` argument revokes the role from all users.

### Examples

[Example 9-90](#) shows how to revoke a role from a user.

### Example 9-90 Revoking a Role From a User

```
DBMCLI> REVOKE ROLE db_monitor FROM USER jdoe
```

## 9.7.9 SET

### Purpose

The `SET` command sets a variable to alter the DBMCLI environment settings for the current session.

### Syntax

```
SET DATEFORMAT {LOCAL | STANDARD}
SET ECHO [ON | OFF]
```

## Usage Notes

The `SET DATEFORMAT` command controls the format of displayed dates. For commands that accept dates, the standard date-time format is recommended. The local format is also accepted. The standard format is recommended for scripts because that format is less sensitive to the time zone, region, and locale changes that might occur when running a script.

The `SET ECHO` command controls whether to echo commands in a script that is run with `@` or `START`. The `ON` option displays the commands on screen. The `OFF` option suppresses the display. The `SET ECHO` command does not affect the display of commands entered interactively or redirected from the operating system.

### Example 9-91 Setting the Date Format with the SET Command

```
SET DATEFORMAT STANDARD
```

## 9.7.10 SPOOL

### Purpose

The `SPOOL` command writes (spools) the results of commands to the specified file.

### Syntax

```
SPO[OL] [file_name [CRE[ATE] | REP[LACE] | APP[END]] | OFF]
```

### Usage Notes

If you issue `SPOOL file_name` with no option, then the output is spooled to that file whether or not the file already exists. The `REPLACE` option is the default behavior.

**Table 9-14** SPOOL Options

Option	Description
APPEND	Adds the results to the end of the file specified.
CREATE	Creates a new file with the name specified, and raises an error if the file exists.
<i>file_name</i>	Names the file to which the results are written. It can be specified with a fully-qualified path name, or with a partially-qualified path name relative to the current directory.
no option	Displays the name of the current spool target file, if any.
OFF	Stops writing (spooling) output to the file.
REPLACE	Replaces the contents of an existing specified file. If the file does not exist, then <code>REPLACE</code> creates the file. This is the default behavior.

## 9.7.11 START

### Purpose

The `START` command runs the DBMCLI commands in the specified script file.

### Syntax

```
STA[RT] file_name
```

### Usage Notes

The `START` option is *file\_name*. It is the name of the script file that contains the DBMCLI commands. If the file name does not include a fully-qualified path, then the DBMCLI utility searches for the file relative to the current directory.

The `START` command is useful when entering long or multiple DBMCLI commands. For example, all the commands in [Example 9-10](#) can be entered in a text file named `alter_dbserver`, then run as follows, assuming that the `alter_dbserver` file is in the current directory:

```
START alter_dbserver
```

# 10

## Using the ExaCLI Utility

The ExaCLI utility enables you to manage database and cell nodes remotely.

- [ExaCLI Command Reference](#)
- [ExaCLI Security](#)  
The authentication process is performed for every ExaCLI session.
- [Creating Users for Use with ExaCLI](#)  
To use ExaCLI, users must be created on the remote node, and the users must be assigned roles that have privileges appropriate for the user.
- [Differences in Commands Between ExaCLI and DBMCLI/CellCLI](#)
- [Certificates for ExaCLI](#)  
Security certificates allow the remote nodes to confirm their identity to ExaCLI.
- [Executing Commands Remotely Using URLs](#)  
In addition to executing commands on a remote node through `exacli`, you can also execute them through URLs.

### 10.1 ExaCLI Command Reference

ExaCLI is a command-line administration tool that runs on database and cell nodes and enables you to manage nodes remotely.

#### Overview

You can use ExaCLI to manage cell and database node configuration and objects in the remote node's environment. ExaCLI is installed when a cell or database node is imaged. ExaCLI supports the same command syntax as DBMCLI and CellCLI. The main difference is that ExaCLI manages cell and database nodes from a remote host whereas DBMCLI runs directly on a database node, and CellCLI runs directly on a cell node.

There are two main reasons to use ExaCLI:

- Many companies, especially government organizations, require root access or the SSH service on the cell and database nodes to be disabled as part of the compliance requirements.
- Service providers that host Exadata machines provide access to a virtual machine running on the cell and database node. Customers are not allowed to SSH into the cell or database node to manage the node using CellCLI or DBMCLI.

For such cases, the only way to manage the cell or database node is by using ExaCLI running on a remote node.

#### Prerequisites

- Users created on the cell or database node that have been granted the necessary privileges. See [Creating Users for Use with ExaCLI](#) for more information.
- Java version 1.8 or later

You can determine the version of Java by running the `java -version` command. In addition, the `JAVA_HOME` environment variable must be set to point to the installation directory of the proper version of Java.

### File Path

```
/usr/local/sbin/exacli
```

### Syntax

You run ExaCLI from the operating system command line.

```
exacli -c [username@]remotehost[:port] [-l username] [--xml]
 [--cookie-jar [filename]] [-e {command | 'command[; command]' |
 @batchfile}]
```

### Options

Option	Description
<code>-c [username@]remotehost</code> or <code>--connect [username@]remotehost[:port]</code>	Specifies the remote node to which you want to connect. ExaCLI prompts for the user name if not specified.  ExaCLI can connect to both cells and compute nodes. The default port for compute node MS is 7879.  You can specify the remote host using an IPv6 address. The IPv6 addresses must be enclosed in square brackets and single quotes as shown in <a href="#">Example 10-2</a> .
<code>-l username</code> or <code>--login-name username</code>	Specifies the user name to log into the remote node.
<code>--xml</code>	Displays the output in XML format.
<code>--cookie-jar [filename]</code>	Specifies the filename of the cookie jar to use. If <i>filename</i> is not specified, the cookie is stored in a default cookie jar located at <i>HOME/.exacli/cookiejar</i> , where <i>HOME</i> is the home directory of the OS user running the ExaCLI command.  The presence of a valid cookie allows the ExaCLI user to execute commands without requiring to login in subsequent ExaCLI sessions.
<code>-e command</code> or <code>-e 'command[; command]'</code> or <code>-e @batchFile</code>	Specifies either the ExaCLI commands to run or a batch file. ExaCLI exits after running the commands.  If specifying multiple commands to run, enclose the commands in single quotes to prevent the shell from interpreting the semi-colon.  Omit this option to start an interactive ExaCLI session.

Option	Description
<code>--cert-proxy proxy[:port]</code>	Specifies the proxy server to use when downloading certificates. If <i>port</i> is omitted, port 80 is used by default.
<code>-n</code> or <code>--no-prompt</code>	Suppresses prompting for user input.

### Usage Notes

- Notes for the `--cookie-jar` option:
  - The user name and password are sent to the remote node for authentication. On successful authentication, the remote node issues a cookie (the login credentials) that is stored in the specified *filename* on the database node. If *filename* is not specified, the cookie is stored in a default cookie jar located at `HOME/.exaccli/cookiejar`, where *HOME* is the home directory of the operating system user running the ExaCLI command.
  - Permissions for the cookie jar file are set to `rw- --- --- (600)`.
  - The operating system user running the ExaCLI command is the owner of the cookie-jar file.
  - A cookie jar can contain multiple cookies from multiple users on multiple nodes in parallel sessions.
  - Cookies are invalidated after 24 hours.
  - If the cookie is not found or is no longer valid, ExaCLI prompts for the password. The new cookie is stored in the cookie jar identified by *filename*, or the default cookie jar if *filename* is not specified.
  - Even without the `--cookie-jar` option, ExaCLI still checks for cookies from the default cookie jar. However, if the cookie does not exist or is no longer valid, the new cookie will *not* be stored in the default cookie jar if the `--cookie-jar` option is not specified.
- Notes for the `-e` option:
  - ExaCLI exits after running the commands.
  - If specifying multiple commands to run, be sure to enclose the commands in single quotes to prevent the shell from interpreting the semi-colon.
  - The batch file is a text file that contains one or more ExaCLI commands to run.
- Notes for the `-n` (`--no-prompt`) option:
  - If ExaCLI needs additional information from the user, for example, if ExaCLI needs to prompt the user for a password (possibly because there were no valid cookies in the cookie-jar) or to prompt the user to confirm the remote node's identity, then ExaCLI prints an error message and exits.

### Examples

The following examples show how use ExaCLI after the users have been created and granted the necessary privileges. See [Creating Users for Use with ExaCLI](#) for examples of creating users and assigning privileges to them.

### Example 10-1 Connecting to a User on a Cell using ExaCLI

Connect to cell node `cellnode01` as the `celladministrator` user. ExaCLI prompts for a password if the default cookie jar does not contain a valid cookie for the `celladministrator` user. Because the `--cookie-jar` option is not specified, the cookie is not stored in the cookie jar.

```
$ exacli -l celladministrator -c cellnode01
```

The following command is equivalent.

```
$ exacli -c celladministrator@cellnode01
```

### Example 10-2 Connecting to a Remote Host Using an IPv6 Address

You can specify the remote host using an IPv6 address. The IPv6 addresses must be enclosed in square brackets and single quotes.

```
$ exacli -c 'scott@[2001:db8:a0b:12f0::1]'
```

```
$ exacli -c '[2001:db8:a0b:12f0::1]' -l scott
```

### Example 10-3 Creating a Cookie for ExaCLI Connections

The user `celladministrator` and password are sent to `cellnode01` for authentication. On successful authentication, the cell node sends back a cookie which is stored in the default cookie jar. The specified commands are then executed on the cell node. Note that

**Tip:**

Multiple commands must be enclosed in single quotes.

```
$ exacli -l celladministrator -c cellnode01 --cookie-jar -e 'list
cell; list celldisk'
Password=*****
```

### Example 10-4 Creating a Cookie for ExaCLI Connections

The user `celladministrator` and password are sent to `cellnode01` for authentication. On successful authentication, the cell node sends back a cookie which is stored in the default cookie jar. The specified commands are then executed on the cell node.

**Tip:**

Multiple commands must be enclosed in single quotes.

```
$ exacli -l celladministrator -c cellnode01 --cookie-jar -e 'list cell; list
celldisk'
Password=*****
```

**Example 10-5 Using a Cookie when Executing ExaCLI Commands**

After creating a cookie, as shown in the previous example, ExaCLI does not prompt for password for subsequent ExaCLI sessions for the `celladministrator` user because it uses the cookie from the default cookie jar.

```
$ exacli -l celladministrator -c cellnode01 -e list griddisk detail
```

If you execute `connect` as a user other than `celladministrator`, and a cookie does not exist for that user, you are prompted for a password.

```
$ exacli -c user1@cellnode01 -e list griddisk detail
Password=*****
```

**Example 10-6 Viewing DBSERVER Details using ExaCLI**

In this example, ExaCLI connects to the `dbnode01` node as the `dbnodeadministrator` user and lists the DBSERVER object attributes in detail.

```
$ exacli -l dbnodeadministrator -c dbnode01 --cookie-jar -e list dbserver
detail
```

The login and password are sent to the database node for authentication. On successful authentication, the database node returns a cookie with the login credentials which gets stored in the cookie jar on the machine running ExaCLI

**Example 10-7 Viewing the ALERTHISTORY for a DBSERVER using ExaCLI**

In this example, ExaCLI connects to the `dbnode01` node as the `dbnodeadministrator` user and lists the ALERTHISTORY object. ExaCLI does not prompt for password if there is a valid cookie in the default cookie jar.

```
$ exacli -c dbnodeadministrator@dbnode01 -e list alerthistory
```

**Example 10-8 Using a Command Batch File with ExaCLI**

In this example, the file `commandFile` is stored in the local directory (from where the ExaCLI command is called) and contains the following information:

```
list dbserver detail
list alerthistory
```



ExaCLI connects as the `dbnodeadministrator` user to the `dbnode01` database node, and, on successful authentication, runs the commands in `commandFile`.

```
$ exacli -l dbnodeadministrator -c dbnode01 -e @commandFile
```

## 10.2 ExaCLI Security

The authentication process is performed for every ExaCLI session.

Because ExaCLI runs DBMCLI and CellCLI commands on a remote node, user access must be authenticated before the commands can be executed. ExaCLI connects to an Oracle Exadata System Software user on the remote node that has been granted the necessary privileges to run the specified commands.

The presence of a valid cookie allows the ExaCLI user to execute commands without requiring to log in for each session. A **cookie** is a token that contains the login credentials issued by the remote node and stored on the host machine running ExaCLI. A cookie issued by the remote node is invalidated after 24 hours. A repository of cookies is called a cookie jar.

The same cookie jar can be used by multiple simultaneous ExaCLI sessions connecting to multiple remote nodes, even when they are using different login names. For example, `exadcli` connects to multiple remote nodes using multiple ExaCLI sessions, but they all use the same cookie jar.

The cookie jar is configured with read and write permissions for the operating system user. The file is not readable, writable, or executable for anyone else.

## 10.3 Creating Users for Use with ExaCLI

To use ExaCLI, users must be created on the remote node, and the users must be assigned roles that have privileges appropriate for the user.

- Use DBMCLI or CellCLI on the remote node to create the users and roles, and grant privileges.

### Note:

When creating a user on the cell or database node, the user password must be 8 to 40 alphanumeric characters or special characters (`!@#$%^&*()_-`) with at least one digit, one lowercase letter and one uppercase letter. The new password cannot be the same as the current password for the user.

### Example 10-9 Create and Configure a User With All Privileges for Use with ExaCLI

Create the users and roles on the remote node using DBMCLI (for database nodes) or CellCLI (for cell nodes). These are the users that are allowed to connect to the remote node using ExaCLI. You use roles to assign privileges to users.

- **Database node example:**

The following example creates an administrator role and grants all privileges to that role. It then creates a user called `dbadministrator` and grants the administrator role to that user.

```
DBMCLI> CREATE ROLE administrator
DBMCLI> GRANT PRIVILEGE ALL ACTIONS ON ALL OBJECTS ALL ATTRIBUTES -
WITH ALL OPTIONS TO ROLE administrator
DBMCLI> CREATE USER dbnodeadministrator password=*
DBMCLI> GRANT ROLE administrator TO USER dbnodeadministrator
```

- **Cell node example:**

This following example creates an administrator role and grants all privileges to that role. It then creates a user called `celladministrator` and grants the administrator role to that user:

```
CellCLI> CREATE ROLE administrator
CellCLI> GRANT PRIVILEGE ALL ACTIONS ON ALL OBJECTS ALL ATTRIBUTES -
WITH ALL OPTIONS TO ROLE administrator
CellCLI> CREATE USER celladministrator password=*
CellCLI> GRANT ROLE administrator TO USER celladministrator
```

#### **Example 10-10 Creating a Cell Node User for Cell Monitoring with ExaCLI**

This example creates a user called `cellmon` who has privileges to only view object details.

```
CellCLI> CREATE ROLE monitor
CellCLI> GRANT PRIVILEGE list ON ALL OBJECTS ALL ATTRIBUTES -
WITH ALL OPTIONS TO ROLE monitor
CellCLI> CREATE USER cellmon password=*
CellCLI> GRANT ROLE monitor TO USER cellmon
```

#### **Example 10-11 Creating a Database Node User for Node Monitoring with ExaCLI**

This example creates a user called `dbnodemon` that has privileges to view object details.

```
DBMCLI> CREATE ROLE monitor
DBMCLI> GRANT PRIVILEGE list ON ALL OBJECTS ALL ATTRIBUTES -
WITH ALL OPTIONS TO ROLE monitor
DBMCLI> CREATE USER dbnodemon password=*
DBMCLI> GRANT ROLE monitor TO USER dbnodemon
```

#### **Related Topics**

- [Creating Exadata Software Users and Roles](#)

## 10.4 Differences in Commands Between ExaCLI and DBMCLI/CellCLI

ExaCLI supports the exact same command syntax as DBMCLI and CellCLI.

Not all CellCLI commands can be executed through ExaCLI. The following commands are not supported on ExaCLI:

- Restarting cell services (the `restart` option in the `ALTER CELL` command)
- Starting cell services (the `startup` option in the `ALTER CELL` command)
- Shutting down cell services (the `shutdown` option in the `ALTER CELL` command)
- User management commands (`CREATE USER`, `ALTER USER`, `DROP USER`)
- Role management commands (`CREATE ROLE`, `DROP ROLE`, `GRANT ROLE`, `REVOKE ROLE`)
- Privilege management commands (`GRANT PRIVILEGE`, `REVOKE PRIVILEGE`)
- `CALIBRATE`
- `DESCRIBE`
- `HELP`
- Getting the following cell attributes: `rsStatus`, `cellsrvStatus`, `msStatus` using the `LIST CELL` command
- `SET`
- `SPOOL`
- `START`

## 10.5 Certificates for ExaCLI

Security certificates allow the remote nodes to confirm their identity to ExaCLI.

All communication between ExaCLI and Management Server (MS) running on the remote node is over HTTPS. Management Server is deployed with a default self-signed security certificate for HTTPS access. Optionally, you can upload a different security certificate issued by a Certificate Authority (CA).

Trusted CA Certificates can be listed by running the following command on the machine where ExaCLI is run:

```
$JAVA_HOME/bin/keytool -list -v -keystore $JAVA_HOME/jre/lib/security/cacerts -storepass password
```

- [Using the Default Self-Signed Certificate](#)  
A security certificate must be signed by a trusted certificate authority (CA) for ExaCLI to accept it without a warning.
- [Using a CA-Certified Security Certificate](#)  
The Oracle Exadata Database Machine administrator can supply CA-certified security certificates, if they are needed for your environment.

### 10.5.1 Using the Default Self-Signed Certificate

A security certificate must be signed by a trusted certificate authority (CA) for ExaCLI to accept it without a warning.

When ExaCLI connects to a remote node that does not have a security certificate signed by a trusted CA, a warning message is generated requesting the user to verify the certificate. When verified, ExaCLI remembers the security certificate and the users are not prompted to verify the certificate again for subsequent commands. Every cell

and compute node is automatically configured with a self-signed certificate during install or Management Server (MS) deployment.

**Note:**

If you are using Oracle Exadata System Software release 19.3 or later, then enter **y** or **n** followed by pressing **Enter** when prompted.

Example of accepting the self-signed security certificate (non-CA certificate):

```
$ exacli -l celladministrator --cookie-jar -c cellnode01
This connection is unsecure. You have asked ExaCLI to connect to cell
cellnode01 securely.
The identity of cellnode01 cannot be verified.
Got certificate from server:
CN=cellnode01,OU=Oracle Exadata,O=Oracle Corporation,L=Redwood
City,ST=California,C=US
Do you want to accept and store this certificate? (Press y/n)
y
```

Example of accepting the self-signed security certificate (non-CA certificate) for a database node:

```
$ exacli -l dbadministrator --cookie-jar -c dbnode01
This connection is unsecure. You have asked ExaCLI to connect to dbserver
dbnode01 securely.
The identity of dbnode01 cannot be verified.
Got certificate from server:
CN=dbnode01,OU=Oracle Exadata,O=Oracle Corporation,L=Redwood
City,ST=California,C=US
Do you want to accept and store this certificate? (Press y/n)
y
```

## 10.5.2 Using a CA-Certified Security Certificate

The Oracle Exadata Database Machine administrator can supply CA-certified security certificates, if they are needed for your environment.

ExaCLI accepts valid CA-Certified security certificates without any prompts and without requiring you to accept the certificates. Each certificate consists of a key pair, which includes a public key and matching private key.

1. Modify security attributes for the storage server or database server.

To upload a CA-certified security certificate, modify the `securityPubKey` and the `securityPrivKey` attributes on the storage server or database server:

- `securityPubKey` specifies a URL for the certificate public key file.
- `securityPrivKey` specifies a URL for the certificate private key file.

The keys must be supplied as PEM-encoded files. Each URL can use the `http`, `https`, or `file` access scheme.

If the private key is encrypted, you can provide the password using the `securityPrivKeyPW` attribute.

The following examples show commands for uploading a CA-certified security certificate:

- Example for a storage server using `http` URLs to access the key files. In this example, the password prompt requests the password for the `celladministrator` user on `cellhost`.

```
$ exacli -l celladministrator -c cellhost -
-e 'alter cell securityPubKey="http://www.example.com/
security/newkey1.pem.crt", -
securityPrivKey="http://www.example.com/security/
newkey1-private.pem", -
securityPrivKeyPW="welcome1"'
```

Password: \*\*\*\*\*

...

- Example for a database server using local `file` URLs to access the key files. In this example, the password prompt requests the password for the `dbadministrator` user on `dbhost`.

```
$ exacli -l dbadministrator -c dbhost -
-e 'alter dbserver securityPubKey="file:///root/
security/newkey2.pem.crt", -
securityPrivKey="file:///root/security/newkey2-
private.pem", -
securityPrivKeyPW="welcome2"'
```

Password: \*\*\*\*\*

...

2. Restart the Management Server (MS) on the storage server or database server.

After you upload the CA-certified security certificate individually to every server in the cluster, you must restart MS before the new security certificate is visible.

For example:

- Restart MS on a storage server.

```
CELLCLI> alter cell restart services ms
```

Restarting MS services...

The RESTART of MS services was successful.

- Restart MS on a database server.

```
DBMCLI> alter dbserver restart services ms
```

Restarting MS services...

The RESTART of MS services was successful.

## 10.6 Executing Commands Remotely Using URLs

In addition to executing commands on a remote node through `exaccli`, you can also execute them through URLs.

Before using URLs to execute commands on the remote node, you have to create users, roles, and privileges, as described in [Creating Users for Use with ExaCLI](#).

The format of the URL is:

```
https://remotenode:port/MS/RESTService/?cmd=command_to_run
```

If the remote node is a cell node, the port is 443. If the remote node is a database node, the port is 7879.

When you submit the URL in a browser, the browser displays a warning about the certificate. If the warning is about a new certificate, the warning is usually acceptable. If the warning is about a changed certificate, you should make sure that the certificate is acceptable.

The browser then prompts for the user name and password. If you do not enter the correct password 5 times in succession, then the account is locked. You have to wait 30 minutes until you can try to login again to the same account.

The results of the command are returned in plain text.

If you submit further requests in the same session, the browser will not display warnings about the certificate and will not prompt for user name and password, unless the session cookie has expired.

The following example runs the `list dbserver detail` command:

```
https://dbnode:7879/MS/RESTService/?cmd=list+dbserver+detail
```

The following example runs the `list cell detail` command:

```
https://cellnode:443/MS/RESTService/?cmd=list+cell+detail
```

You can also specify the URL in a RESTful format. For example:

```
https://cellnode:443/MS/RESTService/celldisk/list?where=size>1g
```

You can also run the URL using `curl`. The steps are:

1. Download and store the server certificate.

```
(openssl s_client -showcerts -connect remote_node:port < /dev/null |
openssl x509 -outform PEM > mycertfile.pem) >& /dev/null
```

You need to do this only once. `curl` will not proceed unless you provide a valid certificate.

**2. Provide a user name and password.**

```
curl -u username:password -c cookiejar --cacert mycertfile.pem
'https://remote_node:port /MS/RESTService/login'
```

**3. Run the command.**

Database node example:

```
curl -b cookiejar --cacert mycertfile.pem 'https://
remote_node:port/MS/RESTService/?cmd=list+dbserver+detail'
```

Cell node example:

```
curl -b cookiejar --cacert mycertfile.pem 'https://
remote_node:port/MS/RESTService/?cmd=list+cell+detail'
```

# 11

## Using the `exadcli` Utility

The `exadcli` utility facilitates centralized management across an Oracle Exadata system by automating the execution of ExaCLI commands.

`exadcli` executes the commands on a set of remote nodes and returns the output to the centralized management location where the `exadcli` utility was run.

- [exadcli Command Reference](#)  
The `exadcli` utility runs commands on multiple remote nodes in parallel threads.
- [Using exadcli for the First Time](#)  
This is a multi-step example that demonstrates how to configure your Oracle Exadata Database Machine to use `exadcli`.
- [exadcli Examples](#)  
These examples show how use ExaCLI after the users have been created and granted the necessary privileges.

### 11.1 `exadcli` Command Reference

The `exadcli` utility runs commands on multiple remote nodes in parallel threads.

- [Overview](#)
- [Prerequisites](#)
- [File Path](#)
- [Syntax](#)
- [Options](#)
- [Usage Notes](#)

#### Overview

`exadcli` enables you to issue an ExaCLI command to be run on multiple remote nodes. Remote nodes are referenced by their host name or IP address. Unlike `dcli`, `exadcli` can only execute ExaCLI commands. Other commands, such as shell commands, cannot be executed using `exadcli`.

The `exadcli` utility is installed when the database node is imaged. `exadcli` does not support an interactive session with a remote application on the remote node.

`exadcli` connects in parallel to the remote nodes and authenticates the user. Command output (stdout and stderr) is collected and displayed after the copy and command execution finishes on the specified remote nodes.

`exadcli` runs ExaCLI in a “no-prompt” mode. This means that if user interaction is needed (for example, if you need to enter a password or if you need to verify that the certificate from a remote node is valid), then `exadcli` exits with an error. The command `exadcli --authenticate` can be used to log into every cell or database node using the same password.



The authentication token is stored in the cookie jar and future `exadcli` sessions will not need to enter an password.

### Prerequisites

- To use `exadcli`, you have to set up users and roles on all the remote nodes.
- Before using `exadcli`, you must use ExaCLI to accept the identity certificate of cell or database nodes. This needs to be done only once per cell (unless the cell is upgraded or a new certificate is uploaded to the remote database or cell node), for example:

#### Note:

If you are using Oracle Exadata System Software release 19.3 or later, then enter **y** or **n** followed by pressing **Enter** when prompted.

```
$ exadcli -l dbnodeadministrator -c dbnode01 --cookie-jar
No cookies found for dbnodeadministrator@dbnode01.
Password: *****
EXA-30016: This connection is not secure. You have asked ExaCLI to
connect to cell dbnode01:7879 securely. The identity of
dbnode01:7879 cannot be verified.
Got certificate from server:
C=US,ST=California,L=Redwood City,O=Oracle Corporation,OU=Oracle
Exadata,CN=localhost
Do you want to accept and store this certificate? (Press y/n) y
exadcli dbnodeadministrator@dbnode01:7879> exit
```

You must accept the certificate on every cell or database node. See [Using exadcli for the First Time](#) for a complete example of configuring access prior to using `exadcli`.

- Java version 1.8 or later  
You can determine the version of Java by running the `java -version` command. In addition, the `JAVA_HOME` environment variable must be set to point to the installation directory of the proper version of Java.

### File Path

```
/usr/local/sbin/exadcli
```

### Syntax

You run ExaCLI from the operating system command line.

```
exadcli options command[; command]
```

## Options

Option	Description
<code>-c host [,host]</code>	Specifies a comma-delimited list of target remote nodes to which ExaCLI commands are sent.
<code>--connect=host [,host]</code>	
<code>-g filename</code>	Specifies a file containing a list of target remote nodes to which commands are sent. The remote nodes can be identified by host names or IP addresses.
<code>--group-file=filename</code>	
<code>-l username</code>	Specifies the user to log in as on the remote nodes. This is a DBMCLI or CellCLI user, not an operating system user.
<code>--login-user=username</code>	
<code>-t</code>	Displays the target remote nodes that are named with the <code>-c</code> option or in the group file identified by the <code>-g</code> option.
<code>--list</code>	
<code>-x filename</code>	Specifies the command file to execute on the remote nodes. The command file contains a list of ExaCLI commands.
<code>--exec-file=filename</code>	<b>Note:</b> <code>exadcli</code> does not have an interactive mode. You can specify the commands to execute using the <code>-x</code> option, or you can specify the commands at the end of the command-line.
<code>--authenticate</code>	Prompts for the user password. The user will be authorized on all remote nodes using the password and the authorization is stored in the default cookie jar.
<code>--max-lines=maxlines</code>	Limits the maximum output lines from a remote node. Default is 100,000 lines.
<code>--xml</code>	Causes the output to be displayed in XML format. This option is passed through to ExaCLI.
<code>-h</code>	Displays the help message.
<code>--help</code>	

Option	Description
<code>command</code>	Specifies ExaCLI commands to execute on the remote nodes. To specify multiple commands, enclose the command string in single quotes and separate each command with the semicolon character <code>;</code> .
or	
<code>'command[; command]'</code>	<b>Note:</b> <code>exadcli</code> does not have an interactive mode. You can specify the commands to execute using the <code>-x</code> option, or you can specify the commands at the end of the command-line.
<code>--cookie-jar [filename]</code>	Specifies the cookie jar file. This option is passed through to ExaCLI. <b>Note:</b> This option is deprecated. Now, <code>exadcli</code> always uses the cookie jar.

### Usage Notes

- Return values from `exadcli` are:
  - 0: Commands were run successfully on all the remote nodes.
  - 1: One or more remote nodes could not be reached or remote execution returned a non-zero status.
  - 2: A local error prevented any command execution.
- If any remote nodes are down or do not respond, then a message is written to `stderr` listing the unresponsive remote nodes. The operation continues on other remote nodes, and the return code after completion is 1.
- If you restart the Management Server (MS), then you must re-authenticate `exadcli` using the `exadcli --authenticate` command.

```
$ exadcli -c dbnode01,dbnode02,dbnode03 -l dbnodeadministrator --
authenticate
```

### Related Topics

- [ExaCLI Command Reference](#)

## 11.2 Using exadcli for the First Time

This is a multi-step example that demonstrates how to configure your Oracle Exadata Database Machine to use `exadcli`.

This example shows how to configure access to cell nodes. The same procedure with minor changes can be used for configuring access to database nodes.

- Use CellCLI to create the `celladministrator` user on each node and grant privileges to those users.

Connect to each remote node and use CellCLI to create the user and assign privileges. In this example, the user is granted all privileges, but you can grant a

limited set of privileges that contains only the privileges needed to execute the `exadcli` commands.

```
CellCLI> CREATE ROLE administrator
```

```
CellCLI> GRANT PRIVILEGE ALL ACTIONS ON ALL OBJECTS ALL ATTRIBUTES -
> WITH ALL OPTIONS TO ROLE administrator
```

```
CellCLI> CREATE USER celladministrator password=*
```

```
CellCLI> GRANT ROLE administrator TO USER celladministrator
```

2. On the node where you plan to run `exadcli`, use `ExaCLI` to accept the security certificate for each node.

 **Note:**

If you are using Oracle Exadata System Software release 19.3 or later, then enter **y** or **n** followed by pressing **Enter** when prompted.

```
exadcli -l celladministrator -c cell01 --cookie-jar
```

```
No cookies found for celladministrator@cell01.
```

```
Password: *****
```

```
EXA-30016: This connection is not secure. You have asked ExaCLI to
connect to
```

```
cell cell01 securely. The identity of cell01 cannot be verified.
```

```
Got certificate from server:
```

```
C=US,ST=California,L=Redwood City,O=Oracle Corporation,OU=Oracle Exadata,
CN=localhost
```

```
Do you want to accept and store this certificate? (Press y/n)
```

```
y
```

```
exadcli celladministrator@cell01> exit
```

Repeat this step for both `cell02` and `cell03`.

3. Use `exadcli` to connect to cells `cell01`, `cell02`, and `cell03` as the `celladministrator` user with the `--authenticate` option.

```
$ exadcli -c cell01,cell02,cell03 -l celladministrator --authenticate
```

```
Password:
```

```
cell01: Password: *****
```

```
cell01: exadcli celladministrator@cell01:7879>
```

```
cell02: Password: *****
```

```
cell02: exadcli celladministrator@cell02:7879>
```

```
cell03: Password: *****
```

```
cell03: exadcli celladministrator@cell03:7879>
```

`exadcli` stores the credentials in `~/.exadcli/cookiejar` so that further `exadcli` commands to the same cells will not prompt for a password.

4. Optional: Create a `mycells` file as the `celladministrator` user that contains a list of the nodes on which you want to run the commands.

```
$ cat mycells
cell01
cell02
cell03
```

5. Use `exadcli` to run a command using the `mycells` file.

```
$ exadcli -g mycells -l celladministrator list cell

cell01: cell01 online
cell02: cell02 online
cell03: cell03 online
```

If you did not create the `mycells` file, you can instead specify the remote nodes on the command line using the `-c` option.

```
$ exadcli -c cell01,cell02,cell03 -l celladministrator list cell
detail
```

## 11.3 exadcli Examples

These examples show how use ExaCLI after the users have been created and granted the necessary privileges.

See [Creating Users for Use with ExaCLI](#) for examples of creating users and assigning privileges to them.

For an example of configuring `exadcli` access to storage cells for the first time, see [Using exadcli for the First Time](#).

### Example 11-1 Using the AUTHENTICATE Option to Re-Authenticate exadcli

Connect to nodes `dbnode01,dbnode02,dbnode03` as the `dbnodeadministrator` user and store the credentials in `~/.exacli/cookiejar` so that further `exadcli` commands to the same nodes do not prompt for a password.

```
$ exadcli -c dbnode01,dbnode02,dbnode03 -l dbnodeadministrator --
authenticate
```

### Example 11-2 Using exadcli to View the DBSERVER Information

You can use `exadcli` to execute a command on multiple node, such as listing DBSERVER information.

```
$ exadcli -c dbnode01,dbnode02,dbnode03 -l dbnodeadministrator list
dbserver
```

**Example 11-3 Using exadcli to View the Detailed Cell Information**

Using `exadcli` you can retrieve the detailed cell information for multiple nodes with a single command.

```
$ exadcli -c cell01,cell02,cell03 -l celladministrator list cell detail
```

**Example 11-4 Using a Node List and Command File with exadcli**

You can use input files with `exadcli` to avoid repeatedly entering the list of nodes to access, or to provide a list of ExaCLI commands to execute. This example assumes you have created a text file named `mycells` or `mydbnodes` that contains the host names of the remote nodes. This example also uses an input file named `cellcommands.txt` and `dbcommands.txt`.

```
$ cat cellcommands.txt
list cell
list griddisk attributes name, size order by size,name desc
list celldisk where status!=normal attributes name, status

$ exadcli -g mycells -l celladministrator -x cellcommands.txt
cell01: cell01 online
cell01: DBFS_DG_CD_11_cell01 29.125G
cell01: DBFS_DG_CD_10_cell01 29.125G
cell01: DBFS_DG_CD_09_cell01 29.125G
...
cell02: cell02 online
cell02: DBFS_DG_CD_11_cell02 29.125G
cell02: DBFS_DG_CD_10_cell02 29.125G
cell02: DBFS_DG_CD_09_cell02 29.125G
...
cell03: cell03 online
cell03: DBFS_DG_CD_11_cell03 29.125G
cell03: DBFS_DG_CD_10_cell03 29.125G
cell03: DBFS_DG_CD_09_cell03 29.125G

$ cat dbcommands.txt
list dbserver
list metriccurrent where objectType='IBPORT'
list physicaldisk

$ exadcli -g mydbnodes -l dbnodeadministrator -x dbcommands.txt
dbnode01: dbnode01 online
dbnode01: N_IB_MB_RCV_SEC HCA-1:1 0.061 MB/sec
dbnode01: N_IB_MB_RCV_SEC HCA-1:2 0.001 MB/sec
...
dbnode01: N_IB_UTIL_TRANS HCA-4:1 0.0 %
dbnode01: N_IB_UTIL_TRANS HCA-4:2 0.0 %
dbnode01: 39:0 71HKPE normal
dbnode01: 39:1 70CFBP normal
...
dbnode02: dbnode02 online
dbnode02: N_IB_MB_RCV_SEC HCA-1:1 0.023 MB/sec
dbnode02: N_IB_MB_RCV_SEC HCA-1:2 0.036 MB/sec
...
```

### Related Topics

- [exadcli Command Reference](#)