Sun Flash Accelerator F20 PCIe Card

Product Notes

Part No.: Part No.: E21357-10
March 2013

Copyright © 2010, 2013, Oracle and/or its affiliates. All rights reserved.

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

U.S. GOVERNMENT END USERS. Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. 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 and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon 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, Opteron, the AMD logo, and the AMD Opteron 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 on 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. 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.

Copyright © 2010, 2013, Oracle et/ou ses affiliés. Tous droits réservés.

Ce logiciel et la documentation qui l'accompagne sont protégés par les lois sur la propriété intellectuelle. Ils sont concédés sous licence et soumis à des restrictions d'utilisation et de divulgation. Sauf disposition de votre contrat de licence ou de la loi, vous ne pouvez pas copier, reproduire, traduire, diffuser, modifier, breveter, transmettre, distribuer, exposer, exécuter, publier ou afficher le logiciel, même partiellement, sous quelque forme et par quelque procédé que ce soit. Par ailleurs, il est interdit de procéder à toute ingénierie inverse du logiciel, de le désassembler ou de le décompiler, excepté à des fins d'interopérabilité avec des logiciels tiers ou tel que prescrit par la loi.

Les informations fournies dans ce document sont susceptibles de modification sans préavis. Par ailleurs, Oracle Corporation ne garantit pas qu'elles soient exemptes d'erreurs et vous invite, le cas échéant, à lui en faire part par écrit.

Si ce logiciel, ou la documentation qui l'accompagne, est concédé sous licence au Gouvernement des Etats-Unis, ou à toute entité qui délivre la licence de ce logiciel ou l'utilise pour le compte du Gouvernement des Etats-Unis, la notice suivante s'applique :

U.S. GOVERNMENT END USERS. Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

Ce logiciel ou matériel a été développé pour un usage général dans le cadre d'applications de gestion des informations. Ce logiciel ou matériel n'est pas conçu ni n'est destiné à être utilisé dans des applications à risque, notamment dans des applications pouvant causer des dommages corporels. Si vous utilisez ce logiciel ou matériel dans le cadre d'applications dangereuses, il est de votre responsabilité de prendre toutes les mesures de secours, de sauvegarde, de redondance et autres mesures nécessaires à son utilisation dans des conditions optimales de sécurité. Oracle Corporation et ses affiliés déclinent toute responsabilité quant aux dommages causés par l'utilisation de ce logiciel ou matériel pour ce type d'applications.

Oracle et Java sont des marques déposées d'Oracle Corporation et/ou de ses affiliés.Tout autre nom mentionné peut correspondre à des marques appartenant à d'autres propriétaires qu'Oracle.

Intel et Intel Xeon sont des marques ou des marques déposées d'Intel Corporation. Toutes les marques SPARC sont utilisées sous licence et sont des marques ou des marques déposées de SPARC International, Inc. AMD, Opteron, le logo AMD et le logo AMD Opteron sont des marques ou des marques déposées d'Advanced Micro Devices. UNIX est une marque déposée d'The Open Group.

Ce logiciel ou matériel et la documentation qui l'accompagne peuvent fournir des informations ou des liens donnant accès à des contenus, des produits et des services émanant de tiers. Oracle Corporation et ses affiliés déclinent toute responsabilité ou garantie expresse quant aux contenus, produits ou services émanant de tiers. En aucun cas, Oracle Corporation et ses affiliés ne sauraient être tenus pour responsables des pertes subies, des coûts occasionnés ou des dommages causés par l'accès à des contenus, produits ou services tiers, ou à leur utilisation.

Please
Recycle

Sun Flash Accelerator F20 PCIe Card Product Notes

This document contains late-breaking information about the Sun Flash Accelerator F20 PCIe card from Oracle. Read this document first. This document is written for technicians, system administrators, authorized service providers (ASPs), and users who have advanced experience troubleshooting and replacing hardware.

For specific installation instructions, see your system installation guide. For late-breaking information about installation and use of the Sun Flash Accelerator F20 PCIe card on your server, see the most recent version of the server product notes.

Supported Servers

This section describes servers that support the Sun Flash Accelerator F20 PCIe card. For detailed information about using this card with your server, see the product notes for your server, available at:

http://www.oracle.com/documentation

Note – The following servers are qualified to work with the Sun Flash Accelerator F20 PCIe card at the *initial release of the card*. Check your server product notes for confirmation that your server has subsequently been qualified.

SPARC Servers	x86 Servers	Other Systems
Sun SPARC Enterprise T5120	Sun Fire X4540	Sun Storage 7120*
Sun SPARC Enterprise T5220	Sun Fire X4170 M2	Exadata V2
Sun SPARC Enterprise T5140	Sun Fire X4270 M2	Exadata X2-2
Sun SPARC Enterprise T5240	Sun Fire X4470	Exadata X2-8

SPARC Servers	x86 Servers	Other Systems
Sun SPARC Enterprise T5440	Sun Fire X4470 M2	
Sun SPARC Enterprise M4000*		
Sun SPARC Enterprise M5000*		
Sun SPARC Enterprise M8000*		
Sun SPARC Enterprise M9000*		
SPARC T3-1		
SPARC T3-2		

* Note - The Sun Storage 7120 and the M series servers only support cards with part numbers: 511-1500-xx.

Supported Operating Systems

The following tables describe which combinations of operating systems and servers are supported for use with the Sun Flash Accelerator F20 PCIe card. Check marks indicate that combination is supported.

Supported OSs for SPARC Servers

SPARC Server	Oracle Solaris 10 SPARC			Oracle SPARC Solaris 11
	10/09	9/10	8/11	11.1
SPARC T3-1	✓	✓	✓	✓
SPARC T3-2	✓	✓	✓	✓
Sun SPARC Enterprise T5120	✓	✓	✓	✓
Sun SPARC Enterprise T5220	✓	✓	✓	✓
Sun SPARC Enterprise T5140	✓	✓	✓	✓
Sun SPARC Enterprise T5240	✓	✓	✓	✓
Sun SPARC Enterprise T5440	✓	✓	✓	✓
Sun SPARC Enterprise M4000	✓	✓	✓	✓
Sun SPARC Enterprise M5000	✓	✓	✓	✓
Sun SPARC Enterprise M8000	✓	✓	✓	✓
Sun SPARC Enterprise M9000	✓	✓	✓	✓

Supported OSs for x86 Servers

Operating System	Sun Fire X4170 M2	Sun Fire X4270 M2	Sun Fire X4470	Sun Fire X4470 M2	Sun Fire X4540
Oracle SPARC Solaris 11.1	✓	✓	✓	✓	✓
Oracle Solaris 10 10/09	✓	✓	✓		✓
Oracle Solaris 10 9/10	✓	✓	✓	✓	✓
Oracle Solaris 10 8/11	✓	✓	✓	✓	
Oracle Linux 5.4	✓	✓			
Oracle Linux 5.5	✓	✓	✓	✓	
Oracle Linux 5.6	✓	✓	✓	✓	
Oracle Linux 5.7					
Oracle Linux 6.0	✓	✓		✓	
Oracle Linux 6.1				✓	
Oracle Linux UEK	✓	✓		✓	
Windows 2008	✓	✓			
Windows 2008 SP2	✓	✓	✓		
Windows 2008 R2	✓	✓	✓	✓	
Windows 2008 R2 SP1	✓	✓		✓	
Windows 2008 R2 SP2				✓	
SuSE SLES 10 SP3	✓	✓	✓		
SuSE SLES 10 SP4	✓	✓		✓	
SuSE 11	✓	✓	✓	✓	
SuSE SLES11 SP1	✓	✓	✓	✓	
RedHat 5.4	✓	✓			
RedHat 5.5	✓	✓	✓	✓	
RedHat 5.6	✓	✓	✓	✓	
RedHat 5.7					
RedHat 6.0	✓	✓		✓	
RedHat 6.1				✓	
ESX/ESX 4.0 U1, U2	✓	✓	✓		
ESX/ESX 4.1 U1	✓	✓	✓	✓	

Operating System	Sun Fire X4170 M2	Sun Fire X4270 M2	Sun Fire X4470	Sun Fire X4470 M2	Sun Fire X4540
Oracle VM Server 2.2.1	✓	✓	✓	✓	
Oracle VM Server 2.2.2	✓	✓		✓	
Oracle VM Server 3.0	✓	✓			

Minimum Supported Card Firmware Versions

Firmware	Version
FMod firmware	D20R D20Y patch 145426-01
SAS/SATA controller firmware	1.27.03 1.27.92 with patch 147206-01

Required Host Software

The Sun Flash Accelerator F20 PCIe card runs with the minimum required host firmware listed below. However, to take advantange of the extended ESM life feature, use the recommended system software versions listed for your system. For more information, see "ESM Expected Life Extended" on page 8.

Host	Minimum Required Host Firmware Version with patches	Recommended System Software Version (patch no.)	Slots Supported for the F20 Card
SPARC T3-1		8.0.4.c (145665)	0, 2, 3, 5
SPARC T3-2		8.0.4.c (145667)	
Sun SPARC Enterprise T5120	Oracle ILOM 3.0 => 139439-39	7.3.0	0,2
Sun SPARC Enterprise T5220 (8 disk)	Oracle ILOM 3.0 => 139439-39	7.3.0	0,2,3,4,5
Sun SPARC Enterprise T5220 (16 disk)	Oracle ILOM 3.0 => 139439-39	7.3.0	0,2,5
Sun SPARC Enterprise T5140	Oracle ILOM 3.0 => 139444-05	7.3.1a	1,2
Sun SPARC Enterprise T5240	Oracle ILOM 3.0 => 139444-05	7.3.1a	2,5
Sun SPARC Enterprise T5440	Oracle ILOM 3.0 => 139444-05	7.3.1a	All
Sun SPARC Enterprise M4000/M5000	XCP 1100	XCP 1102	1,3
Sun SPARC Enterprise M8000/9000	XCP 1100	XCP 1102	All
Sun Fire X4540	Oracle ILOM 3.0.10.13 and BIOS 40	2.3.2	All
Sun Fire X4170 M2	Oracle ILOM 3.0.9.15	1.3.2	0, 2
Sun Fire X4270 M2	Oracle ILOM 3.0.9.15	1.3.2	80W CPU: 0,1,2,4 and 5 130W CPU: 0, 2 and 5
Sun Fire X4470	Oracle ILOM 3.0.14.10.a	See note below.	9,4,8,7, or 6
Sun Fire X4470 M2	Oracle ILOM 3.0.14.10.a	See note below.	9,4,8,7, or 6

Note – The Sun Fire X4470 and Sun Fire X4470 M2 servers do not support ESM monitoring. See the server product notes for manual ESM monitoring instructions for these systems.

Supported Configurations for Using the Card as a Boot Device

The server, firmware, and OS combinations that support using the card as a boot device are described in the following table.

Server	Minimum Supported Operating Systems	Supported Firmware Version
Sun Fire X4170 M2	• Oracle Solaris 10 5/08 x86 (64-bit) • Oracle Solaris 10 5/09 x86 (64-bit) • Oracle Linux 5.5 (64-bit) • Oracle Linux 5.4 (64-bit) • Windows 2008 (64-bit) • Windows 2008 R2 (64-bit)	1.27.92 1.27.92 1.27.03 1.27.03 1.27.03 1.27.03
Sun Fire X4270 M2	• Oracle Solaris 10 5/08 x86 (64-bit) • Oracle Solaris 10 5/09 x86 (64-bit) • Oracle Linux 6.0 (64-bit)	1.27.92 1.27.92 1.27.92
Sun Fire X4540	• Oracle Solaris 10 5/08 x86 (64-bit) • Oracle Solaris 10 5/09 x86 (64-bit)	1.27.92 1.27.92
Sun SPARC Enterprise T5120 Sun SPARC Enterprise T5220 Sun SPARC Enterprise T5140 Sun SPARC Enterprise T5240 Sun SPARC Enterprise T5440	• Oracle Solaris 10 5/08 SPARC (64-bit) • Oracle Solaris 10 5/09 SPARC (64-bit)	1.27.92 1.27.92

Oracle Solaris Performance Patch

Download and install the following performance patch on your Oracle Solaris host:

- For Oracle Solaris 10 SPARC U4-U7, 138880-01 or later with MPT patch 141444
- For Oracle Solaris 10 x86 U4-U7, 138881-01 or later with MPT patch 141737

Available for download from the following web site:

`http://www.oracle.com/us/support/index.html`

Note – To enable maximum throughput from the MPT driver, add `mpt_doneq_thread_n_prop=8;` to `/kernel/drv/mpt.conf` and reboot the system.

Management Pack 2.1 Required Patches

Note – Use Management Pack 2.2 for updating card firmware, if possible.

The Oracle Hardware Management Pack 2.1 (formerly known as Sun Server Hardware Management Pack) contains Oracle Server Management Agents, Oracle Hardware SNMP Plugins, and Oracle Server CLI Tools. Use this tool to update firmare on your Sun Flash Accelerator F20 PCIe card.

Supporting documentation is available at:
http://www.oracle.com/technetwork/documentation/sys-mgmt-networking-190072.html

Download software and patches from My Oracle Support:

http://support.oracle.com/

Once you have installed the Management Pack 2.1 software, download and install the patch required for your OS. For Oracle Solaris:

Patch Request 13175939 Oracle Hardware Management Pack 2.1 - Solaris

ESM Expected Life Extended

ESMs were originally assigned 2 year lifespans. Update your system host software to increase the ESM monitoring lifespan to 3 or 4 years, depending on your host. See "Required Host Software" on page 6 for system host software versions.

Note – When you replace an ESM, you must reset the counter to 0. For instructions, see the "Monitoring ESM Lifespan" section in the *Sun Flash Accelerator F20 PCIe Card User's Guide.*

Host System	Expected ESM Life
Exadata V2	3 years
Exadata X2-2	4 years
Exadata X2-8	4 years
All other supported servers.	3 years

FMod Alignment

You must properly align the FMods before using the card.

Solid state flash devices have block alignments typically aligned on 4-Kbyte boundaries, not the 512-byte boundaries of conventional disks. To maximize performance, partitions need to be aligned on 4-Kbyte boundaries. For directions on aligning the FMods, see the *Sun Flash Accelerator F20 PCIe User's Guide*.

Tuning ZFS

The following best practices are recommended in order to achieve optimal performance of ZFS on the F5100 storage array and F20 Flash Accelerator PCIe Card.

Disable ZFS Cache Flushing

Both the F20 and F5100 can be used as primary low-latency storage. In these cases it is recommended to make the necessary changes to sd.conf to ensure the device is properly treated as one with a non-volatile backed cache. This will improve latency of commit operations (such as regular O_DSYNC writes, NFS commits, and iSCSI operations).

ZFS sends cache flushes before completing commit operations. This is to instruct the storage to ensure the safety of previously completed I/O operations. The F5100 and F20 devices have built-in supercapacitors that provide enough power to ensure time to flush all device caches in the event of a power failure. Thus we need to instruct the F5100 and F20 to ignore these flushes. This setting can be changed only for F5100 and F20 devices by altering the /kernel/drv/sd.conf file.

The following is the entry that should be added (all one line, no line breaks):

```
sd-config-list = "ATA     MARVELL SD88SA02","throttle-max:32,
disksort:false, cache-nonvolatile:true";
```

In the above statement, spacing is absolutely critical; an extra or omitted space in the initial "ATA..." string will result in the directive having no impact.

Alternatively, you can choose to set zfs_nocacheflush on a system-wide basis via adding the following parameter to the /etc/system file. If you have devices in ZFS pools that have caches which are not protected with non-volatile memory, such as a SCSI disk, this is not recommended.

```
set zfs:zfs_nocacheflush = 1
```

Identify the Device(s) for the ZFS Intent Log

Consider using LUNs or low latency disk managed by a controller with persistent memory for the ZFS intent log if available. This option can be considered more cost effective than using flash for low latency commits. The size of the log devices must only be enough to hold 10 seconds of maximum write throughput. Examples would include a storage array based lun, or a disk fronted by an HBA with a battery protected write cache.

If no such device is available, segment a separate pool of F5100 FMods or F20 DOMs and specify as the log device.

The F20 supports 4 independent DOMs (Disks on Modules) of 24GB. It˜2019s ideal to use some DOMs as ZFS log devices to reduce the commit latency particularly if used in an NFS server. 1 single DOM of an F20 used as a ZFS log device will reduce latency of single lightly threaded operation by an order of magnitude (~10X). More F20 DOMs can be stripped together to achieved higher throughput for large amount of synchronous operation for instance when ZFS is exporting iSCSI luns.

Similarly, if using an F5100 both as primary storage and the ZFS log device, it is recommended to segment a few of the flash modules (FMods) for use as the log device. The FMods, like the DOMs, make excellent logging devices when used in this fashion.

An F20 DOM or F5100 FMOD cxtxdx can be added as a ZFS log device (man zpool(1M)) with:

```
zpool add pool log cxtxdx
```

Caution – Be very careful with zpool add commands. Mistakenly adding a log device as a normal pool device is a mistake that will require you to destroy and restore the pool from scratch.

Some of the F20 DOMs / F5100 FMODs not used as log devices can also be used as second level cache devices to both offload IOPS from primary disk storage and to improve read latency for commonly used data recently evicted from the level one ARC cache. In that case one can do this:

```
zpool add pool cache cxtxdx
```

Disable Metadata Compression

If you are running the latest S10 Patches or S11, this step is no longer necessary. ZFS now runs as a 4k-native file system F20 and F5100 devices. Enabling Metadata compression (default Solaris setting) is fine, and has run better test cases.

Flash products achieve optimal performance when subjected to I/O requests that are restricted in size and alignment to multiples of 4k. For this reason, it is recommended to disable metadata compression. Disable metadata compression by adding the following entry to /etc/system:

```
set zfs:zfs_mdcomp_disable = 1
```

Record Size

Large performance gains can be realized by reducing the default recordsize used by ZFS, particularly when running database workloads. The ZFS recordsize should match the database recordsize. Note that the recordsize setting must be in place prior to when data is loaded on zfs. Additional reading on this subject may be found at http://blogs.sun.com/roch/entry/tuning_zfs_recordsize. The following example illustrates how to set the recordsize to 16k:

```
zfs set recordsize=16k mypool/myfs
```

Known Issues

This section describes hardware, firmware, and software issues known to exist at this release of Oracle's Sun Flash Accelerator F20 PCIe card.

CR ID	Description	Workaround
6868341	In some cases, the server fails to recognize the card after a SATA controller firmware upgrade. The upgrade fails with an error message similar to the following: `MPT BIOS Fault 0Dh encountered at adapter PCI(02h,00h,00h)` In addition, the card does not appear in the server device tree.	Reboot the server a second time. The card is recognized after the second reboot.
6881281	FMod support clips may break if pressed too far during FMod removal.	Loosen the clip screws when replacing an FMod, as described in the *Sun Flash Accelerator F20 PCIe User's Guide*.
	Under heavy I/O load in Oracle Solaris, a large number of messages may be seen indicating I/O retries (such as incomplete read- retrying). These messages occur due to a hardware bug in the LSI SAS controller that may incorrectly detect underrun conditions and report them to the driver. Overall data integrity is not compromised, but a performance impact may be observed due to the messages logged and retries required.	Ignore the retry messages or use an external means to throttle I/O throughput down to a level where these messages are not being produced in great numbers. The method required to throttle I/O will be very configuration and workload specific.

CR ID	Description	Workaround
n/a	If there is a significant unexpected drop in card performance, the supercap module on the ESM should be checked.	Look at the back panel of the card to determine if the amber ESM Service Required LED is lit. Service the ESM as described in the *Sun Flash Accelerator F20 PCIe User's Guide*.
6881856	On occasion, the system BIOS may not give the card firmware enough time to load. When this situation occurs, the card cannot be seen by host server.	Reboot the server.
6951778	Repeated power cycling with minimum traffic causes FMod to zombie.	Do not run continuous power cycle tests with minimum traffic to the FMod. If power cycling is needed for qualification either ensure there is no traffic to the FMod including SMART transactions or allow the SuperCap to fully charge and run transactions to the FMod. Because the pre-existing state of a FMod is unknown, it could have been power cycled previously, the number of power cycles cannot be specified.

Documentation Errata

System Label Errata

The label on the card faceplate might read "SuperCap" instead of "ESM".

Getting Started Guide Errata

The printed *Sun Flash Accelerator F20 PCIe Getting Started Guide* (820-7264-11) that ships in the box with the card contains out of date URLs. For valid URLs, see the current *Sun Flash Accelerator F20 PCIe Getting Started Guide* (E29800-01) online at:

http://docs.oracle.com/cd/E19682-01/index.html