



# **Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS**



Sun Microsystems, Inc.  
4150 Network Circle  
Santa Clara, CA 95054  
U.S.A.

Part No: 817-0168-19  
June 2006, Revision A

Copyright 2006 Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 U.S.A. All rights reserved.

Sun Microsystems, Inc. has intellectual property rights relating to technology embodied in the product that is described in this document. In particular, and without limitation, these intellectual property rights may include one or more U.S. patents or pending patent applications in the U.S. and in other countries.

U.S. Government Rights – Commercial software. Government users are subject to the Sun Microsystems, Inc. standard license agreement and applicable provisions of the FAR and its supplements.

This distribution may include materials developed by third parties.

Parts of the product may be derived from Berkeley BSD systems, licensed from the University of California. UNIX is a registered trademark in the U.S. and other countries, exclusively licensed through X/Open Company, Ltd.

Sun, Sun Microsystems, the Sun logo, the Solaris logo, the Java Coffee Cup logo, docs.sun.com, Java, and Solaris are trademarks or registered trademarks of Sun Microsystems, Inc. in the U.S. and other countries. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the U.S. and other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc.

The OPEN LOOK and Sun™ Graphical User Interface was developed by Sun Microsystems, Inc. for its users and licensees. Sun acknowledges the pioneering efforts of Xerox in researching and developing the concept of visual or graphical user interfaces for the computer industry. Sun holds a non-exclusive license from Xerox to the Xerox Graphical User Interface, which license also covers Sun's licensees who implement OPEN LOOK GUIs and otherwise comply with Sun's written license agreements.

Products covered by and information contained in this publication are controlled by U.S. Export Control laws and may be subject to the export or import laws in other countries. Nuclear, missile, chemical or biological weapons or nuclear maritime end uses or end users, whether direct or indirect, are strictly prohibited. Export or reexport to countries subject to U.S. embargo or to entities identified on U.S. export exclusion lists, including, but not limited to, the denied persons and specially designated nationals lists is strictly prohibited.

DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

---

Copyright 2006 Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 U.S.A. Tous droits réservés.

Sun Microsystems, Inc. détient les droits de propriété intellectuelle relatifs à la technologie incorporée dans le produit qui est décrit dans ce document. En particulier, et ce sans limitation, ces droits de propriété intellectuelle peuvent inclure un ou plusieurs brevets américains ou des applications de brevet en attente aux Etats-Unis et dans d'autres pays.

Cette distribution peut comprendre des composants développés par des tierces personnes.

Certains composants de ce produit peuvent être dérivées du logiciel Berkeley BSD, licenciés par l'Université de Californie. UNIX est une marque déposée aux Etats-Unis et dans d'autres pays; elle est licenciée exclusivement par X/Open Company, Ltd.

Sun, Sun Microsystems, le logo Sun, le logo Solaris, le logo Java Coffee Cup, docs.sun.com, Java et Solaris sont des marques de fabrique ou des marques déposées de Sun Microsystems, Inc. aux Etats-Unis et dans d'autres pays. Toutes les marques SPARC sont utilisées sous licence et sont des marques de fabrique ou des marques déposées de SPARC International, Inc. aux Etats-Unis et dans d'autres pays. Les produits portant les marques SPARC sont basés sur une architecture développée par Sun Microsystems, Inc.

L'interface d'utilisation graphique OPEN LOOK et Sun a été développée par Sun Microsystems, Inc. pour ses utilisateurs et licenciés. Sun reconnaît les efforts de pionniers de Xerox pour la recherche et le développement du concept des interfaces d'utilisation visuelle ou graphique pour l'industrie de l'informatique. Sun détient une licence non exclusive de Xerox sur l'interface d'utilisation graphique Xerox, cette licence couvrant également les licenciés de Sun qui mettent en place l'interface d'utilisation graphique OPEN LOOK et qui, en outre, se conforment aux licences écrites de Sun.

Les produits qui font l'objet de cette publication et les informations qu'il contient sont régis par la législation américaine en matière de contrôle des exportations et peuvent être soumis au droit d'autres pays dans le domaine des exportations et importations. Les utilisations finales, ou utilisateurs finaux, pour des armes nucléaires, des missiles, des armes chimiques ou biologiques ou pour le nucléaire maritime, directement ou indirectement, sont strictement interdites. Les exportations ou réexportations vers des pays sous embargo des Etats-Unis, ou vers des entités figurant sur les listes d'exclusion d'exportation américaines, y compris, mais de manière non exclusive, la liste de personnes qui font objet d'un ordre de ne pas participer, d'une façon directe ou indirecte, aux exportations des produits ou des services qui sont régis par la législation américaine en matière de contrôle des exportations et la liste de ressortissants spécifiquement désignés, sont rigoureusement interdites.

LA DOCUMENTATION EST FOURNIE "EN L'ETAT" ET TOUTES AUTRES CONDITIONS, DECLARATIONS ET GARANTIES EXPRESSES OU TACITES SONT FORMELLEMENT EXCLUES, DANS LA MESURE AUTORISEE PAR LA LOI APPLICABLE, Y COMPRIS NOTAMMENT TOUTE GARANTIE IMPLICITE RELATIVE A LA QUALITE MARCHANDE, A L'APTITUDE A UNE UTILISATION PARTICULIERE OU A L'ABSENCE DE CONTREFACON.

# Contents

---

<b>Preface</b> .....	7
<b>1 Introduction to Sun Cluster Hardware</b> .....	15
Installing Sun Cluster Hardware .....	15
▼ Installing Sun Cluster Hardware .....	16
Maintaining Sun Cluster Hardware .....	17
Powering On and Off Sun Cluster Hardware .....	18
Dynamic Reconfiguration Operations For Sun Cluster Nodes .....	18
▼ DR Operations in a Cluster With DR-Enabled Servers .....	18
Local and Multihost Disks in a Sun Cluster Environment .....	19
Removable Media in a Sun Cluster Environment .....	19
SAN Solutions in a Sun Cluster Environment .....	20
Hardware Restrictions .....	20
<b>2 Installing and Configuring the Terminal Concentrator</b> .....	21
Installing the Terminal Concentrator .....	21
▼ How to Install the Terminal Concentrator in a Cabinet .....	22
▼ How to Connect the Terminal Concentrator .....	25
Configuring the Terminal Concentrator .....	25
▼ How to Configure the Terminal Concentrator .....	26
▼ How to Set Terminal Concentrator Port Parameters .....	28
▼ How to Correct a Port Configuration Access Error .....	29
▼ How to Establish a Default Route for the Terminal Concentrator .....	31
Using the Terminal Concentrator .....	33
▼ How to Connect to a Node's Console Through the Terminal Concentrator .....	33
▼ How to Reset a Terminal Concentrator Port .....	34

<b>3</b>	<b>Installing Cluster Interconnect Hardware and Configuring VLANs</b> .....	37
	Interconnect Requirements and Restrictions .....	37
	Cluster Interconnect Speed Requirements .....	37
	Ethernet Switch Configuration When in the Cluster Interconnect .....	38
	SPARC: Requirements When Using Jumbo Frames .....	38
	Restriction on SCI Card Placement .....	38
	Requirements and Restrictions When Using InfiniBand in the Cluster Interconnect .....	39
	Installing Ethernet or InfiniBand Cluster Interconnect Hardware .....	39
	▼ How to Install Ethernet or InfiniBand Transport Cables and Transport Junctions .....	40
	SPARC: Installing PCI-SCI Cluster Interconnect Hardware .....	41
	▼ SPARC: How to Install PCI-SCI Transport Cables and Transport Junction .....	41
	SPARC: Installing Sun Fire Link Cluster Interconnect Hardware .....	46
	Configuring VLANs as Private Interconnect Networks .....	47
<b>4</b>	<b>Maintaining Cluster Interconnect Hardware</b> .....	49
	Maintaining Interconnect Hardware in a Running Cluster .....	49
	▼ How to Add an Interconnect Component .....	50
	▼ How to Replace an Interconnect Component .....	52
	▼ How to Remove an Interconnect Component .....	54
	▼ How to Upgrade Transport Adapter Firmware .....	55
<b>5</b>	<b>Installing and Maintaining Public Network Hardware</b> .....	57
	Requirements When Using Jumbo Frames .....	57
	Installing Public Network Hardware .....	58
	Where to Go From Here .....	58
	Maintaining Public Network Hardware in a Running Cluster .....	58
	Adding Public Network Adapters .....	58
	Replacing Public Network Adapters .....	59
	Removing Public Network Adapters .....	59
	Sun Gigabit Ethernet Adapter Considerations .....	59
	ce Sun Ethernet Driver Considerations .....	60
	GigaSwift Ethernet Driver and Jumbo Frames .....	61
<b>6</b>	<b>Maintaining Platform Hardware</b> .....	63
	Mirroring Internal Disks on Servers that Use Internal Hardware Disk Mirroring or Integrated Mirroring .....	63

▼ How to Configure Internal Disk Mirroring After the Cluster is Established .....	64
▼ How to Remove an Internal Disk Mirror .....	65
Configuring Cluster Nodes With a Single, Dual-Port HBA .....	66
Risks and Trade-offs When Using One Dual-Port HBA .....	67
Supported Configurations When Using a Single, Dual-Port HBA .....	68
Cluster Configuration When Using Solaris Volume Manager and a Single Dual-Port HBA .....	68
Cluster Configuration When Using Solaris Volume Manager for Sun Cluster and a Single Dual-Port HBA .....	69
<b>7 SPARC: Campus Clustering With Sun Cluster Software .....</b>	<b>71</b>
SPARC: Requirements for Designing a Campus Cluster .....	71
SPARC: Selecting Networking Technologies .....	72
SPARC: Connecting to Storage .....	72
SPARC: Sharing Data Storage .....	72
SPARC: Complying With Quorum Device Requirements .....	72
SPARC: Replicating Solaris Volume Manager Disksets .....	73
SPARC: Guidelines for Designing a Campus Cluster .....	73
SPARC: Determining the Number of Rooms in Your Cluster .....	73
SPARC: Deciding How to Use Quorum Devices .....	77
SPARC: Determining Campus Cluster Interconnect Technologies .....	80
SPARC: Cluster Interconnect Technologies .....	80
SPARC: Storage Area Network Technologies .....	80
SPARC: Installing and Configuring Interconnect, Storage, and Fibre-Channel Hardware .....	81
SPARC: Additional Campus Cluster Configuration Examples .....	82
SPARC: Three-Room Examples .....	82
SPARC: Two-Room Examples .....	85
<b>8 Verifying Sun Cluster Hardware Redundancy .....</b>	<b>89</b>
Testing Node Redundancy .....	89
▼ How to Test Nodes Using a Power-Off Method .....	90
Testing Cluster Interconnect and IP Network Multipathing Group Redundancy .....	90
▼ How to Test Cluster Interconnects .....	91
▼ How to Test IP Network Multipathing Groups .....	92
<b>9 Recabling and Replacing Disk Devices .....</b>	<b>93</b>
Replacing a Failed Boot Disk .....	93

▼ How to Replace a Failed Boot Disk in a Running Cluster .....	93
Moving a Disk Cable .....	94
▼ How to Move a Disk Cable to a New Host Adapter .....	95
▼ How to Move a Disk Cable From One Node to Another Node .....	96
▼ How to Update Sun Cluster Software to Reflect Proper Device Configuration .....	97
<b>A Data Replication Approaches .....</b>	<b>99</b>
Using Host-Based Data Replication .....	99
SPARC: Using Storage-Based Data Replication .....	101
SPARC: Recovery Concerns When Using Storage-Based Data Replication .....	103
<b>Index .....</b>	<b>105</b>

# Preface

---

The *Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS* provides a variety of information about how to install and administer basic Sun™ Cluster hardware components. Topics covered in this book include how to install and configure terminal concentrators, the cluster interconnect, public network hardware, campus clustering, and dynamic reconfiguration.

Use this manual with any version of Sun Cluster 3.0 or .31 software. Unless otherwise noted, procedures are the same for all Sun Cluster 3.0 and 3.1 versions.

Sun Cluster software runs on two platforms, SPARC and x86. The information in this document pertains to both platforms unless otherwise specified in a special chapter, section, note, bulleted item, figure, table, or example. Procedures are also the same unless otherwise noted for Sun Cluster 3.1 03/04 and later versions on both SPARC® and x86 based systems.

---

**Note** – In this document, the term “x86” refers to the Intel 32-bit family of microprocessor chips and compatible microprocessor chips made by AMD.

In this document, references to Oracle Real Application Clusters also apply to Oracle Parallel Server unless otherwise stated.

---

See the “[Revision History](#)” on [page 8](#) for a list of changes to this manual.

This book does not include information about configuring servers in a Sun Cluster environment nor does it include specific storage device procedures.

## Who Should Use This Book

This book is for Sun representatives who are performing the initial installation of a Sun Cluster configuration and for system administrators who are responsible for maintaining the system.

This document is intended for experienced system administrators with extensive knowledge of Sun software and hardware. Do not use this document as a planning or presales guide. You should have already determined your system requirements and purchased the appropriate equipment and software before reading this document.

## How This Book Is Organized

The following chapters contain information about hardware used in a Sun Cluster environment.

[Chapter 1](#) provides an overview of installing and administering Sun Cluster hardware.

[Chapter 2](#) describes how to install and configure a terminal concentrator.

[Chapter 3](#) describes how to install cluster interconnect hardware and configure VLANs.

[Chapter 4](#) describes how to maintain cluster interconnect hardware.

[Chapter 5](#) describes how to install and maintain the public network hardware.

[Chapter 7](#) provides guidelines and diagrams on how to configure a campus cluster.

[Chapter 8](#) describes how to verify cluster redundancy.

[Chapter 9](#) provides procedures on how to recable disks.

[Appendix A](#) describes host-based and storage-based data replication in standard and campus clusters.

## Revision History

The following table lists the information that has been revised or added since the initial release of this documentation. The table also lists the revision date for these changes.

**TABLE P-1** Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS

Revision Date	New Information
June 2006	Added restriction on spanning tree algorithm when using Ethernet switches in the cluster private interconnect.
May 2006	Amended information regarding internal disk RAID for servers.
January 2006	Added information on single, dual-port HBA configurations.
December 2005	Added information about using InfiniBand on the cluster interconnect.



## Related Documentation

The following books provide conceptual information or procedures to administer hardware and applications. If you plan to use this documentation in a hardcopy format, ensure that you have these books available for your reference.

The following Sun Cluster books support the Sun Cluster 3.1, 2004, and 2005 releases. If you are maintaining a different version of Sun Cluster software, refer to the appropriate documentation. All Sun Cluster documentation is available at <http://docs.sun.com>.

Documentation that is not available at <http://docs.sun.com> is listed with the appropriate URL.

**TABLE P-2** Hardware Documentation

Title	Part Number
<i>Sun Fire Link Software Installation Guide</i> <a href="http://docs.sun.com/db/doc/817-0401-11">http://docs.sun.com/db/doc/817-0401-11</a> .	817-0401
<i>Sun Fire Link Service Manual</i> <a href="http://docs.sun.com/db/doc/806-1394-10">http://docs.sun.com/db/doc/806-1394-10</a> .	806-1394
<i>Sun Fire Link Fabric Administrator's Guide</i> <a href="http://docs.sun.com/db/doc/806-1405-10">http://docs.sun.com/db/doc/806-1405-10</a> .	806-1405
<i>Sun Fire Link Hardware Installation Guide</i> Available on <a href="http://docs.sun.com/db/doc/806-1396-11">http://docs.sun.com/db/doc/806-1396-11</a> .	806-1396
<i>Sun StorEdge SAN Foundation 4.2 Installation Guide</i> Available on <a href="http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_Solutions/SAN">http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_Solutions/SAN</a>	817-1244
<i>Sun StorEdge SAN Foundation 4.2 Configuration Guide</i> Available on <a href="http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_Solutions/SAN">http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_Solutions/SAN</a>	817-1245
<i>Sun StorEdge SAN Foundation 4.2 Guide to Documentation</i> Available on <a href="http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_Solutions/SAN">http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_Solutions/SAN</a>	817-1576
<i>Sun StorEdge SAN Foundation 4.2 Release Notes</i> Available on <a href="http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_Solutions/SAN">http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_Solutions/SAN</a>	817-1246
<i>IP Network Multipathing Administration Guide</i>	816-0850

**TABLE P-2 Hardware Documentation** (Continued)

Title	Part Number
<i>Sun Gigabit Ethernet/P 2.0 Adapter Installation and User's Guide</i>	804-6140
Available on <a href="http://www.sun.com/products-n-solutions/hardware/docs">http://www.sun.com/products-n-solutions/hardware/docs</a>	
<i>Sun StorEdge Long Wave GBIC Interface Converter Guide</i>	805-6965
Available on <a href="http://www.sun.com/products-n-solutions/hardware/docs">http://www.sun.com/products-n-solutions/hardware/docs</a>	
<i>Sun StorEdge Traffic Manager Installation and Configuration Guide</i>	816-1420
Available on <a href="http://www.sun.com/products-n-solutions/hardware/docs">http://www.sun.com/products-n-solutions/hardware/docs</a>	

**TABLE P-3 Sun Cluster Documentation**

Application	Title
Concepts	<i>Sun Cluster Concepts Guide for Solaris OS</i>
Overview	<i>Sun Cluster Overview for Solaris OS</i>
Hardware administration	<i>Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS</i> Individual hardware administration guides
Software installation	<i>Sun Cluster Software Installation Guide for Solaris OS</i>
Data service administration	<i>Sun Cluster Data Services Planning and Administration Guide for Solaris OS</i> Individual data service guides
Data service development	<i>Sun Cluster Data Services Developer's Guide for Solaris OS</i>
System administration	<i>Sun Cluster System Administration Guide for Solaris OS</i>
Error messages	<i>Sun Cluster Error Messages Guide for Solaris OS</i>
Command and function reference	<i>Sun Cluster Reference Manual for Solaris OS</i>
Release Notes	<i>Sun Cluster 3.1 8/05 Release Notes for Solaris OS</i> <i>Sun Cluster 3.0-3.1 Release Notes Supplement</i>

## Using UNIX Commands

This document contains information about commands that are used to install, configure, or upgrade a Sun Cluster configuration. This document might not contain complete information about basic UNIX® commands and procedures such as shutting down the system, booting the system, and configuring devices.

See one or more of the following sources for this information:

- Online documentation for the Solaris™ Operating System (Solaris OS)
- Other software documentation that you received with your system
- Solaris Operating System man pages

## Getting Help

If you have problems installing or using Sun Cluster, contact your service provider and provide the following information.

- Your name and email address (if available)
- Your company name, address, and phone number
- The model number and serial number of your systems
- The release number of the operating environment (for example, Solaris 9)
- The release number of Sun Cluster (for example, Sun Cluster 3.1 8/05)

Use the following commands to gather information about your system for your service provider.

Command	Function
<code>prtconf -v</code>	Displays the size of the system memory and reports information about peripheral devices
<code>psrinfo -v</code>	Displays information about processors
<code>showrev -p</code>	Reports which patches are installed
<code>prtdiag -v</code>	Displays system diagnostic information
<code>/usr/cluster/bin/scinstall -pv</code>	Displays Sun Cluster release and package version information

Also have available the contents of the `/var/adm/messages` file.

## Documentation, Support, and Training

The Sun web site provides information about the following additional resources:

- Documentation (<http://www.sun.com/documentation/>)
- Support (<http://www.sun.com/support/>)
- Training (<http://www.sun.com/training/>)

## Typographic Conventions

The following table describes the typographic conventions that are used in this book.

TABLE P-4 Typographic Conventions

Typeface	Meaning	Example
AaBbCc123	The names of commands, files, and directories, and onscreen computer output	Edit your <code>.login</code> file. Use <code>ls -a</code> to list all files. <code>machine_name%</code> you have mail.
<b>AaBbCc123</b>	What you type, contrasted with onscreen computer output	<code>machine_name%</code> <b>su</b> Password:
<i>aabcc123</i>	Placeholder: replace with a real name or value	The command to remove a file is <i>rm filename</i> .
<i>AaBbCc123</i>	Book titles, new terms, and terms to be emphasized	Read Chapter 6 in the <i>User's Guide</i> . <i>A cache</i> is a copy that is stored locally. Do <i>not</i> save the file. <b>Note:</b> Some emphasized items appear bold online.

## Shell Prompts in Command Examples

The following table shows the default UNIX system prompt and superuser prompt for the C shell, Bourne shell, and Korn shell.

TABLE P-5 Shell Prompts

Shell	Prompt
C shell	<code>machine_name%</code>
C shell for superuser	<code>machine_name#</code>
Bourne shell and Korn shell	<code>\$</code>
Bourne shell and Korn shell for superuser	<code>#</code>

## Product Training

Sun Microsystems offers training in many Sun technologies through a variety of instructor-led courses and self-paced courses. For information about the training courses that Sun offers and to enroll in a class, visit Sun Training at <http://training.sun.com/>.



# Introduction to Sun Cluster Hardware

---

This chapter provides overview information on cluster hardware. The chapter also provides overviews of the tasks that are involved in installing and maintaining this hardware specifically in a Sun™ Cluster environment.

This chapter contains the following information:

- “Installing Sun Cluster Hardware” on page 15
- “Maintaining Sun Cluster Hardware” on page 17
- “Powering On and Off Sun Cluster Hardware” on page 18
- “Dynamic Reconfiguration Operations For Sun Cluster Nodes” on page 18
- “Local and Multihost Disks in a Sun Cluster Environment” on page 19
- “Removable Media in a Sun Cluster Environment” on page 19
- “SAN Solutions in a Sun Cluster Environment” on page 20
- “Hardware Restrictions” on page 20

## Installing Sun Cluster Hardware

The following procedure lists the tasks for installing a cluster and the sources for instructions.

**TABLE 1-1** Task Map: Installing Cluster Hardware

Task	For Instructions
Plan for cluster hardware capacity, space, and power requirements.	The site planning documentation that shipped with your nodes and other hardware
Install the nodes.	The documentation that shipped with your nodes
Install the administrative console.	The documentation that shipped with your administrative console

**TABLE 1-1 Task Map: Installing Cluster Hardware** (Continued)

Task	For Instructions
Install a console access device.  Use the procedure that is indicated for your type of console access device. For example, Sun Enterprise E10000 servers use a System Service Processor (SSP) as a console access device, rather than a terminal concentrator.	“Installing the Terminal Concentrator” on page 21 or The documentation that shipped with your Sun Enterprise E10000 hardware
Install the cluster interconnect hardware.	Chapter 3
Install the public network hardware.	Chapter 5
Install and configure the shared disk storage arrays.	Refer to the Sun Cluster manual that pertains to your storage device as well as to the device’s own documentation.
Install the Solaris Operating System and Sun Cluster software.	Sun Cluster software installation documentation
Configure the cluster interconnects.	Sun Cluster software installation documentation

## ▼ Installing Sun Cluster Hardware

### 1 Plan for cluster hardware capacity, space, and power requirements.

For more information, see the site planning documentation that shipped with your servers and other hardware. See “[Hardware Restrictions](#)” on page 20 for critical information about hardware restrictions with Sun Cluster.

### 2 Install the nodes.

For server installation instructions, see the documentation that shipped with your servers.

### 3 Install the administrative console.

For more information, see the documentation that shipped with your administrative console.

### 4 Install a console access device.

Use the procedure that is indicated for your type of console access device. For example, Sun Enterprise E10000 servers use a System Service Processor (SSP) as a console access device, rather than a terminal concentrator.

For installation instructions, see “[Installing the Terminal Concentrator](#)” on page 21 or the documentation that shipped with your Sun Enterprise E10000 hardware.

### 5 Install the cluster interconnect and public network hardware.

For installation instructions, see [Chapter 3](#).



**6 Install and configure the storage arrays.**

Perform the service procedures that are indicated for your type of storage hardware.

**7 Install the Solaris Operating System and Sun Cluster software.**

For more information, see Sun Cluster software installation documentation .

**8 Plan, install, and configure resource groups and data services.**

For more information, see the Sun Cluster data services collection.

## Maintaining Sun Cluster Hardware

*Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS* augments documentation that ships with your hardware components by providing information on maintaining the hardware *specifically in a Sun Cluster environment*. Table 1–2 describes some of the differences between maintaining cluster hardware and maintaining standalone hardware.

TABLE 1–2 Sample Differences Between Servicing Standalone and Cluster Hardware

Task	Standalone Hardware	Cluster Hardware
Shutting down a node	Use the <code>shutdown(1M)</code> command.	To perform an orderly node shutdown, first use the <code>scswitch(1M)</code> command to switch device groups and resource groups to another node. Then shut down the node by running the <code>shutdown(1M)</code> command.
Adding a disk	Run <code>boot - r</code> on a SPARC based system, <code>b - r</code> on an x86 based system, or <code>devfsadm(1M)</code> to assign a logical device name to the disk. You also need to run volume manager commands to configure the new disk if the disks are under volume management control.	Use the <code>devfsadm(1M)</code> , <code>scgdevs(1M)</code> , and <code>scdidadm(1M)</code> commands. You also need to run volume manager commands to configure the new disk if the disks are under volume management control.
Adding a transport adapter or public network adapter	Perform an orderly node shutdown, then install the public network adapter. After you install the network adapter, update the <code>/etc/hostname.adapter</code> and <code>/etc/inet/hosts</code> files.	Perform an orderly node shutdown, then install the public network adapter. After you install the public network adapter, update the <code>/etc/hostname.adapter</code> and <code>/etc/inet/hosts</code> files. Finally, add this public network adapter to an IP Network Multipathing group.

## Powering On and Off Sun Cluster Hardware

Consider the following when powering on and powering off cluster hardware.

- Use shut down and boot procedures in your Sun Cluster system administration documentation for nodes in a running cluster.
- Use the power-on and power-off procedures in the manuals that shipped with the hardware *only* for systems that are newly installed or are in the process of being installed.



---

**Caution** – After the cluster is online and a user application is accessing data on the cluster, do not use the power-on and power-off procedures listed in the manuals that came with the hardware.

---

## Dynamic Reconfiguration Operations For Sun Cluster Nodes

The Sun Cluster environment supports Solaris dynamic reconfiguration (DR) operations on qualified servers. Throughout the *Sun Cluster Hardware Administration Collection for Solaris OS*, there are procedures that require that the user add or remove transport adapters or public network adapters in a cluster node. Contact your service provider for a list of storage arrays that are qualified for use with DR-enabled servers.

---

**Note** – Review the documentation for the Solaris DR feature on your hardware platform *before* you use the DR feature with Sun Cluster software. All of the requirements, procedures, and restrictions that are documented for the Solaris DR feature also apply to Sun Cluster DR support (except for the operating environment quiescence operation).

---

Documentation for DR on currently qualified server platforms are listed here.

- *Sun Enterprise 10000 Dynamic Reconfiguration User Guide*
- *Sun Enterprise 10000 Dynamic Reconfiguration Reference Manual*
- *Sun Fire 6800, 4810, 4800, and 3800 Systems Dynamic Reconfiguration User Guide*
- *Sun Fire 6800, 4810, 4800, and 3800 Systems Dynamic Reconfiguration Release Notes*
- *Sun Fire 15K Dynamic Reconfiguration (DR) User Guide*
- *Sun Fire 15K Dynamic Reconfiguration Release Notes*
- *Sun Fire 880 Dynamic Reconfiguration User's Guide*
- *Sun Fire V1280/Netra 1280 System Administration Guide*

### ▼ DR Operations in a Cluster With DR-Enabled Servers

Some procedures within the *Sun Cluster Hardware Administration Collection for Solaris OS* instruct the user to shut down and power off a cluster node before you add, remove, or replace a transport adapter or a public network adapter (PNA).

However, if the node is a server that is enabled with the DR feature, the user does *not* have to power off the node before you add, remove, or replace the transport adapter or PNA. Instead, do the following:

- 1 **Follow the procedure steps in *Sun Cluster Hardware Administration Collection for Solaris OS*, including any steps for disabling and removing the transport adapter or PNA from the active cluster interconnect.**  
See the Sun Cluster system administration documentation for instructions about how to remove transport adapters or PNAs from the cluster configuration.
- 2 **Skip any step that instructs you to power off the node, where the purpose of the power-off is to add, remove, or replace a transport adapter or PNA.**
- 3 **Perform the DR operation (add, remove, or replace) on the transport adapter or PNA.**
- 4 **Continue with the next step of the procedure in *Sun Cluster Hardware Administration Collection for Solaris OS*.**

For conceptual information about Sun Cluster support of the DR feature, see your Sun Cluster concepts documentation document.

## Local and Multihost Disks in a Sun Cluster Environment

Two sets of storage arrays reside within a cluster: local disks and multihost disks.

- Local disks are directly connected to a single node and hold the Solaris Operating System for each node.
- Multihost disks are connected to more than one node and hold client application data and other files that need to be accessed from multiple nodes.

For more conceptual information on multihost disks and local disks, see the Sun Cluster concepts documentation.

## Removable Media in a Sun Cluster Environment

Removable media include tape and CD-ROM drives, which are local devices. *Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS* does not contain procedures for adding, removing, or replacing removable media as highly available storage arrays. Although tape and CD-ROM drives are global devices, these drives do not have more than one port and do not have multi-initiator firmware support that would enable these devices as highly available. Thus, *Sun Cluster 3.0-3.1 Hardware Administration Manual for Solaris OS* focuses on disk drives as global devices.

Although tape and CD-ROM drives cannot be highly available at this time, in a cluster environment, you can access tape and CD-ROM drives that are not local to your system. All the various density extensions (such as h, b, l, n, and u) are mapped so that the tape drive can be accessed from any node in the cluster.

Install, remove, replace, and use tape and CD-ROM drives as you would in a noncluster environment. For procedures about how to install, remove, and replace tape and CD-ROM drives, see the documentation that shipped with your hardware.

## SAN Solutions in a Sun Cluster Environment

You cannot have a single point of failure in a SAN configuration that is in a Sun Cluster environment. For information on how to install and configure a SAN configuration, see your [SAN documentation](#).

## Hardware Restrictions

The following restrictions apply to hardware in all Sun Cluster configurations.

- Multihost tape, CD-ROM, and DVD-ROM are not supported.
- Alternate pathing (AP) is not supported.
- Storage devices with more than a single path from a given cluster node to the enclosure are not supported except for the following storage devices:
  - Sun StorEdge™ A3500, for which two paths are supported to each of two nodes.
  - Devices using Sun StorEdge Traffic Manager.
  - EMC storage devices that use EMC PowerPath software.
  - Sun StorEdge 9900 storage devices that use HDLM.
- If you are using a Sun Enterprise™ 420R server with a PCI card in slot J4701, the motherboard must be at dash-level 15 or higher (501–5168–15 or higher). To find the motherboard part number and revision level, look at the edge of the board closest to PCI slot 1.
- System panics have been observed in clusters when UDWIS I/O cards are used in slot 0 of a board in a Sun Enterprise 10000 server; do not install UDWIS I/O cards in slot 0 of this server.
- Sun VTS™ software is not supported.

# Installing and Configuring the Terminal Concentrator

---

This chapter provides the hardware and software procedures for installing and configuring a terminal concentrator as a console access device in a Sun Cluster environment. This chapter also includes information on how to use a terminal concentrator.

This chapter contains the following procedures:

- “How to Install the Terminal Concentrator in a Cabinet” on page 22
- “How to Connect the Terminal Concentrator” on page 25
- “How to Configure the Terminal Concentrator” on page 26
- “How to Set Terminal Concentrator Port Parameters” on page 28
- “How to Correct a Port Configuration Access Error” on page 29
- “How to Establish a Default Route for the Terminal Concentrator” on page 31
- “How to Connect to a Node’s Console Through the Terminal Concentrator” on page 33
- “How to Reset a Terminal Concentrator Port” on page 34

For conceptual information on console access devices, see your Sun Cluster concepts documentation.

## Installing the Terminal Concentrator

This section describes the procedure for installing the terminal concentrator hardware and for connecting cables from the terminal concentrator to the administrative console and to the cluster nodes.

## ▼ How to Install the Terminal Concentrator in a Cabinet

This procedure provides step-by-step instructions for rack-mounting the terminal concentrator in a cabinet. For convenience, you can rack-mount the terminal concentrator even if your cluster does not contain rack-mounted nodes.

- To rack-mount your terminal concentrator, go to the first step of the following procedure.
- If you do not want to rack-mount your terminal concentrator, place the terminal concentrator in its standalone location, connect the unit power cord into a utility outlet, and go to [“How to Connect the Terminal Concentrator”](#) on page 25.

### 1 Install the terminal concentrator bracket hinge onto the primary cabinet:

- a. Locate the bracket hinge portion of the terminal concentrator bracket assembly (see [Figure 2-1](#)).
- b. Loosely install two locator screws in the right-side rail of the rear of the cabinet.  
Thread the screws into holes 8 and 29, as shown in [Figure 2-1](#). The locator screws accept the slotted holes in the hinge piece.
- c. Place the slotted holes of the hinge over the locator screws, and let the hinge drop into place.
- d. Install the screws into holes 7 and 28.  
Tighten these screws, and the screws in holes 8 and 29, as shown in [Figure 2-1](#).

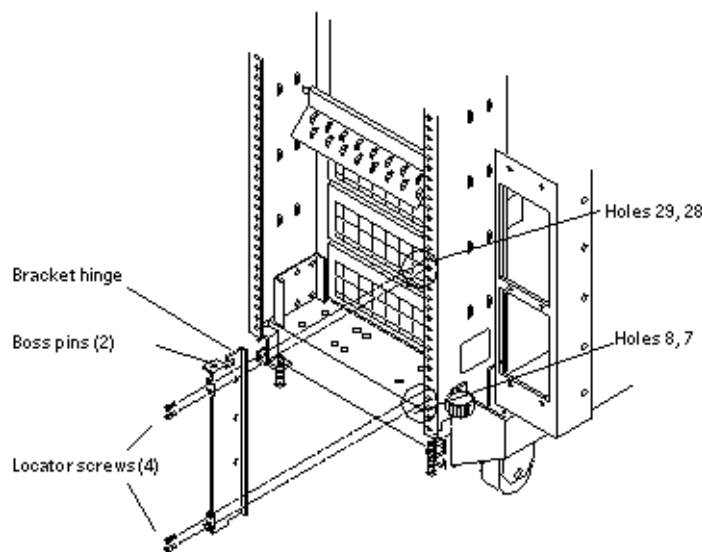


FIGURE 2-1 Installing the Terminal Concentrator Bracket Hinge to the Cabinet

### 2 Install the terminal concentrator into the bracket.

- a. Place the side pieces of the bracket against the terminal concentrator, as shown in [Figure 2-2](#).
- b. Lower the terminal concentrator (with side pieces) onto the bottom plate, aligning the holes in the side pieces with the threaded studs on the bottom plate.
- c. Install and tighten three nuts on the three threaded studs that penetrate through each side plate.

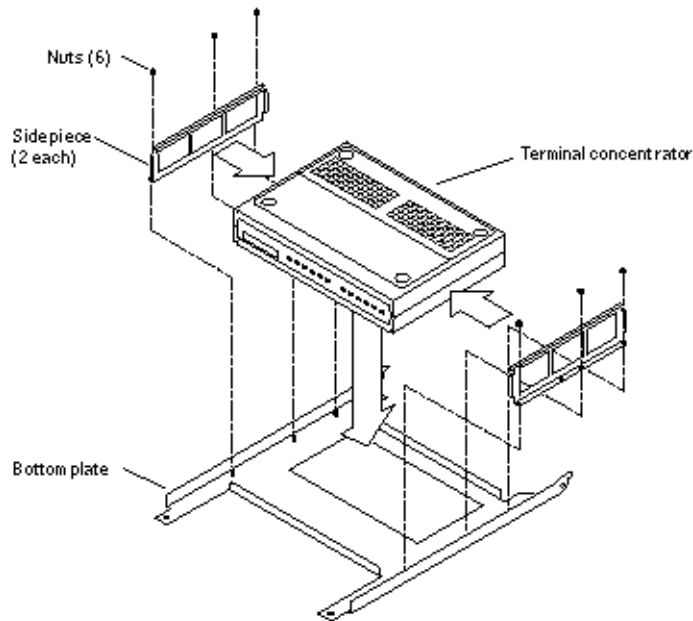


FIGURE 2-2 Installing the Terminal Concentrator Into the Bracket

- 3 Install the terminal concentrator bracket onto the bracket hinge that is already installed on the cabinet.
  - a. Turn the terminal concentrator bracket on its side so the hinge holes and cable connectors face toward the bracket hinge (see [Figure 2-3](#)).
  - b. Align the bracket holes with the boss pins on the bracket hinge and install the bracket onto the hinge.
  - c. Install the keeper screw in the shorter boss pin to ensure the assembly cannot be accidentally knocked off the hinge.

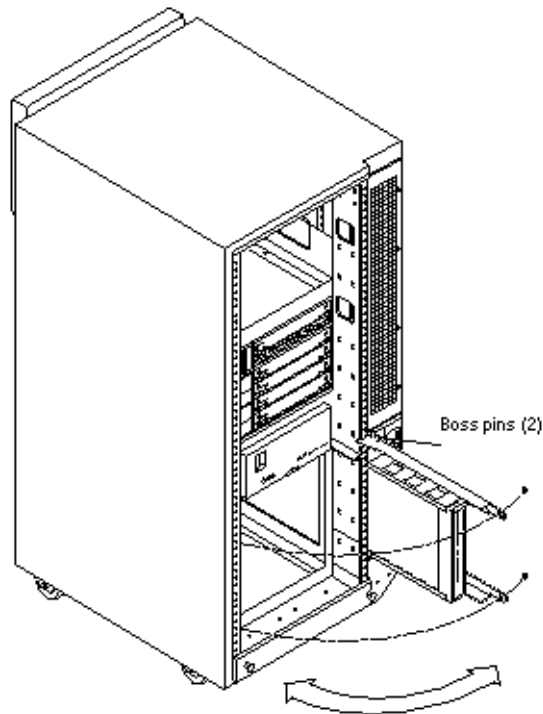


FIGURE 2-3 Terminal Concentrator Bracket Installed on the Hinge

- 4 Connect one end of the power cord to the terminal concentrator, as shown in [Figure 2-4](#). Connect the other end of the power cord to the power distribution unit.

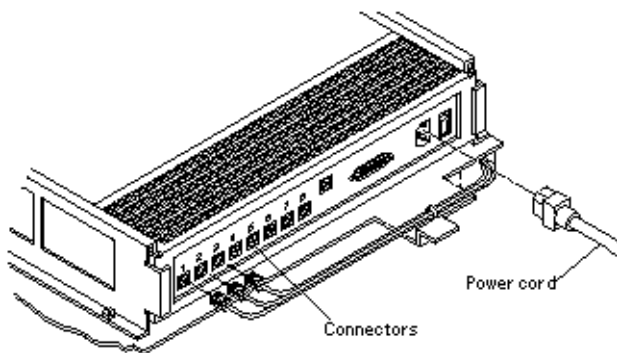


FIGURE 2-4 Terminal Concentrator Cable Connector Locations

**See Also** To cable the terminal concentrator, go to [“How to Connect the Terminal Concentrator”](#) on page 25.



## ▼ How to Connect the Terminal Concentrator

- 1 **Connect a DB-25 to RJ-45 serial cable (part number 530-2152-01 or 530-2151-01) from serial port A on the administrative console to serial port 1 on the terminal concentrator, as shown in [Figure 2-5](#).**

This cable connection from the administrative console enables you to configure the terminal concentrator. You can remove this connection after you set up the terminal concentrator.

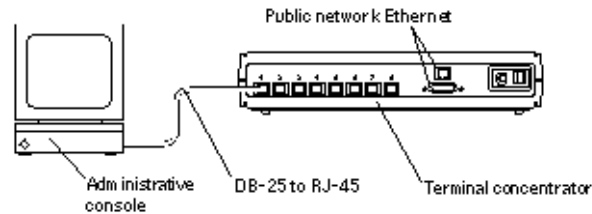


FIGURE 2-5 Connecting the Administrative Console

- 2 **Connect the cluster nodes to the terminal concentrator by using serial cables.**

The cable connections from the concentrator to the nodes enable you to access the ok prompt or OpenBoot™ PROM (OBP) mode by using the Cluster Console windows from the Cluster Control Panel (CCP). For more information on how to use the CCP, see your Sun Cluster system administration documentation.

- 3 **Connect the public network Ethernet cable to the appropriate connector on the terminal concentrator.**

---

**Note** – The terminal concentrator requires a 10-Mbit/sec Ethernet connection.

---

- 4 **Close the terminal concentrator bracket, and install screws in holes 8 and 29 on the left-side rear rail of the cabinet (see [Figure 2-3](#)).**

**Next Steps** Go to “[Configuring the Terminal Concentrator](#)” on page 25.

## Configuring the Terminal Concentrator

This section describes the procedure for configuring the terminal concentrator’s network addresses and ports.

## ▼ How to Configure the Terminal Concentrator

- 1 From the administrative console, add the following entry to the `/etc/remote` file.

```
tc:\
:dv=/dev/term/a:br#9600:
```

- 2 Verify that the server and the terminal concentrator are powered on and that the cabinet keyswitch (if applicable) is in the ON position.

- 3 Establish a connection to the terminal concentrator's serial port:

```
# tip tc
```

- 4 Hold down the terminal concentrator Test button (Figure 2-6) until the power LED flashes (about three seconds), then release the Test button.

- 5 Hold down the terminal concentrator Test button again for one second, then release it.

The terminal concentrator performs a self-test, which lasts about 30 seconds. Messages display on the administrative console. If the network connection is not found, press the Q key to stop the message.

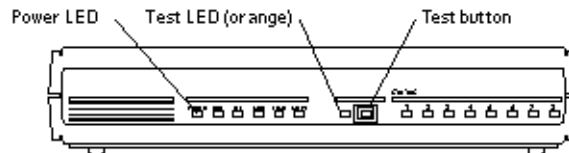


FIGURE 2-6 Terminal Concentrator Test Button and LEDs

- 6 Observe the terminal concentrator front-panel LEDs and use the information in the following table to decide your course of action.

Power (Green)	Unit (Green)	Net (Green)	Attn (Amber)	Load (Green)	Active (Green)	Test (Orange)
ON	ON	ON	ON	OFF	Intermittent blinking	ON

- If the front-panel LEDs light up as shown in the table above and the administrative console displays a `monitor:: prompt`, go to [Step 7](#).
- If the front-panel LEDs do not light up as shown in the table above, or the administrative console does not display a `monitor:: prompt`, use the following table and the documentation that shipped with your terminal concentrator to troubleshoot the problem.

Mode	Power (Green)	Unit (Green)	Net (Green)	Attn (Amber)	Load (Green)	Active (Green)
Hardware failure	ON	Blinking	OFF	Blinking	OFF	OFF
Network test failure	ON	ON	Blinking	OFF	OFF	Intermittent blinking
Network test aborted, or net command failed	ON	ON	OFF	Blinking	OFF	Intermittent blinking
Booted wrong image	ON	ON	ON	Blinking	OFF	OFF
Other failure	One or more Status LEDs (1-8) are ON					

**7 Use the `addr` command to assign an IP address, subnet mask, and network address to the terminal concentrator.**

In the following example (Class B network, Class C subnet), the broadcast address is the terminal concentrator's address with the host portion set to 255 (all binary 1's).

```
monitor:: addr
Enter Internet address [<uninitialized>]: 172.25.80.6
Internet address: 172.25.80.6
Enter Subnet mask [255.255.0.0]: 255.255.255.0
Subnet mask: 255.255.255.0
Enter Preferred load host Internet address [<any host>]: 172.25.80.6
*** Warning: Load host and Internet address are the same ***
Preferred load host address: 172.25.80.6
Enter Broadcast address [0.0.0.0]: 172.25.80.255
Broadcast address: 172.25.80.255
Enter Preferred dump address [0.0.0.0]: 172.25.80.6
Preferred dump address: 172.25.80.6
Select type of IP packet encapsulation (ieee802/ethernet) [<ethernet>]:
Type of IP packet encapsulation: <ethernet>
Load Broadcast Y/N [Y]: n
Load Broadcast: N
```

**8 After you finish the `addr` session, power-cycle the terminal concentrator.**

The Load and Active LEDs should briefly blink, then the Load LED should turn off.

**9 Use the `ping(1M)` command to confirm that the network connection works.**

- 10 Exit the `tip` utility by pressing Return and typing a tilde, followed by a period.

```
<Return>~.
~
[EOT]
#
```

**Next Steps** Go to “[How to Set Terminal Concentrator Port Parameters](#)” on page 28.

## ▼ How to Set Terminal Concentrator Port Parameters

This procedure explains how to determine if the port type variable must be set and how to set this variable.

The port type parameter must be set to `dial_in`. If the parameter is set to `hardwired`, the cluster console might be unable to detect when a port is already in use.

- 1 Locate the Sun serial number label on the top panel of the terminal concentrator (Figure 2–7).
- 2 Check if the serial number is in the lower serial-number range. The serial number consists of 7 digits, followed by a dash and 10 more digits.
  - If the numbers after the dash start with 9520 or higher, the port type variable is set correctly. Go to Step 4.
  - If the numbers after the dash start with 9519 or lower, you must change the port type variable. Go to Step 3.

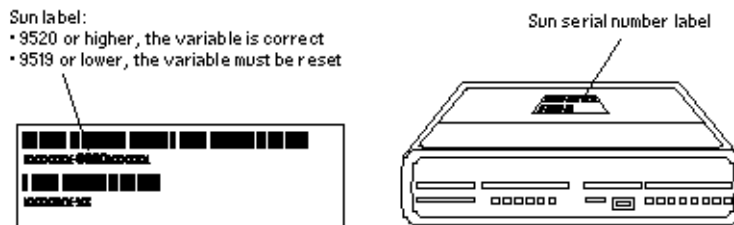


FIGURE 2-7 Determining the Version From the Serial Number Label

- 3 Use the administrative console to change the port type variable to `dial_in` by setting the port parameters, then reboot the terminal concentrator as shown in the following example.

The boot command causes the changes to take effect. The terminal concentrator is unavailable for approximately one minute.

```
admin-ws# telnet tc_name
Trying terminal concentrator IP address
Connected to tc_name
```

```

Escape character is "^]".
Rotaries Defined:
    cli
Enter Annex port name or number: cli
Annex Command Line Interpreter * Copyright 1991 Xylogics, Inc.
annex: su
Password: password
(The default password is the terminal concentrator IP address)
annex# admin
Annex administration MICRO-XL-UX R7.0.1, 8 ports
admin : set port=1-8 type dial_in imask_7bits Y
    You may need to reset the appropriate port, Annex subsystem or
        reboot the Annex for changes to take effect.
admin : set port=1-8 mode slave
admin : quit
annex# boot
bootfile: <return>
warning: <return>

```

---

**Note** – Ensure that the terminal concentrator is powered on and has completed the boot process before you proceed.

---

#### 4 Verify that you can log in from the administrative console to the consoles of each node.

For information on how to connect to the nodes' consoles, see [“How to Connect to a Node's Console Through the Terminal Concentrator”](#) on page 33.

## ▼ How to Correct a Port Configuration Access Error

A misconfigured port that does not accept network connections might return a `Connect : Connection refused` message when you use `telnet(1)`. Use the following procedure to correct the port configuration.

### 1 Connect to the terminal concentrator without specifying a port.

```
# telnet tc_name
```

*tc\_name* Specifies the hostname of the terminal concentrator

### 2 Press Return again after you make the connection, then specify the port number.

```

Trying ip_address ..
Connected to 192.9.200.1
Escape character is "^]".
...
[RETURN]
Rotaries Defined:

```

```
cli
Enter Annex port name or number: 2
```

- If you see a Port(s) busy, do you wish to wait? (y/n) message, answer **N** and go to “[How to Reset a Terminal Concentrator Port](#)” on page 34.
- If you see an Error: Permission denied message, the port mode is configured incorrectly to the command-line interface and must be set to slave. Go to [Step 3](#).

### 3 Select the terminal concentrator’s command-line interface.

```
...
Enter Annex port name or number: cli
annex:
```

### 4 Type the su command and password.

The default password is the terminal concentrator’s IP address.

```
annex: su
Password:
```

### 5 Reset the port.

```
annex# admin
Annex administration MICRO-XL-UX R7.0.1, 8 ports
admin: port 2
admin: set port mode slave
```

You may need to reset the appropriate port, Annex subsystem or reboot the Annex for changes to take effect.

```
admin: reset 2
```

## Example 2-1 Correcting a Terminal Concentrator Port Configuration Access Error

The following example shows how to correct an access error on the terminal concentrator port 4.

```
admin-ws# telnet tc1
Trying 192.9.200.1 ...
Connected to 192.9.200.1.
Escape character is '^]'.
[Return]
Enter Annex port name or number: cli
```

```
...
annex: su
Password: root_password
annex# admin
Annex administration MICRO-XL-UX R7.0.1, 8 ports
admin: port 4
admin: set port mode slave
```

You may need to reset the appropriate port, Annex subsystem or

reboot the Annex for changes to take effect.  
 admin: **reset 4**

## ▼ How to Establish a Default Route for the Terminal Concentrator

---

**Note** – This procedure is optional. By setting a default route, you prevent possible problems with routing table overflows (see the following paragraphs). Routing table overflow is not a problem for connections that are made from a host that resides on the same network as the terminal concentrator.

---

A routing table overflow in the terminal concentrator can cause network connections to be intermittent or lost altogether. Symptoms include connection timeouts and routes that are reestablished, then disappear, even though the terminal concentrator itself has not rebooted.

The following procedure fixes this problem by establishing a default route within the terminal concentrator. To preserve the default route within the terminal concentrator, you must also disable the routed feature.

### 1 Connect to the terminal concentrator.

```
# telnet tc_name
```

*tc\_name* Specifies the name of the terminal concentrator

### 2 Press Return again after you make the connection, then select the command-line interface to connect to the terminal concentrator.

...

Enter Annex port name or number: **cliannex:**

### 3 Type the su command and password.

The default password is the terminal concentrator's IP address.

```
annex: su
```

```
Password:
```

### 4 Start the editor to change the config.annex file.

```
annex# edit config.annex
```

---

**Note** – The keyboard commands for this editor are Control-W: save and exit, Control-X: exit, Control-F: page down, and Control-B: page up.

---

The `config.annex` file, which is created in the terminal concentrator's EEPROM file system, defines the default route. The `config.annex` file can also define rotaries that enable a symbolic name to be used instead of a port number.

**5 Add the following lines to the file.**

Substitute the appropriate IP address for your default router.

```
%gateway
net default gateway 192.9.200.2 metric 1 active ^W
```

**6 Disable the local routed feature.**

```
annex# admin set annex routed n
```

**7 Reboot the terminal concentrator.**

```
annex# boot
bootfile: <reboot>
warning: <return>
```

While the terminal concentrator is rebooting, you cannot access the node consoles.

## Example 2-2 Establishing a Default Route for the Terminal Concentrator

The following example shows how to establish a default route for the terminal concentrator.

```
admin-ws# telnet tc1
Trying 192.9.200.1 ...
Connected to 192.9.200.1.
Escape character is '^]'.
[Return]
Enter Annex port name or number: cli
...
annex: su
Password: root_password
annex: edit config.annex
(Editor starts)
Ctrl-W:save and exit Ctrl-X:exit Ctrl-F:page down Ctrl-B:page up
%gateway
net default gateway 192.9.200.2 metric 1 active ^W
annex# admin set annex routed n
You may need to reset the appropriate port, Annex subsystem or
reboot the Annex for changes to take effect.
annex# boot
```



# Using the Terminal Concentrator

This section describes the procedures about how to use the terminal concentrator in a cluster.

TABLE 2-1 Task Map: Using the Terminal Concentrator

Task	For Instructions
Connect to a node's console through the terminal concentrator	<a href="#">“How to Connect to a Node's Console Through the Terminal Concentrator” on page 33</a>
Reset a terminal concentrator port	<a href="#">“How to Reset a Terminal Concentrator Port” on page 34</a>

## ▼ How to Connect to a Node's Console Through the Terminal Concentrator

The following procedure enables remote connections from the administrative console to a cluster node's console by first connecting to the terminal concentrator.

- 1 Connect to a node by starting a session with the terminal concentrator port that the node is cabled to.**

```
# telnet tc_name tc_port_number
```

*tc\_name* Specifies the name of the terminal concentrator.

*tc\_port\_number* Specifies the port number on the terminal concentrator. Port numbers are configuration dependent. Typically, ports 2 and 3 (5002 and 5003) are used for the first cluster that is installed at a site.

---

**Note** – If you set up node security, you are prompted for the port password.

---

- 2 Log in to the node's console.**

After establishing the telnet connection, the system prompts you for the login name and password.

- 3 Set the terminal type, based on the type of window that was used in Step 1.**

```
# TERM=xterm
# export TERM
```

### Example 2-3 Connecting to a Node's Console Through the Terminal Concentrator

The following example shows how to connect to a cluster node in a configuration that uses a terminal concentrator. A Shell tool has already been started by using an xterm window.

```
admin-ws# telnet tc1 5002
Trying 192.9.200.1 ...
Connected to 192.9.200.1.
Escape character is '^]'.
[Return]
pys-palindrome-1 console login: root
password: root_password
(for sh or ksh)
phys-palindrome-1# TERM=xterm; export TERM
(for csh)
phys-palindrome-1# set term=xterm
```

## ▼ How to Reset a Terminal Concentrator Port

When a port on the terminal concentrator is busy, you can reset the port to disconnect its user. This procedure is useful if you need to perform an administrative task on the busy port.

A busy port returns the following message when you try to connect to the terminal concentrator.

```
telnet: Unable to connect to remote host: Connection refused
```

If you use the port selector, you might see a port busy message. See [“How to Correct a Port Configuration Access Error” on page 29](#) for details on the port busy message.

### 1 Connect to the terminal concentrator port.

```
# telnet tc_name
```

*tc\_name* Specifies the name of the terminal concentrator

### 2 Press Return again after you make the connection and select the command-line interface to connect to the terminal concentrator.

```
Enter Annex port name or number: cli
annex:
```

### 3 Type the su command and password.

The default password is the terminal concentrator's IP address.

```
annex: su
Password:
```

### 4 Determine which port to reset.

The who command shows ports that are in use.

```
annex# who
```

**5 Reset the port that is in use.**

```
annex# admin reset port_number
```

**6 Disconnect from the terminal concentrator.**

```
annex# hangup
```

You can now connect to the port.

**Example 2-4** Resetting a Terminal Concentrator Connection

The following example shows how to reset the terminal concentrator connection on port 2.

```
admin-ws# telnet tc1
Trying 192.9.200.1 ...
Connected to 192.9.200.1.
Escape character is '^]'.
[Return]
...
Enter Annex port name or number: cli
...
annex: su
Password: root_password
annex: who
Port   What   User   Location      When   Idle   Address
2      PSVR   ---   ---           ---   1:27   192.9.75.12
v1     CLI    ---   ---           ---           192.9.76.10
annex# admin reset 2
annex# hangup
```



# Installing Cluster Interconnect Hardware and Configuring VLANs

---

This chapter describes the procedures to install cluster interconnect hardware. Where appropriate, this chapter includes separate procedures for the interconnects that Sun Cluster software supports:

- Ethernet
- InfiniBand
- SPARC: PCI-SCI
- Sun Fire Link

This chapter contains the following information:

- “Installing Ethernet or InfiniBand Cluster Interconnect Hardware” on page 39
- “SPARC: Installing PCI-SCI Cluster Interconnect Hardware” on page 41
- “SPARC: Installing Sun Fire Link Cluster Interconnect Hardware” on page 46
- “Configuring VLANs as Private Interconnect Networks” on page 47

Use the following information to learn more about cluster interconnects:

- For conceptual information about cluster interconnects, see “Cluster Interconnect” in *Sun Cluster Concepts Guide for Solaris OS*.
- For information about how to administer cluster interconnects, see Chapter 6, “Administering Cluster Interconnects and Public Networks,” in *Sun Cluster System Administration Guide for Solaris OS*.

## Interconnect Requirements and Restrictions

This section contains requirements on interconnect operation when using certain special features.

### Cluster Interconnect Speed Requirements

An interconnect path is one network step in the cluster private network: from a node to a node, from a node to a switch, or from the switch to another node. Each path in your cluster interconnect must use the same networking technology, whether Ethernet, PCI-SCI, or Sun Fire Link.

All interconnect paths must also operate at the same speed. This means, for example, that if you are using Ethernet components that are capable of operating at different speeds, and if your cluster configuration does not allow these components to automatically negotiate a common network speed, you must configure them to operate at the same speed.

## Ethernet Switch Configuration When in the Cluster Interconnect

When configuring Ethernet switches for your cluster private interconnect, disable the spanning tree algorithm on ports used for the interconnect.

## SPARC: Requirements When Using Jumbo Frames

If you are using Ethernet to implement the private interconnect and your cluster is configured with scalable services, then you must configure jumbo frames for both the public and the private networks. The private interconnect must be configured with the same or greater maximum transfer unit (MTU) as the public network.

If you are using SCI or Sun Fire Link to implement the private interconnect, you can use jumbo frames on the public network with no restrictions. For information about how to configure jumbo frames, see [the Sun GigaSwift documentation](#).

All of your cluster interconnect paths must be configured with jumbo frames on when using jumbo frames.

Certain patches are required to use jumbo frames with Sun Cluster. Use the PatchPro tool (<http://www.sun.com/PatchPro/>) to get these patches.

## Restriction on SCI Card Placement

Do not place an SCI card in the 33 MHz PCI slot (slot 1) of the hot swap PCI+ (hsPCI+) I/O assembly. This placement can cause a system panic.

## Requirements and Restrictions When Using InfiniBand in the Cluster Interconnect

The following requirements and restrictions apply to Sun Cluster configurations that use InfiniBand adapters:

- A two-node cluster must use InfiniBand switches. You cannot directly connect the InfiniBand adapters to each other.
- Sun InfiniBand switches support up to nine nodes in a cluster.
- Jumbo frames are not supported on a cluster that uses InfiniBand adapters.
- If only one InfiniBand adapter is installed on a cluster node, each of its two ports must be connected to a different InfiniBand switch.
- If two InfiniBand adapters are installed in a cluster node, leave the second port on each adapter unused. For example, connect port 1 on HCA 1 to switch 1 and connect port 1 on HCA 2 to switch 2.
- VLANs are not supported on a cluster that uses InfiniBand switches.

## Installing Ethernet or InfiniBand Cluster Interconnect Hardware

[Table 3–1](#) lists procedures about how to install Ethernet cluster interconnect hardware. Perform the procedures in the order that the procedures are listed. This section contains a procedure about how to install cluster hardware during an *initial installation* of a cluster, before you install Sun Cluster software.

TABLE 3–1 Task Map: Installing Ethernet Cluster Interconnect Hardware

Task	For Instructions
Install the transport adapters.	The documentation that shipped with your nodes and host adapters
Install the transport cables.	<a href="#">“How to Install Ethernet or InfiniBand Transport Cables and Transport Junctions”</a> on page 40
If you have a greater than two-node cluster, install a transport junction (switch).	<a href="#">“How to Install Ethernet or InfiniBand Transport Cables and Transport Junctions”</a> on page 40

## ▼ How to Install Ethernet or InfiniBand Transport Cables and Transport Junctions

Use this procedure to install Ethernet or InfiniBand transport cables and transport junctions (switches).

### 1 If not already installed, install transport adapters in your cluster nodes.

For the procedure about how to install transport adapters, see the documentation that shipped with your host adapters and node hardware.

### 2 Install the transport cables. If necessary, install transport junctions.

- **(Ethernet Only)** As [Figure 3–1](#) shows, a cluster with only two nodes can use a point-to-point connection, requiring no transport junctions.

**(Ethernet Only)** For a point-to-point connection, you can use either UTP or fibre. With fibre, use a standard patch cable. A crossover cable is unnecessary. With UTP, see your network interface card documentation to determine whether you need a crossover cable.

---

**Note – (Ethernet Only)** You can optionally use transport junctions in a two-node cluster. If you use a transport junction in a two-node cluster, you can more easily add additional nodes later. Anytime you use a transport junction, you must have two transport junctions.

---

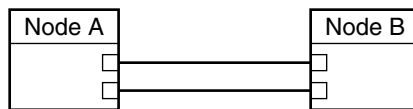


FIGURE 3-1 (Ethernet Only) Typical Two-Node Cluster Interconnect

- As [Figure 3–2](#) shows, cluster with more than two nodes and all clusters using InfiniBand on the interconnect require transport junctions. These transport junctions are Ethernet or InfiniBand switches (customer-supplied).

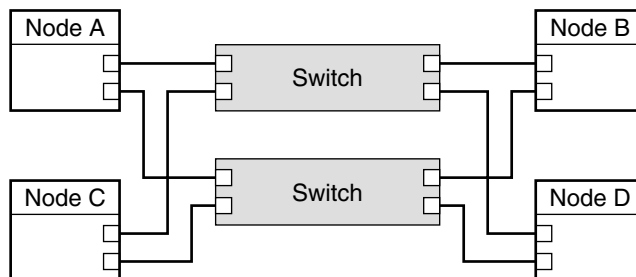


FIGURE 3-2 Typical Four-Node Cluster Interconnect

**See Also** To install the Sun Cluster software and configure the Sun Cluster software with the new interconnect,



see Chapter 2, “Installing and Configuring Sun Cluster Software,” in *Sun Cluster Software Installation Guide for Solaris OS*.

To configure jumbo frames on the interconnect, review the requirements in “SPARC: Requirements When Using Jumbo Frames” on page 38 and see the [Sun GigaSwift documentation](#) for instructions.

## SPARC: Installing PCI-SCI Cluster Interconnect Hardware

Table 3–2 lists procedures about how to install PCI-SCI cluster interconnect hardware. Perform the procedures in the order that the procedures are listed. This section contains a procedure about how to install cluster hardware during an *initial installation* of a cluster before you install Sun Cluster software.

TABLE 3–2 SPARC: Task Map: Installing PCI-SCI Cluster Interconnect Hardware

Task	For Instructions, Go To
Install the transport adapters.	The documentation that shipped with your nodes and host adapters
Install the PCI-SCI transport cables.	“SPARC: How to Install PCI-SCI Transport Cables and Transport Junction” on page 41
If you have a three-node or four-node cluster, install a PCI-SCI transport junction (switch).	“SPARC: How to Install PCI-SCI Transport Cables and Transport Junction” on page 41

### ▼ SPARC: How to Install PCI-SCI Transport Cables and Transport Junction

Use this procedure to install PCI-SCI transport cables and transport junctions (switches).

When you perform this procedure, the following error messages are displayed on your console.

- If you are using Solaris 8:

```
Nov 13 20:11:43 e04a ip: ip_rput_dlpi(scid0): DL_ERROR_ACK for
DL_ENABMULTI_REQ(29), errno 7, unix 0
Nov 13 20:11:43 e04a ip: ip: joining multicasts failed (7) on scid0
- will use link layer broadcasts for multicast
```

These error messages are displayed because the associated driver does not support the multicast feature. These error messages are displayed when the `ip` module probes the driver. Sun Cluster software does not use the multicast feature on the private interconnect. You can safely ignore these error messages.

- If you are using Solaris 9:

```
Dec  4 17:40:14 e03a in.routed[132]: write(rt_sock) RTM_ADD 172.17.0.128/25
-->172.17.0.129    metric=0 flags=0: File exists
Dec  4 17:40:19 e03a in.routed[132]: interface scid0 to 172.17.0.129 broken:
in=0 ierr=0 out=0 oerr=4
```

These error messages are responses to the way Solaris 9 handles SCI dLpi interfaces. Solaris 9 uses the `in.routed` routing protocol as the default routing protocol. You can safely ignore these error messages. The `in.routed` routing protocol is the source of these error messages.

- If you are using Solaris 10, no error messages appear in this situation.

## 1 If not already installed, install PCI-SCI transport adapters in your cluster nodes.

For the procedure about how to install PCI-SCI transport adapters and set their DIP switches, see the documentation that shipped with your PCI-SCI host adapters and node hardware.

---

**Note** – Sbus-SCI host adapters are not supported by Sun Cluster software. If you are upgrading from a Sun Cluster 2.2 cluster, remove any Sbus-SCI host adapters from the cluster nodes. If you do not remove these adapters, you might see panic error messages during the SCI self test.

---

## 2 Install the PCI-SCI transport cables and optionally, transport junctions, depending on how many nodes are in your cluster.

- [Configuration With Point-to-Point Connections:](#)

A two-node cluster can use a point-to-point connection, requiring no transport junction.

- a. Connect the ends of the cables that are marked SCI Out to the I connectors on the adapters.
- b. Connect the ends of the cables that are marked SCI In to the O connectors of the adapters as shown in the previous diagram.

See the following diagrams for cabling details.

- [Configuration With Transport Junction:](#)

A three-node or four-node cluster requires SCI transport junctions.

- a. Set the Unit selectors on the fronts of the SCI transport junctions to F. Do not use the X-Ports on the SCI transport junctions.
- b. Connect the ends of the cables that are marked SCI Out to the I connectors on the adapters and the Out connectors on the transport junctions.
- c. Connect the ends of the cables that are marked SCI In to the O connectors of the adapters and In connectors on the transport junctions, as shown in the previous diagram.

See the following diagrams for cabling details. For the procedure about how to install and cable, see the SCI switch documentation that shipped with your hardware switches.

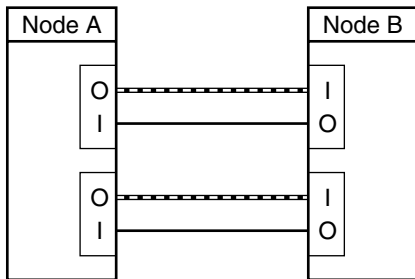


FIGURE 3-3 Configuration With Point-to-Point Connections: Two Interconnects

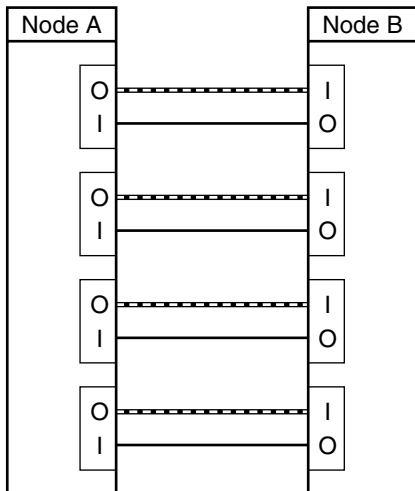


FIGURE 3-4 Configuration With Point-to-Point Connections: Four Interconnects

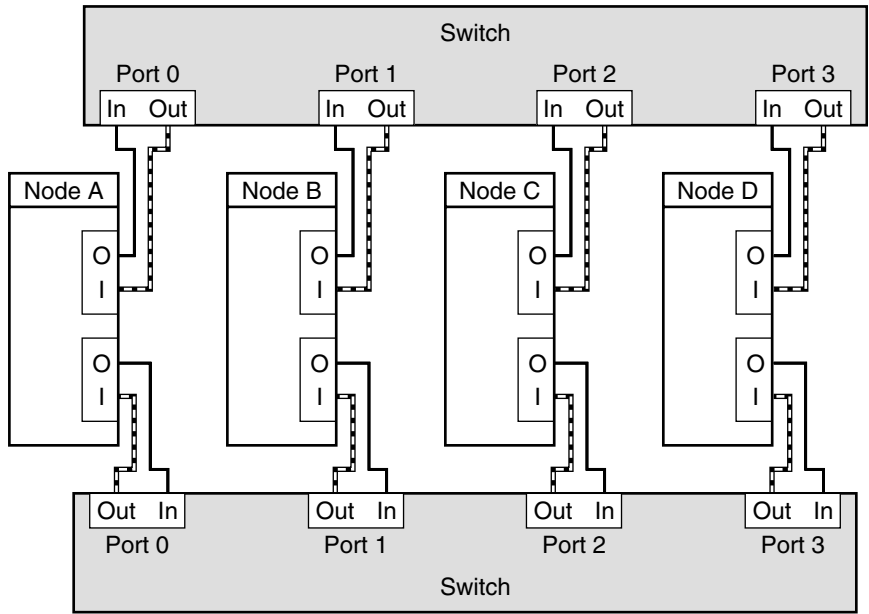


FIGURE 3-5 Configuration With Transport Junction: Two Interconnects

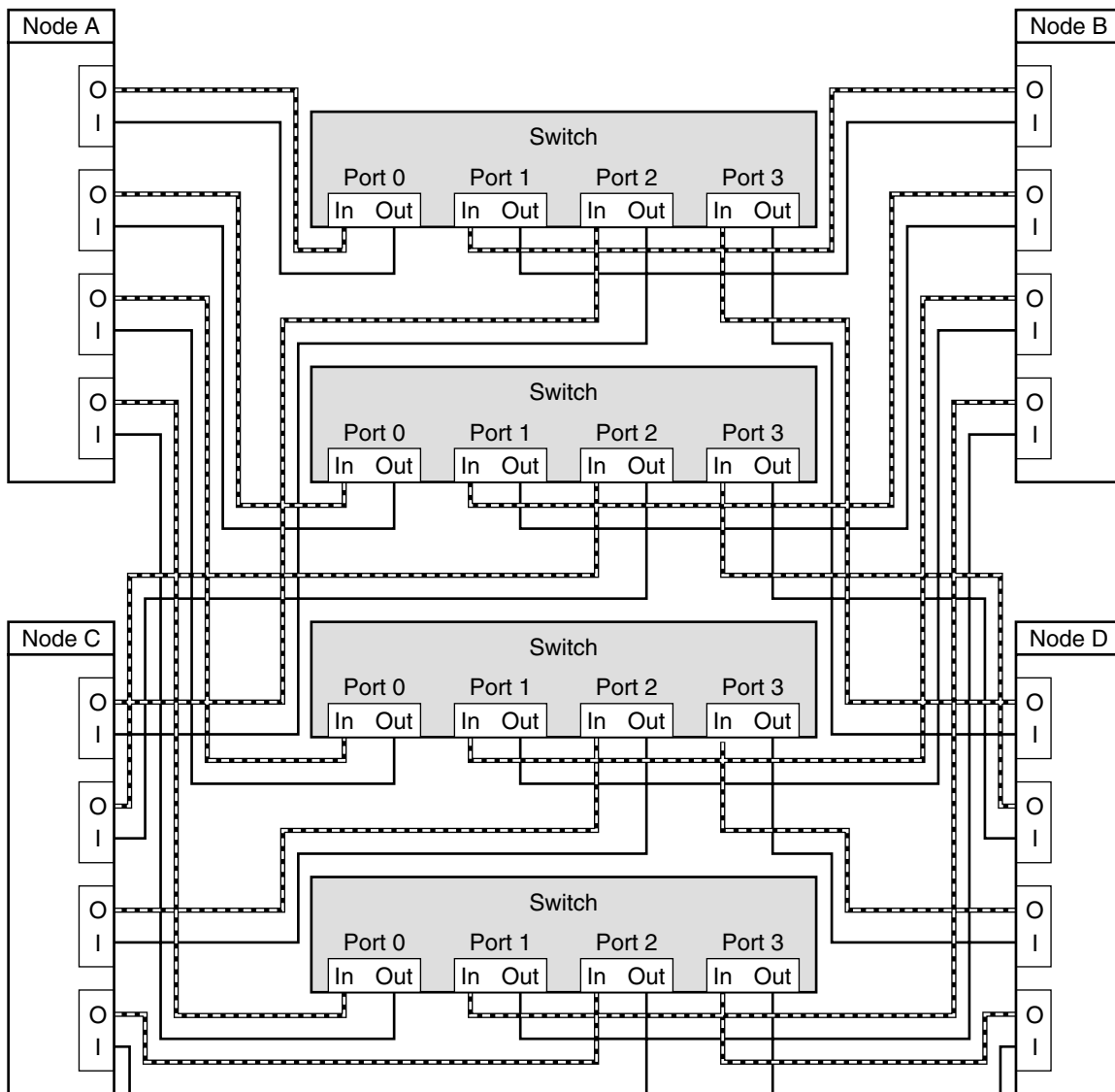


FIGURE 3-6 Configuration With Transport Junction: Four Interconnects

**Troubleshooting** If you have problems with your PCI-SCI interconnect, perform the following tasks:

- Verify that the LED on the PCI-SCI transport adapter is blinking green rapidly. For detailed LED interpretations and actions, see the documentation that shipped with your host adapter.
- Verify that the PCI-SCI transport adapter card's DIP switch settings are correct. For more information, see the documentation that shipped with your PCI-SCI host adapter.

- Verify that the PCI-SCI cables are correctly connected. The PCI-SCI cable connects to the connector that is marked **SCI In** on the PCI-SCI adapter cards. If you are using transport junctions, the PCI-SCI cable also connects to the **Out** ports on the SCI transport junctions.
- Verify that the PCI-SCI cables are correctly connected. The PCI-SCI cable connects to the connector that is marked **SCI Out** on the PCI-SCI adapter cards. If you are using transport junctions, the PCI-SCI cable also connects to the **In** ports on the SCI transport junctions.
- Verify that the PCI-SCI switch unit selectors are set to **F**.

- See Also**
- To increase Oracle Real Application Clusters performance, set the *max-vc-number* parameter in the `/etc/system` file for each node. Choose the value that corresponds to the number of interconnects in your configuration:
    - (2 PCI-SCI interconnects) *max-vc-number* = 32768
    - (3 PCI-SCI interconnects) *max-vc-number* = 49152
    - (4 PCI-SCI interconnects) *max-vc-number* = 65536
  - To install the Sun Cluster software and configure the Sun Cluster software with the new interconnect, see Chapter 2, “Installing and Configuring Sun Cluster Software,” in *Sun Cluster Software Installation Guide for Solaris OS*.

## SPARC: Installing Sun Fire Link Cluster Interconnect Hardware

Table 3–3 lists procedures about how to install Sun Fire Link cluster interconnect hardware. Perform the procedures in the order that the procedures are listed.

TABLE 3–3 SPARC: Task Map: Installing Sun Fire Link Cluster Interconnect Hardware

Task	For Instructions
Install the transport adapters (Sun Fire Link optical module).	<i>Sun Fire Link Hardware Installation Guide</i>
Install the Sun Fire Link transport cables (Sun Fire Link cables)	<i>Sun Fire Link Hardware Installation Guide</i>
If you have a three-node or four-node cluster, install a Sun Fire Link transport junction (Sun Fire Link switch).	<i>Sun Fire Link Hardware Installation Guide</i>
Perform the Sun Fire Link software installation.	<i>Sun Fire Link Software Installation Guide</i>
Create and activate a dual-controller Sun Fire Link fabric.	<i>Sun Fire Link Fabric Administrator's Guide</i>

# Configuring VLANs as Private Interconnect Networks

Sun Cluster software supports the use of private interconnect networks over switch-based virtual local area networks (VLAN). In a switch-based VLAN environment, Sun Cluster software enables multiple clusters and nonclustered systems to share Ethernet transport junction (switch) in two different configurations.

The implementation of switch-based VLAN environments is vendor-specific. Because each switch manufacturer implements VLAN differently, the following guidelines address Sun Cluster software requirements about how to configure VLANs with cluster interconnects.

- You must understand your capacity needs before you set up a VLAN configuration. You must know the minimum bandwidth necessary for your interconnect and application traffic.

For the best results, set the Quality of Service (QOS) level for each VLAN to accommodate basic cluster traffic plus the desired application traffic. Ensure that the bandwidth that is allocated to each VLAN extends from node to node.

To determine the basic cluster traffic requirements, use the following equation. In this equation,  $n$  equals the number of nodes in the configuration, and  $s$  equals the number of switches per VLAN.

$$n (s-1) \times 10\text{Mb}$$

- Interconnect traffic must be placed in the highest-priority queue.
- All ports must be equally serviced, similar to a round robin or first-in, first-out model.
- You must verify that you have properly configured your VLANs to prevent path timeouts.

The first VLAN configuration enables nodes from multiple clusters to send interconnect traffic across one pair of Ethernet transport junctions. Sun Cluster software requires the use of at least two transport junctions to eliminate a single point of failure. The following figure is an example of the first VLAN configuration in a two-node cluster. VLAN configurations are not limited to two-node clusters.

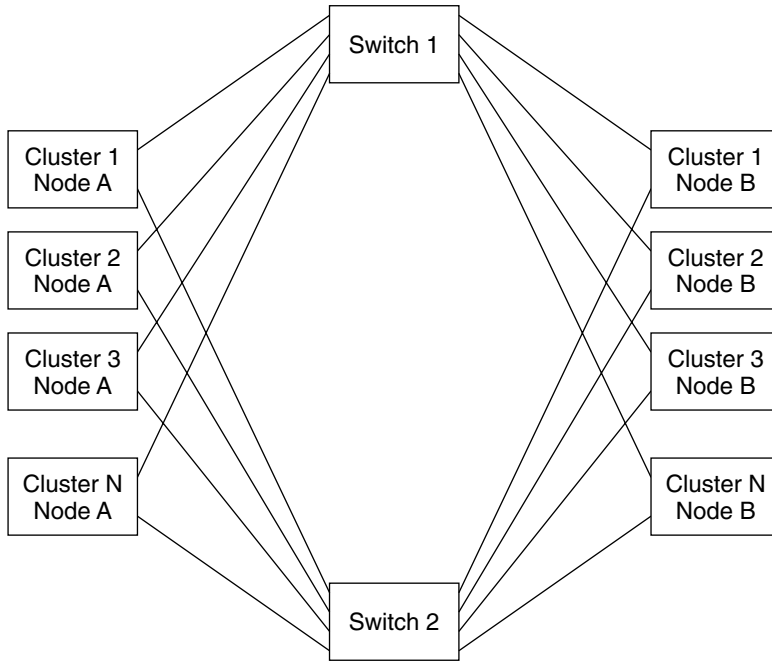


FIGURE 3-7 First VLAN Configuration

The second VLAN configuration also uses the same transport junctions for the interconnect traffic of multiple clusters. However, the second VLAN configuration requires two pairs of transport junctions that are connected by links. This configuration enables VLANs to be supported in a campus cluster configuration with the same restrictions as other campus cluster configurations. The following figure illustrates the second VLAN configuration.

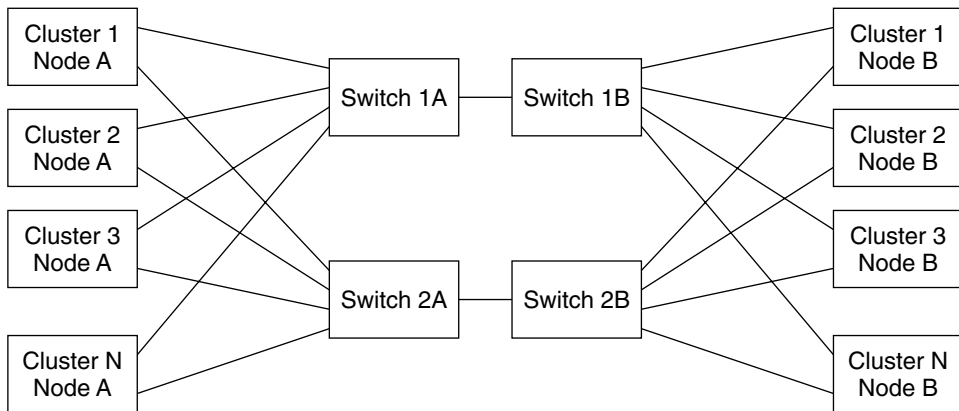


FIGURE 3-8 Second VLAN Configuration



# Maintaining Cluster Interconnect Hardware

---

This chapter describes the procedures to maintain cluster interconnect hardware. The procedures in this chapter apply to all interconnects that Sun Cluster software supports:

- Ethernet
- InfiniBand
- SPARC: Peripheral Component Interconnect-Scalable Coherent Interface (PCI-SCI)
- SPARC: Sun Fire Link

This chapter contains the following procedures:

- “How to Add an Interconnect Component” on page 50
- “How to Replace an Interconnect Component” on page 52
- “How to Remove an Interconnect Component” on page 54
- “How to Upgrade Transport Adapter Firmware” on page 55

For more information, see the following documentation:

- For conceptual information about cluster interconnects, see “Cluster Interconnect” in *Sun Cluster Concepts Guide for Solaris OS*.
- For information about how to administer cluster interconnects, see “Administering the Cluster Interconnects” in *Sun Cluster System Administration Guide for Solaris OS*.

## Maintaining Interconnect Hardware in a Running Cluster

The following table lists procedures about how to maintain cluster interconnect hardware.

TABLE 4-1 Task Map: Maintaining Cluster Interconnect Hardware

Task	Instructions
Add an interconnect component.  Interconnect components include the following components: <ul style="list-style-type: none"> <li>■ Transport adapter</li> <li>■ Transport cable</li> <li>■ Transport junction (switch)</li> </ul>	<a href="#">“How to Add an Interconnect Component” on page 50</a>
Replace an interconnect component.  Interconnect components include the following components: <ul style="list-style-type: none"> <li>■ Transport adapter</li> <li>■ Transport cable</li> <li>■ Transport junction (switch)</li> </ul>	<a href="#">“How to Replace an Interconnect Component” on page 52</a>
Remove an interconnect component.  Interconnect components include the following components: <ul style="list-style-type: none"> <li>■ Transport adapter</li> <li>■ Transport cable</li> <li>■ Transport junction (switch)</li> </ul>	<a href="#">“How to Remove an Interconnect Component” on page 54</a>
Upgrade transport adapter firmware	<a href="#">“How to Upgrade Transport Adapter Firmware” on page 55</a>

## ▼ How to Add an Interconnect Component

This procedure defines interconnect component as any one of the following components:

- Transport adapter
- Transport cable
- Transport junction (switch)

This section contains the procedure about how to add transport adapters to nodes in a running cluster.

**Before You Begin** This procedure relies on the following prerequisites and assumptions:

- Your cluster is operational and all nodes are powered on.
- If Virtual Local Area Networks (VLANs) are configured, more than one cluster might be impacted by removing a transport junction. Ensure that all clusters are prepared for the removal of a transport junction. Also, record the configuration information of the transport junction you plan to replace and configure the new transport junction accordingly.

For more information about how to configure VLANs, see [“Configuring VLANs as Private Interconnect Networks” on page 47](#).

**1 Determine if you need to shut down and power off the node that is to be connected to the interconnect component you are adding.**

- If you are adding a transport junction, you do not need to shut down and power off the node. Proceed to [Step 2](#).
- If you are adding a transport cable, you do not need to shut down and power off the node. Proceed to [Step 2](#).
- If your node has Dynamic Reconfiguration (DR) enabled and you are replacing a transport adapter, you do not need to shut down and power off the node. Proceed to [Step 2](#).
- If your node does *not* have DR enabled and you are adding a transport adapter, shut down and power off the node with the transport adapter you are adding.

For the full procedure about how to shut down a node, see Chapter 3, “Shutting Down and Booting a Cluster,” in *Sun Cluster System Administration Guide for Solaris OS*.

**2 Install the interconnect component.**

- If you are using an Ethernet or InfiniBand interconnect, see [“How to Install Ethernet or InfiniBand Transport Cables and Transport Junctions” on page 40](#) for cabling diagrams and considerations.
- If you are using a PCI-SCI interconnect, see [“SPARC: How to Install PCI-SCI Transport Cables and Transport Junction” on page 41](#) for cabling diagrams and considerations.
- For the procedure about how to install transport adapters or set transport adapter DIP switches, see the documentation that shipped with your host adapter and node hardware.
- If your interconnect uses jumbo frames, review the requirements in [“SPARC: Requirements When Using Jumbo Frames” on page 38](#) and see the [Sun GigaSwift documentation](#) for instructions.

**3 If you shut down the node in [Step 1](#), perform a reconfiguration boot to update the new Solaris device files and links. Otherwise, skip this step.**

- See Also**
- To increase Oracle Real Application Clusters performance, set the *max-vc-number* parameter. See [“SPARC: Installing PCI-SCI Cluster Interconnect Hardware” on page 41](#) for more information.

- To reconfigure Sun Cluster software with the new interconnect component, see Chapter 6, “Administering Cluster Interconnects and Public Networks,” in *Sun Cluster System Administration Guide for Solaris OS*.

## ▼ How to Replace an Interconnect Component

This procedure defines interconnect component as any one of the following components:

- Transport adapter
- Transport cable
- Transport junction (switch)




---

**Caution** – You must maintain at least one cluster interconnect between the nodes of a cluster. The cluster does not function without a working cluster interconnect. You can check the status of the interconnect with the command `scstat -w`. For more details about checking the status of the cluster interconnect, see “How to Check the Status of the Cluster Interconnect” in *Sun Cluster System Administration Guide for Solaris OS*.

---

You might perform this procedure in the following scenarios:

- You need to replace a failed transport adapter.
- You need to replace a failed transport cable.
- You need to replace a failed transport junction.

For conceptual information about transport adapters, transport cables, and transport junction, see “Cluster Interconnect” in *Sun Cluster Concepts Guide for Solaris OS*.

### Before You Begin

This procedure relies on the following prerequisites and assumptions.

- Install a temporary crossover cable before you remove a transport junction. This temporary crossover cable eliminates the single point of failure that exists when you remove a transport junction.
- Your cluster is operational and all nodes are powered on.
- Identify the interconnect component that you want to replace. Remove that interconnect component from the cluster configuration by using the procedure in “How to Remove Cluster Transport Cables, Transport Adapters, and Transport Junctions” in *Sun Cluster System Administration Guide for Solaris OS*.

---

**Note** – PCI-SCI interconnect requires that you remove the interconnect component from the cluster configuration because the SCI driver needs to be updated. If the SCI driver is not updated, the nodes cannot use that interconnect to communicate. Although this practice is optional for all other interconnect technology, remove the interconnect component from the cluster configuration to prevent error messages from appearing on your console.

---

- If Virtual Local Area Networks (VLANs) are configured, more than one cluster might be impacted by removing a transport junction. Ensure that all clusters are prepared for the removal of a transport junction. Also, record the configuration information of the transport junction you plan to replace and configure the new transport junction accordingly.

For more information about how to configure VLANs, see [“Configuring VLANs as Private Interconnect Networks”](#) on page 47.

**1 Determine if you need to shut down and power off the node that is connected to the interconnect component you are replacing.**

- If you are replacing a transport junction, you do not need to shut down and power off the node. Proceed to [Step 2](#).
- If you are replacing a transport cable, you do not need to shut down and power off the node. Proceed to [Step 2](#).
- If your node has DR enabled and you are replacing a transport adapter, you do not need to shut down and power off the node. Proceed to [Step 2](#).
- If your node does *not* have DR enabled and you are replacing a transport adapter, shut down and power off the node with the transport adapter you are replacing.

For the full procedure about how to shut down a node, see Chapter 3, “Shutting Down and Booting a Cluster,” in *Sun Cluster System Administration Guide for Solaris OS*.

**2 Disconnect the failed interconnect component from other cluster devices.**

For the procedure about how to disconnect cables from transport adapters, see the documentation that shipped with your host adapter and node.

**3 Connect the new interconnect component to other cluster devices.**

- If you are replacing an Ethernet or InfiniBand interconnect, see [“How to Install Ethernet or InfiniBand Transport Cables and Transport Junctions”](#) on page 40 for cabling diagrams and considerations.
- If you are replacing a PCI-SCI interconnect, see [“SPARC: How to Install PCI-SCI Transport Cables and Transport Junction”](#) on page 41 for cabling diagrams and considerations.
- If your interconnect uses jumbo frames, review the requirements in [“SPARC: Requirements When Using Jumbo Frames”](#) on page 38 and see the [Sun GigaSwift documentation](#) for instructions. Refer to [“ce Sun Ethernet Driver Considerations”](#) on page 60 for details of how to edit the `ce.conf` file according to the GigaSwift documentation’s instructions.

**4 If you shut down the node in [Step 1](#), perform a reconfiguration boot to update the new Solaris device files and links. Otherwise, skip this step.**

**See Also** To reconfigure Sun Cluster software with the new interconnect component, see [“How to Add Cluster Transport Cables, Transport Adapters, or Transport Junctions”](#) in *Sun Cluster System Administration Guide for Solaris OS*.

## ▼ How to Remove an Interconnect Component

This procedure defines interconnect component as any one of the following components:

- Transport adapter
- Transport cable
- Transport junction (switch)




---

**Caution** – You must maintain at least one cluster interconnect between the nodes of a cluster. The cluster does not function without a working cluster interconnect. You can check the status of the interconnect with the command `scstat -w`. For more details about checking the status of the cluster interconnect, see “How to Check the Status of the Cluster Interconnect” in *Sun Cluster System Administration Guide for Solaris OS*.

---

You might perform this procedure in the following scenarios:

- You need to remove an unused transport adapter.
- You need to remove an unused transport cable.
- You need to remove an unused transport junction.
- You want to migrate from a two–node cluster that uses switches to a point-to-point configuration.

For conceptual information about transport adapters, transport cables, and transport junctions, see “Cluster Interconnect” in *Sun Cluster Concepts Guide for Solaris OS*.

**Before You Begin** This procedure relies on the following prerequisites and assumptions:

- If you are migrating from a two–node cluster that uses switches to a point-to-point configuration, install a crossover cable before you remove a switch.
- Your cluster is operational and all nodes are powered on.
- Identify the interconnect component that you want to remove. Remove that interconnect component from the cluster configuration by using the procedure in “How to Remove Cluster Transport Cables, Transport Adapters, and Transport Junctions” in *Sun Cluster System Administration Guide for Solaris OS*.

---

**Note** – PCI-SCI interconnect requires that you remove the interconnect component from the cluster configuration because the SCI driver needs to be updated. If the SCI driver is not updated, the nodes cannot use that interconnect to communicate. Although this practice is optional for all other interconnect technology, remove the interconnect component from the cluster configuration to prevent error messages from appearing on your console.

---

- If you plan to use virtual local area networks (VLANs) in your cluster interconnect, configure the transport junction. For more information about how configure VLANs, see “[Configuring VLANs as Private Interconnect Networks](#)” on page 47.

- 1 **Determine if you need to shut down and power off the node that is connected to the interconnect component you are removing.**
  - If you are removing a transport junction you, do not need to shut down and power off the node. Proceed to [Step 2](#).
  - If you are removing a transport cable you, do not need to shut down and power off the node. Proceed to [Step 2](#).
  - If your node has DR enabled and you are removing a transport adapter, you do not need to shut down and power off the node. Proceed to [Step 2](#).
  - If your node does *not* have DR enabled and you are removing a transport adapter, shut down and power off the node with the transport adapter you are removing.  
For the full procedure about how to shut down a node, see Chapter 3, “Shutting Down and Booting a Cluster,” in *Sun Cluster System Administration Guide for Solaris OS*.
- 2 **Disconnect the interconnect component from other cluster devices.**  
For the procedure about how to disconnect cables from transport adapters, see the documentation that shipped with your host adapter and node.
- 3 **Remove the interconnect component.**  
For the procedure about how to remove interconnect component, see the documentation that shipped with your host adapter, nodes, or switch.
- 4 **If you shut down the node in [Step 1](#), perform a reconfiguration boot to update the new Solaris device files and links. Otherwise, skip this step.**

**See Also** To reconfigure Sun Cluster software with the new interconnect component, see “How to Add Cluster Transport Cables, Transport Adapters, or Transport Junctions” in *Sun Cluster System Administration Guide for Solaris OS*.

## ▼ How to Upgrade Transport Adapter Firmware

You might perform this procedure in the following scenarios:

- You want to use firmware bug fixes.
- You want to use new firmware features.

Use this procedure to update transport adapter firmware.

- 1 **Determine the resource groups and the device groups that are running on the node. This node is the node on which you are upgrading transport adapter firmware.**

```
# scstat
```

**2 Migrate the resource groups and device groups off the node on which you plan to upgrade the firmware.**

```
# scswitch -S -h from-node
```

**3 Perform the firmware upgrade.**

For the procedure about how to upgrade your transport adapter firmware, see the patch documentation.

**4 Return the resource groups and device groups to their original node.**

```
# scswitch -z -g resource-group -h nodename
```

```
# scswitch -z -D device-group -h nodename
```



# Installing and Maintaining Public Network Hardware

---

This chapter contains information about how to maintain public network hardware. This chapter covers the following topics.

- “Requirements When Using Jumbo Frames” on page 57
- “Installing Public Network Hardware” on page 58
- “Maintaining Public Network Hardware in a Running Cluster” on page 58
- “Sun Gigabit Ethernet Adapter Considerations” on page 59
- “Sun Ethernet Driver Considerations” on page 60

For conceptual information on cluster interconnects and public network interfaces, see your Sun Cluster concepts documentation.

For information on how to administer public network interfaces, see your Sun Cluster system administration documentation

## Requirements When Using Jumbo Frames

If you are using Ethernet to implement the private interconnect and your cluster is configured with scalable services, then you must configure jumbo frames for both the public and the private networks. The private interconnect must be configured with the same or greater maximum transfer unit (MTU) as the public network.

If you are using SCI or Sun Fire Link to implement the private interconnect, you can use jumbo frames on the public network with no restrictions. For information about how to configure jumbo frames, see [the Sun GigaSwift documentation](#).

All of your cluster interconnect paths must be configured with jumbo frames on when using jumbo frames.

Certain patches are required to use jumbo frames with Sun Cluster. Use the PatchPro tool (<http://www.sun.com/PatchPro/>) to get these patches.

## Installing Public Network Hardware

This section covers installing cluster hardware during an *initial cluster installation, before Sun Cluster software is installed*.

Physically installing public network adapters to a node in a cluster is no different from adding public network adapters in a noncluster environment.

For the procedure about how to add public network adapters, see the documentation that shipped with your nodes and public network adapters.

### Where to Go From Here

Install the cluster software and configure the public network hardware after you have installed all other hardware. To review the task map about how to install cluster hardware, see [“Installing Sun Cluster Hardware” on page 15](#).

If your network uses jumbo frames, review the requirements in [“Requirements When Using Jumbo Frames” on page 57](#) and see the [Sun GigaSwift documentation](#) for information about how to configure jumbo frames.

## Maintaining Public Network Hardware in a Running Cluster

The following table lists procedures about how to maintain cluster interconnect and public network hardware. The interconnect maintenance procedures in this section are for both Ethernet-based and PCI-SCI interconnects.

TABLE 5-1 Task Map: Maintaining Public Network Hardware

Task	Information
Add public network adapters.	<a href="#">“Adding Public Network Adapters” on page 58</a>
Replace public network adapters.	<a href="#">“Replacing Public Network Adapters” on page 59</a>
Remove public network adapters.	<a href="#">“Removing Public Network Adapters” on page 59</a>

### Adding Public Network Adapters

Physically adding public network adapters to a node in a cluster is no different from adding public network adapters in a noncluster environment. For the procedure about how to add public network adapters, see the hardware documentation that shipped with your node and public network adapters.

Once the adapters are physically installed, Sun Cluster requires that they be configured in an IPMP group. For the procedure on configuring an IPMP group, see [“How to Configure Internet Protocol \(IP\) Network Multipathing Groups” in \*Sun Cluster Software Installation Guide for Solaris OS\*](#).

If your network uses jumbo frames, review the requirements in [“Requirements When Using Jumbo Frames” on page 57](#) and see [the Sun GigaSwift documentation](#) for information about how to configure jumbo frames.

## Where to Go From Here

To add a new public network adapter to an IP Network Multipathing group, see the *IP Network Multipathing Administration Guide*.

## Replacing Public Network Adapters

For cluster-specific commands and guidelines about how to replace public network adapters, see your Sun Cluster system administration documentation.

For procedures about how to administer public network connections, see the *IP Network Multipathing Administration Guide*.

For the procedure about removing public network adapters, see the hardware documentation that shipped with your node and public network adapters.

## Where to Go From Here

To add the new public network adapter to a IP Network Multipathing group, see your Sun Cluster system administration documentation.

## Removing Public Network Adapters

For cluster-specific commands and guidelines about how to remove public network adapters, see your Sun Cluster system administration documentation.

For procedures about how to administer public network connections, see the *IP Network Multipathing Administration Guide*.

For the procedure about how to remove public network adapters, see the hardware documentation that shipped with your node and public network adapters.

# Sun Gigabit Ethernet Adapter Considerations

Some Gigabit Ethernet switches require some device parameter values to be set differently than the defaults. Chapter 3 of the *Sun Gigabit Ethernet/P 2.0 Adapter Installation and User's Guide* describes the procedure about how to change device parameters. If you are using an operating system earlier than the Solaris 10 OS, the procedure that you use on nodes that are running Sun Cluster software varies slightly from the procedure that is described in the guide. In particular, the difference is in how you derive parent names for use in the `ge.conf` file from the `/etc/path_to_inst` file.

Chapter 3 of the *Sun Gigabit Ethernet/P 2.0 Adapter Installation and User's Guide* describes the procedure on how to change ge device parameter values. This change occurs through entries in the `/kernel/drv/ge.conf` file. The procedure to derive the parent name from the `/etc/path_to_inst` listing, which is be used in ge.conf entries, appears in *Setting Driver Parameters Using a ge.conf File*. For example, from the following `/etc/path_to_inst` line, you can derive the parent name for ge2 to be `/pci@4,4000`.

```
"/pci@4,4000/network@4" 2 "ge"
```

On Sun Cluster nodes, a `/node@nodeid` prefix appears in the `/etc/path_to_inst` line. Do *not* consider the `/node@nodeid` prefix when you derive the parent name. For example, on a cluster node, an equivalent `/etc/path_to_inst` entry would be the following:

```
"/node@1/pci@4,4000/network@4" 2 "ge"
```

The parent name for ge2, to be used in the ge.conf file is still `/pci@4,4000` in this instance.

## ce Sun Ethernet Driver Considerations

The Sun Cluster software supports the ce Sun Ethernet driver for cluster interconnect and public network applications. Consult your Sun sales representative for details on which network interface products are supported. These network interfaces can only be used as new interconnects and public network interfaces. As such, you cannot upgrade from another type of adapter card to those network interfaces by using the ce Sun Ethernet driver.

When you use the ce Sun Ethernet driver for public network connections, add the following line to your `/etc/system` file.

```
set ce:ce_reclaim_pending=1
```

Apply any required patches, and then reboot the system. See the Sun Cluster system administration documentation for instructions on rebooting a cluster node.

The ce driver version must be at least 1.115. To determine the ce driver version, run the following command.

```
# modinfo | grep CE
```

---

**Note** – For the current list of patches that are required for the ce Sun Ethernet driver, refer to PatchPro. PatchPro is available online to Sun service providers and to customers with service contracts at the PatchPro site: <http://patchpro.sun.com>.

---

## GigaSwift Ethernet Driver and Jumbo Frames

If you are using jumbo frames, you must edit the `ce.conf` file to configure them, as explained in [the Sun GigaSwift documentation](#).

The driver documentation instructs you to `grep` certain entries from the `/etc/path_to_inst` file to determine your entries for the `ce.conf` file. If you are using an operating system earlier than the Solaris 10 OS, the OS modifies the entries on Sun Cluster nodes, adding a `node-identifier` prefix to them. For example, an entry modified for a Sun Cluster node resembles the following:

```
# grep ce /etc/path_to_inst
"/node@1/pci@8,600000/network@1" 0 "ce"
```

When editing the `ce.conf` file, remove the `/node@nodeID` identifier prefix from the entries that you put into the driver configuration file. For the example above, the entry to put into the configuration file is:

```
"/pci@8,600000/network@1" 0 "ce"
```



# Maintaining Platform Hardware

---

This chapter contains information about node hardware in a cluster environment. It contains the following topics:

- “Mirroring Internal Disks on Servers that Use Internal Hardware Disk Mirroring or Integrated Mirroring” on page 63
- “Configuring Cluster Nodes With a Single, Dual-Port HBA” on page 66

## Mirroring Internal Disks on Servers that Use Internal Hardware Disk Mirroring or Integrated Mirroring

Some servers support the mirroring of internal hard drives (internal hardware disk mirroring or integrated mirroring) to provide redundancy for node data. To use this feature in a cluster environment, follow the steps in this section.

Depending on the version of the Solaris operating system you use, you might need to install a patch to correct change request 5023670 and ensure the proper operation of internal mirroring. Check the PatchPro site to find the patch for your server.

The best way to set up hardware disk mirroring is to perform RAID configuration during cluster installation, before you configure multipathing. (For instructions on performing this configuration, see Chapter 1, “Sun Cluster 3.1 8/05 Release Notes Supplement,” in *Sun Cluster 3.0-3.1 Release Notes Supplement*.) If you need to change your mirroring configuration after you have established the cluster, you must perform some cluster-specific steps to clean up the device IDs.

---

**Note** – Specific servers might have additional restrictions. See the documentation that shipped with your server hardware.

---

For specifics about how to configure your server’s internal disk mirroring, refer to the documents that shipped with your server and the `raidctl(1M)` man page.

This section contains the following procedures:

- “How to Configure Internal Disk Mirroring After the Cluster is Established” on page 64
- “How to Remove an Internal Disk Mirror” on page 65

## ▼ How to Configure Internal Disk Mirroring After the Cluster is Established

### Before You Begin

This procedure assumes that you have already installed your hardware and software and have established the cluster. To configure an internal disk mirror during cluster installation, see the *Sun Cluster Installation Guide*.

Check the PatchPro site for any patches required for using internal disk mirroring.

PatchPro is a patch-management tool that eases the selection and download of patches required for installation or maintenance of Sun Cluster software. PatchPro provides an Interactive Mode tool especially for Sun Cluster. The Interactive Tool makes the installation of patches easier. PatchPro’s Expert Mode tool helps you to maintain your configuration with the latest set of patches. Expert Mode is especially useful for obtaining all of the latest patches, not just the high availability and security patches.

To access the PatchPro tool for Sun Cluster software, go to <http://www.sun.com/PatchPro/>, click Sun Cluster, then choose either Interactive Mode or Expert Mode. Follow the instructions in the PatchPro tool to describe your cluster configuration and download the patches.

For third-party firmware patches, see the SunSolve<sup>SM</sup> Online site at <http://sunsolve.ebay.sun.com>.

### 1 If necessary, prepare the node for establishing the mirror.

#### a. Determine the resource groups and device groups that are running on the node.

Record this information because you use this information in later in this procedure to return resource groups and device groups to the node.

```
# scstat
```

#### b. If necessary, move all resource groups and device groups off the node.

```
# scswitch -S -h fromnode
```

### 2 Configure the internal mirror.

```
# raidctl -c clt0d0 clt1d0
```

```
-c clt0d0 clt1d0
```

Creates the mirror of primary disk to the mirror disk. Enter the name of your primary disk as the first argument. Enter the name of the mirror disk as the second argument.



**3 Boot the node into single user mode.**

```
# reboot -- -S
```

**4 Clean up the device IDs.**

```
# sctdidadm -R /dev/rdisk/clt0d0
```

`-R /dev/rdisk/clt0d0` Updates the cluster's record of the device IDs for the primary disk. Enter the name of your primary disk as the argument.

**5 Confirm that the mirror has been created and only the primary disk is visible to the cluster.**

```
# sctdidadm -l
```

The command lists only the primary disk as visible to the cluster.

**6 Boot the node back into cluster mode.**

```
# reboot
```

**7 If you are using Solstice DiskSuite or Solaris Volume Manager and if the state database replicas are on the primary disk, recreate the state database replicas.**

```
# metadb -afc 3 /dev/rdisk/clt0d0s4
```

**8 If you moved device groups off the node in Step 1, move all device groups back to the node.**

Perform the following step for each device group you want to return to the original node.

```
# scswitch -z -D devicegroup -h nodename
```

In these commands, *devicegroup* is one or more device groups that are returned to the node.

**9 If you moved resource groups off the node in Step 1, move all resource groups back to the node.**

```
# scswitch -z -g resourcegroup -h nodename
```

**▼ How to Remove an Internal Disk Mirror****1 If necessary, prepare the node for removing the mirror.****a. Determine the resource groups and device groups that are running on the node.**

Record this information because you use this information later in this procedure to return resource groups and device groups to the node.

```
# scstat
```

**b. If necessary, move all resource groups and device groups off the node.**

```
# scswitch -S -h fromnode
```

**2 Remove the internal mirror.**

```
# raidctl -d clt0d0
```

-d *clt0d0* Deletes the mirror of primary disk to the mirror disk. Enter the name of your primary disk as the argument.

**3 Boot the node into single user mode.**

```
# reboot -- -S
```

**4 Clean up the device IDs.**

```
# scdidadm -R /dev/rdisk/clt0d0
```

```
# scdidadm -R /dev/rdisk/clt1d0
```

```
-R /dev/rdisk/clt0d0
```

-R /dev/rdisk/clt1d0 Updates the cluster's record of the device IDs. Enter the names of your disks separated by spaces.

Confirm that the mirror has been deleted and that both disks are visible.

```
# scdidadm -l
```

The command lists both disks as visible to the cluster.

**5 Boot the node back into cluster mode.**

```
# reboot
```

**6 If you are using Solstice DiskSuite or Solaris Volume Manager and if the state database replicas are on the primary disk, recreate the state database replicas.**

```
# metadb -c 3 -ag /dev/rdisk/clt0d0s4
```

**7 If you moved device groups off the node in [Step 1](#), return the device groups to the original node.**

```
# scswitch -z -D devicegroup -h nodename
```

**8 If you moved resource groups off the node in [Step 1](#), return the resource groups and device groups to the original node.**

```
# scswitch -z -g resourcegroup -h nodename
```

## Configuring Cluster Nodes With a Single, Dual-Port HBA

This section explains the use of dual-port host bus adapters (HBAs) to provide both connections to shared storage in the cluster. While Sun Cluster supports this configuration, it is less redundant than the recommended configuration. You *must* understand the risks that a dual-port HBA configuration poses to the availability of your application, if you choose to use this configuration.

This section contains the following topics:

- “Risks and Trade-offs When Using One Dual-Port HBA” on page 67
- “Supported Configurations When Using a Single, Dual-Port HBA” on page 68
- “Cluster Configuration When Using Solaris Volume Manager and a Single Dual-Port HBA” on page 68
- “Cluster Configuration When Using Solaris Volume Manager for Sun Cluster and a Single Dual-Port HBA” on page 69

## Risks and Trade-offs When Using One Dual-Port HBA

Sun recommends that you strive for as much separation and hardware redundancy as possible when connecting each cluster node to shared data storage. This approach provides the following advantages to your cluster:

- The best assurance of high availability for your clustered application
- Good failure isolation
- Good maintenance robustness

Sun Cluster is usually layered on top of a volume manager, mirrored data with independent I/O paths, or a multipathed I/O link to a hardware RAID arrangement. Therefore, the cluster software does not expect a node ever to ever lose access to shared data. These redundant paths to storage ensure that the cluster can survive any single failure.

Sun Cluster does support certain configurations that use a single, dual-port HBA to provide the required two paths to the shared data. However, using a single, dual-port HBA for connecting to shared data increases the vulnerability of your cluster. If this single HBA fails and takes down both ports connected to the storage device, the node is unable to reach the stored data. How the cluster handles such a dual-port failure depends on several factors:

- the cluster configuration
- the volume manager configuration
- the node on which the failure occurs
- the state of the cluster when the failure occurs

If you choose one of these configurations for your cluster, you must understand that the supported configurations mitigate the risks to high availability and the other advantages. The supported configurations do not eliminate these previously mentioned risks.

## Supported Configurations When Using a Single, Dual-Port HBA

Sun Cluster supports the following volume manager configurations when you use a single, dual-port HBA for connecting to shared data:

- Solaris Volume Manager with more than one disk in each diskset and no dual-string mediators configured. For details about this configuration, see [“Cluster Configuration When Using Solaris Volume Manager and a Single Dual-Port HBA”](#) on page 68.
- Solaris Volume Manager for Sun Cluster. For details about this configuration, see [“Cluster Configuration When Using Solaris Volume Manager for Sun Cluster and a Single Dual-Port HBA”](#) on page 69.

## Cluster Configuration When Using Solaris Volume Manager and a Single Dual-Port HBA

If the Solaris Volume Manager metadbs lose replica quorum for a diskset on a cluster node, the volume manager panics the cluster node. Sun Cluster then takes over the diskset on a surviving node and your application fails over to a secondary node.

To ensure that the node panics and is fenced off if it loses its connection to shared storage, configure each metaset with at least two disks. In this configuration, the metadbs stored on the disks create their own replica quorum for each diskset.

Dual-string mediators are not supported in Solaris Volume Manager configurations that use a single dual-port HBA. Using dual-string mediators prevents the service from failing over to a new node.

### Configuration Requirements

When configuring Solaris Volume Manager metasets, ensure that each metaset contains at least two disks. Do not configure dual-string mediators.

### Expected Failure Behavior

When a dual-port HBA fails with both ports in this configuration, the cluster behavior depends on whether the affected node is primary for the diskset.

- If the affected node is primary for the diskset, Solaris Volume Manager panics that node because it lacks required state database replicas. Your cluster reforms with the nodes that achieve quorum and brings the diskset online on a new primary node.
- If the affected node is not primary for the diskset, your cluster remains in a degraded state.

## Failure Recovery

Follow the instructions for replacing an HBA in your storage device documentation.

## Cluster Configuration When Using Solaris Volume Manager for Sun Cluster and a Single Dual-Port HBA

Because Solaris Volume Manager for Sun Cluster uses raw disks only and is specific to Oracle RAC, no special configuration is required.

### Expected Failure Behavior

When a dual-port HBA fails and takes down both ports in this configuration, the cluster behavior depends on whether the affected node is the current master for the multi-owner diskset.

- If the affected node is the current master for the multi-owner diskset, the node does not panic. If any other node fails or is rebooted, the affected node will panic when it tries to update the replicas. The volume manager chooses a new master for the diskset if the surviving nodes can achieve quorum.
- If the affected node is not the current master for the multi-owner diskset, the node remains up but the device group is in a degraded state. If an additional failure affects the master node and Solaris Volume Manager for Sun Cluster attempts to remaster the diskset on the node with the failed paths, that node will also panic. A new master will be chosen if any surviving nodes can achieve quorum.

## Failure Recovery

Follow the instructions for replacing an HBA in your storage device documentation.



# SPARC: Campus Clustering With Sun Cluster Software

---

In campus clustering, nodes or groups of nodes are located in separate rooms, sometimes several kilometers apart. In addition to providing the usual benefits of using a Sun cluster, properly designed campus clusters can generally survive the loss of any single room and continue to provide their services.

This chapter introduces the basic concepts of campus clustering and provides some configuration and setup examples. The following topics are covered:

- “SPARC: Requirements for Designing a Campus Cluster” on page 71
- “SPARC: Guidelines for Designing a Campus Cluster” on page 73
- “SPARC: Determining Campus Cluster Interconnect Technologies” on page 80
- “SPARC: Installing and Configuring Interconnect, Storage, and Fibre-Channel Hardware” on page 81
- “SPARC: Additional Campus Cluster Configuration Examples” on page 82

This chapter does not attempt to explain clustering, provide information about clustering administration, or furnish details about hardware installation and configuration. For conceptual information about clustering, see your Sun Cluster concepts documentation and your Sun Cluster system administration documentation.

## SPARC: Requirements for Designing a Campus Cluster

When designing your campus cluster, all the requirements for a noncampus cluster still apply. Plan your cluster to eliminate any single point of failure in nodes, cluster interconnect, data storage, and public network. Just as in the standard cluster, a campus cluster requires redundant connections and switches. Disk multipathing is useful to help ensure that each node can always access each shared storage device. These concerns are universal for Sun Cluster.

After you have a valid cluster plan, follow the remaining requirements in this section to ensure a proper campus cluster. To achieve maximum benefits from your campus cluster, consider implementing the “SPARC: Guidelines for Designing a Campus Cluster” on page 73.

## SPARC: Selecting Networking Technologies

Your campus cluster must observe all requirements and limitations of the technologies you choose to use. “[SPARC: Determining Campus Cluster Interconnect Technologies](#)” on page 80 provides a list of tested technologies and their known limitations.

When planning your cluster interconnect, remember that campus clustering requires redundant physical (not logical) network connections.

## SPARC: Connecting to Storage

A campus cluster must include at least two rooms using two independent SANs to connect to the shared storage. See [Figure 7-1](#) for an illustration of this configuration.

Additional rooms need not be fully connected to the shared storage. However, if you are using Oracle Real Application Clusters (RAC), all nodes which support Oracle RAC must be fully connected to the shared storage devices. See “[SPARC: Quorum in Clusters With Four Rooms or More](#)” on page 78 for a description of a campus cluster with both direct and indirect storage connections.

## SPARC: Sharing Data Storage

Your campus cluster must use SAN-supported storage devices for shared storage. When planning the cluster, ensure that it adheres to the SAN requirements for all storage connections. See the [SAN Solutions](#) documentation site for information on SAN requirements.

You must mirror a campus cluster’s shared data. If one room of the cluster is lost, another room must be able to provide access to the data. Therefore, data replication between shared disks must always be performed across rooms, rather than within rooms. Both copies of the data should never be in a single room. Host-based mirroring is required for all campus cluster configurations, because hardware RAID alone does not lend itself to providing data redundancy across rooms.

In addition to mirroring your data, you can add storage-based data replication if you judge that your campus cluster needs the additional data redundancy. See [Appendix A](#) for more information on storage-based data replication.

## SPARC: Complying With Quorum Device Requirements

You must use a quorum device for a two-node cluster. For larger clusters, a quorum device is optional. These are standard cluster requirements.

In addition, you can configure quorum devices to ensure that specific rooms can form a cluster in the event of a failure. For guidelines about where to locate your quorum device, see “[SPARC: Deciding How to Use Quorum Devices](#)” on page 77.



## SPARC: Replicating Solaris Volume Manager Disksets

If you use Solstice DiskSuite/Solaris Volume Manager as your volume manager for shared device groups, carefully plan the distribution of your replicas. In two-room configurations, all disksets should be configured with an additional replica in the room that houses the cluster quorum device. For example, in three-room two-node configurations, a single room houses both the quorum device and at least one extra disk configured into each of the disksets. Each diskset should have extra replicas in the third room.

---

**Note** – You can use a quorum disk for these replicas.

---

Refer to your Solstice DiskSuite/Solaris Volume Manager documentation for details on configuring diskset replicas.

## SPARC: Guidelines for Designing a Campus Cluster

In planning a campus cluster, your goal is to build a cluster that can at least survive the loss of a room and continue to provide services. The concept of a room must shape your planning of redundant connectivity, storage replication, and quorum. Use the following guidelines to help you manage these design considerations.

### SPARC: Determining the Number of Rooms in Your Cluster

The concept of a room, or location, adds a layer of complexity to the task of designing a campus cluster. Think of a **room** as a functionally independent hardware grouping, such as a node and its attendant storage or a quorum device physically separated from any nodes. Each room is separated from other rooms to increase the likelihood of failover and redundancy in case of accident or failure. The definition of a room therefore depends on the type of failure to safeguard against, as described in [Table 7-1](#).

**TABLE 7-1** SPARC: Definitions of “Room”

Failure Scenario	Sample Definitions of “Room”
Power-line failure	Isolated and independent power supplies
Minor accidents, furniture collapse, water seepage	Different parts of a physical room
Small fire, fire sprinklers starting	Different physical areas (for example, sprinkler zone)
Structural failure, building-wide fire	Different buildings

**TABLE 7-1 SPARC: Definitions of “Room”** *(Continued)*

Failure Scenario	Sample Definitions of “Room”
Large-scale natural disaster (for example, earthquake or flood)	Different corporate campuses up to several kilometers apart

Your campus cluster can consist of two rooms, each containing a combination of one or more nodes and storage devices. However, a properly configured cluster with three or more rooms will be more resilient if a failure occurs.

Whenever a two-room campus cluster loses a room, it has only a 50 percent chance of remaining available. If the room with fewest quorum votes is the surviving room, the surviving nodes cannot form a cluster. In this case, your cluster requires manual intervention from your Sun service provider before it can become available.

The advantage of a three-room or larger cluster is that, if any one of the three rooms is lost, automatic failover can be achieved. **Only a properly configured three-room or larger campus cluster can guarantee system availability if an entire room is lost (assuming no other failures).**

Sun Cluster does support two-room campus clusters. These clusters are valid and might offer nominal insurance against disasters. However, consider adding a small third room, possibly even a secure closet or vault (with a separate power supply and proper cabling), to contain the quorum device or a third server.

## SPARC: Three-Room Campus Cluster Examples

A three-room campus cluster configuration supports up to eight nodes. Three rooms enable you to arrange your nodes and quorum device so that your campus cluster can reliably survive the loss of a single room and still provide cluster services. The following example configurations all follow the campus cluster requirements and the design guidelines described in this chapter.

- **Figure 7-1** shows a three-room, two-node campus cluster. In this arrangement, two rooms each contain a single node and an equal number of disk arrays to mirror shared data. The third room contains at least one disk subsystem, attached to both nodes and configured with a quorum device.
- **Figure 7-2** shows an alternative three-room, two-node campus cluster.
- **Figure 7-3** shows a three-room, three-node cluster. In this arrangement, two rooms each contain one node and an equal number of disk arrays. The third room contains a small server, which eliminates the need for a storage array to be configured as a quorum device.

---

**Note** – These examples illustrate general configurations and are not intended to indicate required or recommended setups. For simplicity, the diagrams and explanations concentrate only on features unique to understanding campus clustering. For example, public-network Ethernet connections are not shown.

---

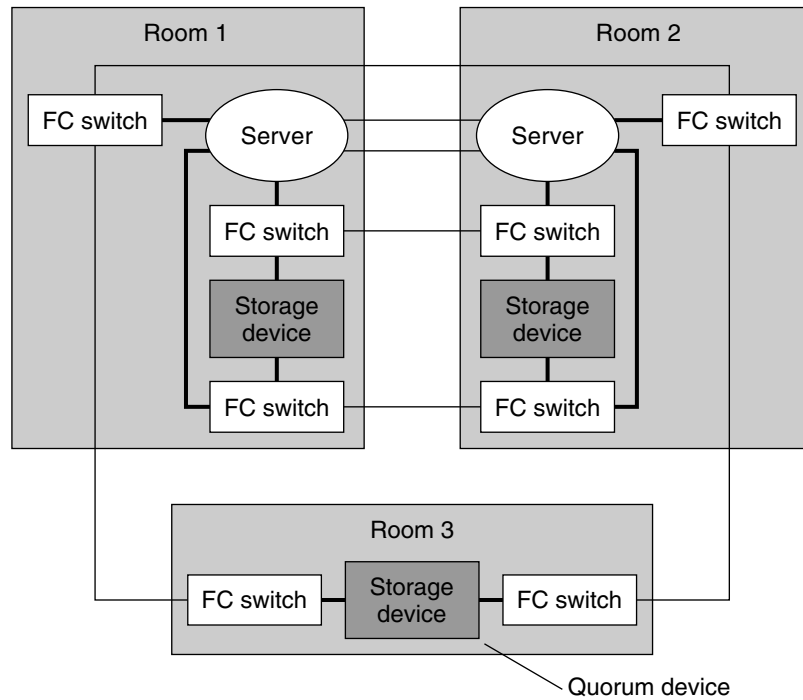
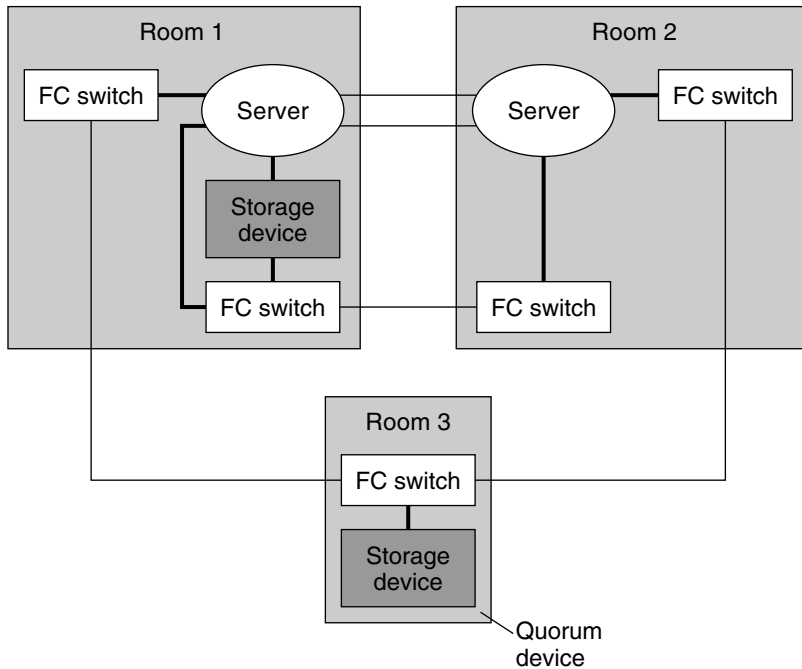


FIGURE 7-1 SPARC: Basic Three-Room, Two-Node Campus Cluster Configuration With Multipathing

In the configuration shown in [Figure 7-1](#), if at least two rooms are up and communicating, recovery is automatic. Only three-room or larger configurations can guarantee that the loss of any one room can be handled automatically. Loss of two rooms requires the replacement or rebuilding of one room and typically requires Sun service provider intervention.



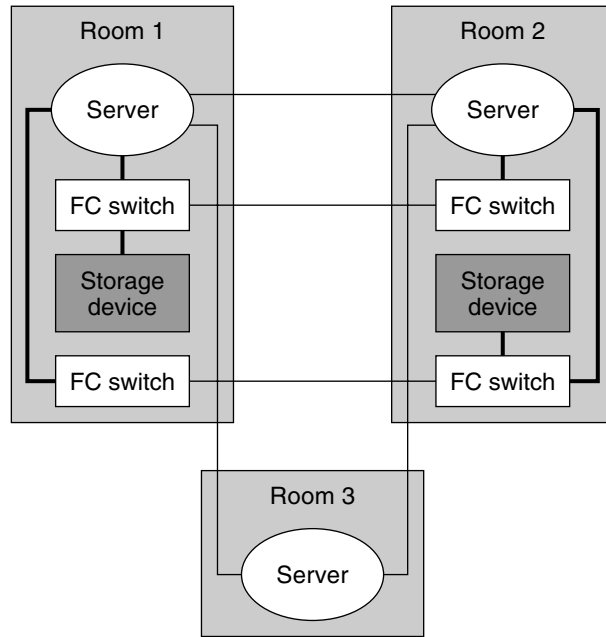
—— Single-mode fiber

—— Multimode fiber

**FIGURE 7-2** SPARC: Minimum Three-Room, Two-Node Campus Cluster Configuration Without Multipathing

In the configuration shown in [Figure 7-2](#), one room contains one node and shared storage. A second room contains a cluster node only. The third room contains shared storage only. A LUN or disk of the storage device in the third room is configured as a quorum device.

This configuration provides the reliability of a three-room cluster with minimum hardware requirements. This campus cluster can survive the loss of any single room without requiring manual intervention.



— Single-mode fiber

— Multimode fiber

FIGURE 7-3 SPARC: Three-Room, Three-Node Campus Cluster Configuration

In the configuration shown in [Figure 7-3](#), a server acts as the quorum vote in the third room. This server does not necessarily support data services. Instead, it replaces a storage device as the quorum device.

## SPARC: Deciding How to Use Quorum Devices

When adding quorum devices to your campus cluster, your goal should be to balance the number of quorum votes in each room. No single room should have a much larger number of votes than the other rooms, because loss of that room can bring the entire cluster down.

For campus clusters with more than three rooms and three nodes, quorum devices are optional. Whether you use quorum devices in such a cluster, and where you place them, depends on your assessment of the following:

- Your particular cluster topology
- The specific characteristics of the rooms involved
- Resiliency requirements for your cluster

As with two-room clusters, locate the quorum device in a room you determine is more likely to survive any failure scenario. Alternatively, you can locate the quorum device in a room that you *want* to form a cluster, in the event of a failure. Use your understanding of your particular cluster requirements to balance these two criteria.

Refer to your Sun Cluster concepts documentation for general information about quorum devices and how they affect clusters that experience failures. If you decide to use one or more quorum devices, consider the following recommended approach:

1. For each room, total the quorum votes (nodes) for that room.
2. Define a quorum device in the room that contains the lowest number of votes and that contains a fully connected shared storage device.

When your campus cluster contains more than two nodes, *do not* define a quorum device if each room contains the same number of nodes.

The following sections discuss quorum devices in various sizes of campus clusters.

- [“SPARC: Quorum in Clusters With Four Rooms or More” on page 78](#)
- [“SPARC: Quorum in Three-Room Configurations” on page 80](#)
- [“SPARC: Quorum in Two-Room Configurations” on page 80](#)

## **SPARC: Quorum in Clusters With Four Rooms or More**

[Figure 7-4](#) illustrates a four-node campus cluster with fully connected storage. Each node is in a separate room. Two rooms also contain the shared storage devices, with data mirrored between them.

Note that the quorum devices are marked *optional* in the illustration. This cluster does not require a quorum device. With no quorum devices, the cluster can still survive the loss of any single room.

Consider the effect of adding *Quorum Device A*. Because the cluster contains four nodes, each with a single quorum vote, the quorum device receives three votes. Four votes (one node and the quorum device, or all four nodes) are required to form the cluster. This configuration is not optimal, because the loss of *Room 1* brings down the cluster. The cluster is not available after the loss of that single room.

If you then add *Quorum Device B*, both *Room 1* and *Room 2* have four votes. Six votes are required to form the cluster. This configuration is clearly better, as the cluster can survive the random loss of any single room.

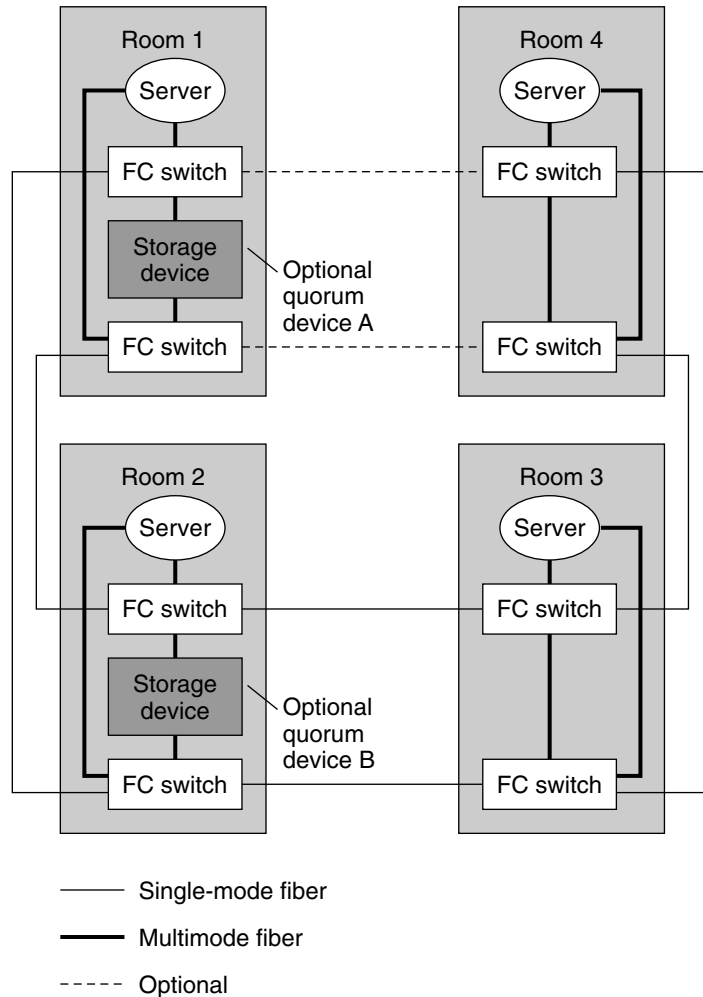


FIGURE 7-4 SPARC: Four-Room, Four-Node Campus Cluster

**Note** – In Figure 7-4, the cluster interconnect is not shown.

Consider the optional I/O connection between *Room 1* and *Room 4*. Although fully connected storage is preferable for reasons of redundancy and reliability, fully redundant connections might not always be possible in campus clusters. Geography might not accommodate a particular connection, or the project's budget might not cover the additional fiber.

In such a case, you can design a campus cluster with indirect access between some nodes and the storage. In Figure 7-4, if the optional I/O connection is omitted, *Node 4* must access the storage indirectly.

## SPARC: Quorum in Three-Room Configurations

In three-room, two-node campus clusters, you should use the third room for the quorum device (Figure 7-1) or a server (Figure 7-3). Isolating the quorum device gives your cluster a better chance to maintain availability after the loss of one room. If at least one node and the quorum device remain operational, the cluster can continue to operate.

## SPARC: Quorum in Two-Room Configurations

In two-room configurations, the quorum device occupies the same room as one or more nodes. Place the quorum device in the room that is more likely to survive a failure scenario if all cluster transport and disk connectivity are lost between rooms. If **only** cluster transport is lost, the node that shares a room with the quorum device is not necessarily the node that reserves the quorum device first. For more information about quorum and quorum devices, see the Sun Cluster concepts documentation.

# SPARC: Determining Campus Cluster Interconnect Technologies

This section lists the supported technologies for the private cluster interconnect and the public networks and their various distance limits. You must observe these limits when constructing a campus cluster.

## SPARC: Cluster Interconnect Technologies

Table 7-2 lists supported node-to-node link technologies and their limitations.

TABLE 7-2 SPARC: Campus Cluster Interconnect Technologies and Distance Limitations

Link Technology	Maximum Distance
100 Mbps Ethernet, unshielded twisted pair (UTP)	100 meters per segment
1000 Mbps Ethernet, UTP	100 meters per segment
1000 Mbps Ethernet, 62.5/125 micron multimode fiber (MMF)	260 meters per segment
1000 Mbps Ethernet, 50/125 micron MMF	550 meters per segment
DWDM	200 kilometers and up

Always check your vendor documentation for technology-specific requirements and limitations.

## SPARC: Storage Area Network Technologies

Table 7-3 lists link technologies for the cluster public network and the distance limits for a single ISL.



TABLE 7-3 SPARC: Interswitch Link-Length Limits

Link Technology	Maximum Distance	Comments
Fibre-channel (FC) short-wave gigabit interface converter (GBIC)	500 meters at 1 Gbps	50/125 micron MMF
FC long-wave GBIC	10 kilometers at 1 Gbps	9/125 micron single-mode fiber (SMF)
FC short-wave small form-factor pluggable (SFP)	300 meters at 2 Gbps	62.5/125 micron MMF
FC short-wave SFP	500 meters at 2 Gbps	62.5/125 micron MMF
FC long-wave SFP	10 kilometers at 2 Gbps	9/125 micron single-mode fiber (SMF)

## SPARC: Installing and Configuring Interconnect, Storage, and Fibre-Channel Hardware

Generally, using interconnect, storage, and FC hardware does not differ markedly from noncampus cluster configurations.

The steps for installing Ethernet-based campus cluster interconnect hardware are the same as the steps for noncampus clusters. Refer to [“Installing Ethernet or InfiniBand Cluster Interconnect Hardware” on page 39](#). When installing the media converters, consult the accompanying documentation, including requirements for fiber connections.

The guidelines for installing virtual local area networks interconnect networks are the same as the guidelines for noncampus clusters. See [“Configuring VLANs as Private Interconnect Networks” on page 47](#).

The steps for installing Sun StorEdge A5x00 and Sun StorEdge T3 or T3+ arrays are the same as the steps for noncampus clusters. Refer to the *Sun Cluster Hardware Administration Collection for Solaris OS* for those steps.

However, when installing Sun StorEdge A5x00 arrays at distances greater than 500m, install the Sun Long Wave GBICs as indicated in the *Sun StorEdge Long Wave GBIC Interface Converter Guide*. This manual also includes single-mode fiber specifications.

Campus clusters require FC switches to mediate between multimode and single-mode fibers. The steps for configuring the settings on the FC switches are very similar to the steps for noncampus clusters.

If your switch supports flexibility in the buffer allocation mechanism, (for example the QLogic switch with donor ports), then make certain to allocation a sufficient number of buffers to the ports which are dedicated to ISLs. If your switch has a fixed number of frame buffers (or buffer credits) per port, you do not have this flexibility.

Use the following rules of thumb for determining the number of buffers you might need:

- For 1 Gbps, calculate buffer credits as:

$$(\text{length-in-km}) \times (0.6)$$

Round the result up to the next whole number. For example, a 10 km connection requires 6 buffer credits and a 7 km connection requires 5 buffer credits.

- For 2 Gbps, calculate buffer credits as:

$$(\text{length-in-km}) \times (1.2)$$

Round the result up to the next whole number. For example, a 10 km connection requires 12 buffer credits while a 7 km connection requires 9 buffer credits.

Refer to your switch documentation for details about computing buffer credits.

## SPARC: Additional Campus Cluster Configuration Examples

While detailing all of the configurations possible in campus clustering is far beyond the scope of this document, the following illustrations depict variations on the configurations previously shown.

The following are three-room examples:

- Three-room campus cluster with a multipathing solution implemented (Figure 7-5)
- Three-room campus cluster with Sun StorEdge A5x00 disk arrays (Figure 7-6)

The following are two-room examples:

- Two-room campus cluster with a multipathing solution implemented (Figure 7-7)
- Two-room campus cluster without a multipathing solution implemented (Figure 7-8)
- Two-room cluster with Sun StorEdge A5x00 disk arrays (Figure 7-9)

---

**Note** – Configurations that include Sun StorEdge A5x00 disk arrays have no FC switches. The SE A5x00 arrays can contain long-wave GBICS, and thus do not require the switches.

---

## SPARC: Three-Room Examples

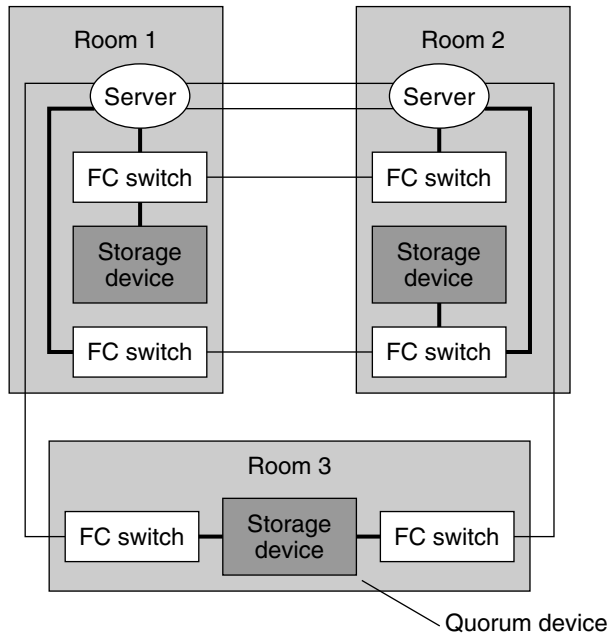
The figures in this section show three-room campus cluster configurations with and without multipathing implementations. Each example configuration follows the campus cluster requirements and design guidelines described in this chapter.

---

**Note** – Configurations include Sun StorEdge A5x00 disk arrays have no FC switches. The SE A5x00 arrays can contain long-wave GBICS, and thus may not require the switches.

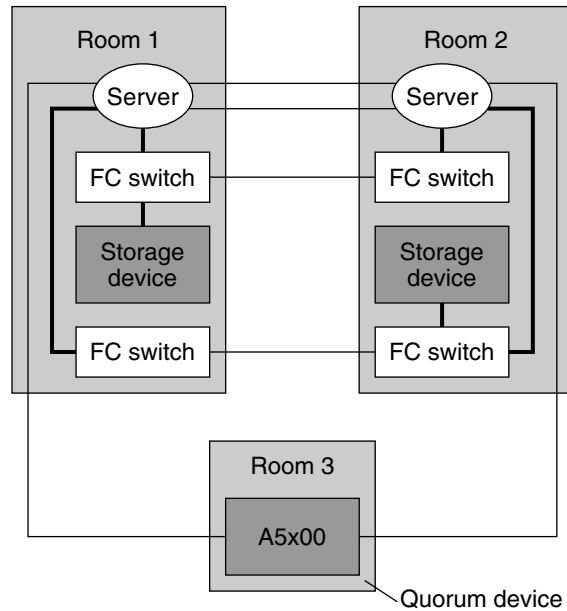
---





— Single-mode fiber  
 — Multimode fiber

FIGURE 7-5 SPARC: Three-Room Campus Cluster With a Multipathing Solution Implemented



— Single-mode fiber

— Multimode fiber

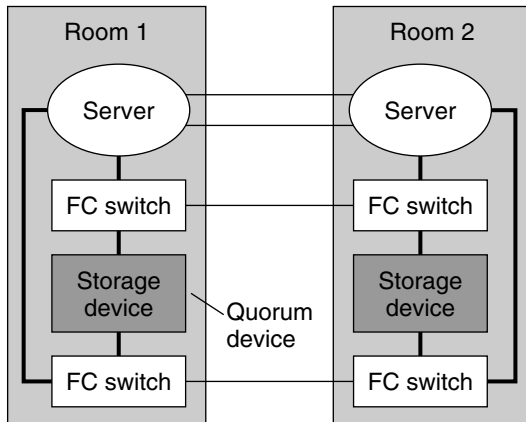
FIGURE 7-6 SPARC: Three-Room Campus Cluster (Sun StorEdge A5x00 as Quorum Device)

In [Figure 7-6](#), note the absence of FC switches in *Room 3*. The storage array connection is made through a Sun StorEdge Dual FC Network Adapter and long-wave GBICs.

## SPARC: Two-Room Examples

[Figure 7-7](#) shows a two-room campus cluster that uses partner pairs of storage devices and four FC switches, with a multipathing solution implemented. The four switches are added to the cluster for greater redundancy and potentially better I/O throughput. This configuration could be implemented by using Sun StorEdge T3 partner groups or Sun StorEdge 9910/9960 arrays with Sun StorEdge Traffic Manager software installed.

For information about Traffic Manager software, see the *Sun StorEdge Traffic Manager Installation and Configuration Guide* at <http://www.sun.com/products-n-solutions/hardware/docs/>.

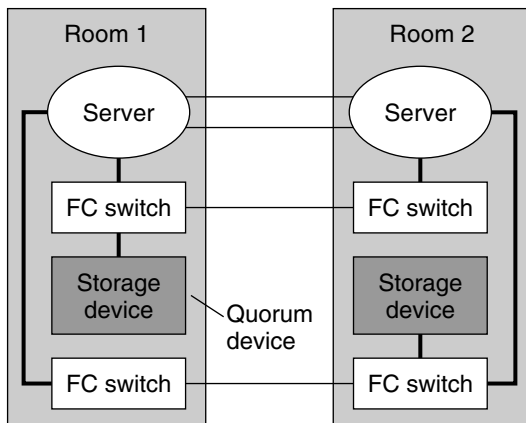


— Single-mode fiber

— Multimode fiber

FIGURE 7-7 SPARC: Two-Room Campus Cluster With a Multipathing Solution Implemented

The configuration in [Figure 7-8](#) could be implemented by using Sun StorEdge T3 or T3+ arrays in single-controller configurations, rather than partner groups.



— Single-mode fiber

— Multimode fiber

FIGURE 7-8 SPARC: Two-Room Campus Cluster Without a Multipathing Solution Implemented

[Figure 7-9](#) depicts a two-room campus cluster that uses Sun StorEdge A5x00s. Note the absence of switches.

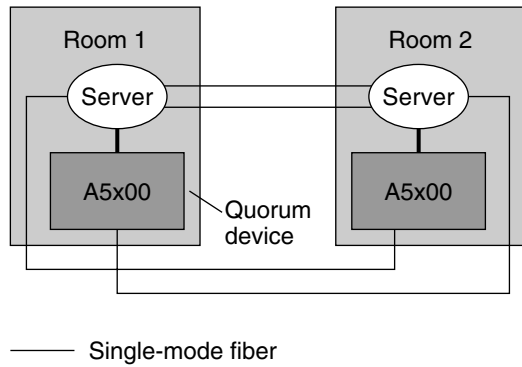


FIGURE 7-9 SPARC: Two-Room Configuration (Sun StorEdge A5x00s)





# Verifying Sun Cluster Hardware Redundancy

---

This chapter describes the tests for verifying and validating the high availability (HA) of your Sun Cluster configuration. The tests in this chapter assume that you installed Sun Cluster hardware, the Solaris Operating System, and Sun Cluster software. All nodes should be booted as cluster members.

This chapter contains the following procedures.

- “How to Test Nodes Using a Power-Off Method” on page 90
- “How to Test Cluster Interconnects” on page 91
- “How to Test IP Network Multipathing Groups” on page 92

If your cluster passes these tests, your hardware has adequate redundancy. This redundancy means that your nodes, cluster transport cables, and Network Adapter Failover (NAFO) groups are not single points of failure.

To perform the tests in “How to Test Nodes Using a Power-Off Method” on page 90 and “How to Test Cluster Interconnects” on page 91, you must first identify the device groups that each node masters. Perform these tests on all cluster pairs that share a disk device group. Each pair has a primary and a secondary for a particular device group. Use the `scstat(1M)` command to determine the initial primary and secondary.

For conceptual information on primary, secondary, failover, device groups, or cluster hardware, see your Sun Cluster concepts documentation.

## Testing Node Redundancy

This section provides the procedure about how to test node redundancy and high availability of device groups. Perform the following procedure to confirm that the secondary takes over the device group that is mastered by the primary when the primary fails.

## ▼ How to Test Nodes Using a Power-Off Method

**Before You Begin** To perform these tests, you must first identify the device groups that each node masters. Perform these tests on all cluster pairs that share a disk device group. Each pair has a primary and a secondary for a particular device group. Use the `scstat(1M)` command to determine the initial primary and secondary.

**1 Power off the primary node.**

Cluster interconnect error messages appear on the consoles of the existing nodes.

**2 On another node, verify that the secondary took ownership of the device group that is mastered by the primary.**

Look for the output that shows the device group ownership.

```
# scstat
```

**3 Power on the initial primary. Boot the node into cluster mode.**

Wait for the system to boot. The system automatically starts the membership monitor software. The node then rejoins the Sun Cluster configuration.

**4 Do you have the device group failback option enabled?**

Use the `scconf -p` command to determine if your device group has the device group failback option enabled.

- **If yes, skip to Step 6.**

The system boot process moves ownership of the device group back to the initial primary.

- **If no, proceed to Step 5.**

**5 From the initial primary, move ownership of the device group back to the initial primary.**

```
# scswitch -S -h nodename
```

**6 Verify that the initial primary has ownership of the device group.**

Look for the output that shows the device group ownership.

```
# scstat
```

## Testing Cluster Interconnect and IP Network Multipathing Group Redundancy

This section provides the procedure about how to test cluster interconnect and IP Network Multipathing group redundancy.

## ▼ How to Test Cluster Interconnects

**Before You Begin** To perform these tests, you must first identify the device groups that each node masters. Perform these tests on all cluster pairs that share a disk device group. Each pair has a primary and a secondary for a particular device group. Use the `scstat(1M)` command to determine the initial primary and secondary.

**1 Disconnect one of the cluster transport cables from a primary node that masters a device group.**

Messages appear on the consoles of each node, and error messages appear in the `/var/adm/messages` file. If you run the `scstat(1M)` command, the Sun Cluster software assigns a `faulted` status to the cluster transport path that you disconnected. This fault does not result in a failover.

**2 Disconnect the remaining cluster transport cable from the primary node that you identified in Step 1.**

Messages appear on the consoles of each node, and error messages appear in the `/var/adm/messages` file. If you run the `scstat` command, the Sun Cluster software assigns a `faulted` status to the cluster transport path that you disconnected. This action causes the primary node to shutdown. This shutdown results in a partitioned cluster.

For conceptual information on failure fencing or *split brain*, see your Sun Cluster concepts documentation.

**3 On another node, verify that the secondary node took ownership of the device group that was mastered by the primary.**

```
# scstat
```

**4 Reconnect all cluster transport cables.**

**5 Boot the initial primary, which you identified in Step 1, into cluster mode.**

For the procedure about how to boot a node, see your Sun Cluster system administration documentation.

**6 Do you have the device group failback option enabled?**

Use the `scconf -p` command to determine if your device group has the device group failback option enabled.

- **If yes, skip to Step 8.**

The system boot process returns ownership of the resource groups and device groups to the initial primary.

- **If no, proceed to Step 7.**

**7 Move all resource groups and device groups off the current primary.**

```
# scswitch -S -h from-node
```

- 8 **Verify that the Sun Cluster software assigned a path online status to each cluster transport path that you reconnected in Step 4.**

```
# scstat
```

## ▼ How to Test IP Network Multipathing Groups

If you perform this test, you can verify that IP addresses failover from one adapter to another adapter within the same IP Network Multipathing group.

- 1 **Verify that all network adapters are active.**

```
# scstat -i
```

- 2 **Disconnect one public network cable from an active network adapter.**

- 3 **Verify that the network adapter status displays as Failed.**

```
# scstat -i
```

- 4 **Ensure the IP address failed over to another adapter.**

```
# ifconfig -a
```

- 5 **Reconnect the public network cable to the network adapter.**

- If `failback` is set to `yes` in the `/etc/default/mpathd` file, the IP address automatically returns to the original network adapter.
- If `failback` is not set to `yes`, the IP address remains with the new adapter until you perform a manual switchover.

For more procedures about how to move an IP address to another adapter, see your Sun Cluster system administration documentation.

- 6 **Repeat Step 1 to Step 5 for each active adapter.**

# Recabling and Replacing Disk Devices

---

This chapter contains the procedures for recabling and replacing disk devices. This chapter provides the following procedures.

- “How to Replace a Failed Boot Disk in a Running Cluster” on page 93
- “How to Move a Disk Cable to a New Host Adapter” on page 95
- “How to Move a Disk Cable From One Node to Another Node” on page 96
- “How to Update Sun Cluster Software to Reflect Proper Device Configuration” on page 97

## Replacing a Failed Boot Disk

This section contains the procedure about how to replace a failed boot disk.

### ▼ How to Replace a Failed Boot Disk in a Running Cluster

Use this procedure to replace a failed boot disk that is mirrored and that is on a node in a running cluster. Use this procedure for both Solstice DiskSuite/Solaris Volume Manager and VERITAS Volume Manager. This procedure also assumes that one mirror is available. This procedure defines Node N as the node on which you are replacing the failed boot disk.

If you do not have a mirror that is available, see your Sun Cluster system administration documentation to restore data on the boot disk.

- 1 Is Node N up and running?
  - If no, skip to [Step 5](#)
  - If yes, proceed to [Step 2](#)
- 2 Is your disk drive hot-pluggable?
  - If yes, skip to [Step 5](#)

- **If no, proceed to [Step 3](#)**
- 3 Determine the resource groups and device groups running on Node N.**  
Record this information because you use this information in [Step 6](#) of this procedure to return resource groups and device groups to these nodes.  

```
# scstat
```

  
For more information, see your Sun Cluster system administration documentation.
  - 4 Move all resource groups and device groups off Node N.**  

```
# scswitch -S -h from-node
```

  
For more information, see your Sun Cluster system administration documentation.
  - 5 Replace the failed boot disk by using the procedure that is outlined in your volume manager documentation.**  
For more information, see your Solstice DiskSuite/Solaris Volume Manager or VERITAS Volume Manager documentation. For the procedure about how to replace a boot disk, see “Recovering From Boot Problems” in *Solaris Volume Manager Administration Guide*.
  - 6 If you moved all resource groups and device groups off Node N in [Step 4](#), return the resource groups and device groups that you identified in [Step 3](#) to Node N.**  

```
# scswitch -z -g resource-group -h nodename
```

```
# scswitch -z -D device-group-name -h nodename
```

  
For more information, see your Sun Cluster system administration documentation.

## Moving a Disk Cable

You can move a disk cable to a different host adapter on the same bus because of a failed host adapter. However, the best solution is to replace the failed host adapter rather than recable the disk cable to a different host adapter. You might want to move a disk cable to a different host adapter on the same bus to improve performance.

This section provides the following two procedures for moving a disk cable.

- “[How to Move a Disk Cable to a New Host Adapter](#)” on page 95
- “[How to Move a Disk Cable From One Node to Another Node](#)” on page 96

Use one of these two procedures to prevent interference with normal operation of your cluster when you want to move a disk cable to a different host adapter on the same bus. If you do not follow these procedures correctly, you might see an error the next time you run the `scdidadm -r` command or the `scgdevs` command. If you see an error message that says `did reconfiguration discovered invalid diskpath`, go to “[How to Update Sun Cluster Software to Reflect Proper Device Configuration](#)” on page 97.

## ▼ How to Move a Disk Cable to a New Host Adapter

Use this procedure to move a disk cable to a new host adapter within a node.



**Caution** – Failure to follow this cabling procedure might introduce invalid device IDs and render the devices inaccessible.

### 1 Stop all I/O to the affected disk(s).

For more information, see your Solstice DiskSuite/Solaris Volume Manager or VERITAS Volume Manager documentation.

### 2 Unplug the cable from the old host adapter.

### 3 From the local node, unconfigure all drives that are affected by the recabling.

```
# cfgadm
```

Or reboot the local node.

```
# reboot -- -r
```

### 4 From the local node, update the Solaris device link.

```
# devfsadm -C
```

### 5 From the local node, update the device ID path.

```
# sddidadm -C
```

### 6 Connect the cable to the new host adapter.

### 7 From the local node, configure the drives in the new location.

```
# cfgadm
```

Or reboot the local node.

```
# reboot -- -r
```

### 8 Add the new device ID path.

```
# scgdevs
```

#### Troubleshooting

If you did not follow this procedure correctly, you might see an error the next time you run the `sddidadm -r` command or the `scgdevs` command. If you see an error message that says `did reconfiguration discovered invalid diskpath`, go to [“How to Update Sun Cluster Software to Reflect Proper Device Configuration”](#) on page 97.

## ▼ How to Move a Disk Cable From One Node to Another Node

Use this procedure to move a disk cable from one node to another node.




---

**Caution** – Failure to follow this cabling procedure might introduce invalid device IDs and render the devices inaccessible.

---

- 1 Delete all references to the device ID path that you are removing from all volume manager and data service configurations.**

For more information, see your Sun Cluster data services collection and your Solstice DiskSuite/Solaris Volume Manager or VERITAS Volume Manager documentation.

- 2 Stop all I/O to the affected disk(s).**

For more information, see your Solstice DiskSuite/Solaris Volume Manager or VERITAS Volume Manager documentation.

- 3 Unplug the cable from the old node.**

- 4 From the old node, unconfigure all drives that are affected by the recabling.**

```
# cfgadm
```

Or reboot the old node.

```
# reboot -- -r
```

- 5 From the old node, update the Solaris device link.**

```
# devfsadm -C
```

- 6 From the old node, update the device ID path.**

```
# sctdidadm -C
```

- 7 Connect the cable to the new node.**

- 8 From the new node, configure the drives in the new location.**

```
# cfgadm
```

Or reboot the new node.

```
# reboot -- -r
```

- 9 From the new node, create the new Solaris device links.**

```
# devfsadm
```



**10 From the new node, add the new device ID path.**

```
# scgdevs
```

**11 Add the device ID path on the new node to any volume manager and data service configurations that are required.**

When you configure data services, check that your node failover preferences are set to reflect the new configuration.

For more information, see your Sun Cluster data services collection and your Solstice DiskSuite/Solaris Volume Manager or VERITAS Volume Manager documentation.

**Troubleshooting**

If you did not follow this procedure correctly, you might see an error the next time you run the `scdidadm -r` command or the `scgdevs` command. If you see an error message that says `did reconfiguration discovered invalid diskpath`, go to [“How to Update Sun Cluster Software to Reflect Proper Device Configuration”](#) on page 97.

## ▼ How to Update Sun Cluster Software to Reflect Proper Device Configuration

If you see the following error when you run the `scdidadm -r` command or the `scgdevs` command, the Sun Cluster software does not reflect the proper device configuration because of improper device recabling.

```
did reconfiguration discovered invalid diskpath.
This path must be removed before a new path
can be added. Please run did cleanup (-C)
then re-run did reconfiguration (-r).
```

Use this procedure to ensure that the Sun Cluster software becomes aware of the new configuration and to guarantee device availability.

**1 Ensure that your cluster meets the following conditions.**

- The cable configuration is correct.
- The cable that you are removing is detached from the old node.
- The old node is removed from any volume manager or data service configurations that are required.

For more information, see your Sun Cluster data services collection and your Solstice DiskSuite/Solaris Volume Manager or VERITAS Volume Manager documentation.

**2 From all nodes, one node at a time, unconfigure all drives.**

```
# cfgadm
```

Or reboot all nodes, one node at a time.

```
# reboot -- -r
```

**3 From all nodes, one node at a time, update the Solaris device link.**

```
# devfsadm -C
```

**4 From all nodes, one node at a time, update the device ID path.**

```
# scdidadm -C
```

**5 From all nodes, one node at a time, configure all drives.**

```
# cfgadm
```

Or reboot all nodes, one node at a time.

```
# reboot -- -r
```

**6 From any node, add the new device ID path.**

```
# scgdevs
```

**7 From all nodes that are affected by the recabling, verify that SCSI reservations are in the correct state.**

```
# scdidadm -R device
```

## Data Replication Approaches

---

This appendix describes available data replication approaches with Sun Cluster. Host-based data replication is important in all clusters. Storage-based data replication, using special software, provides additional data redundancy for any type of cluster.

This appendix includes the following sections:

- “Using Host-Based Data Replication” on page 99
- “SPARC: Using Storage-Based Data Replication” on page 101

---

**Note** – The examples in this appendix illustrate general campus cluster configurations and are not intended to indicate required or recommended setups. For simplicity, the diagrams and explanations concentrate only on features unique to understanding campus clustering. For example, public-network Ethernet connections are not shown.

---

### Using Host-Based Data Replication

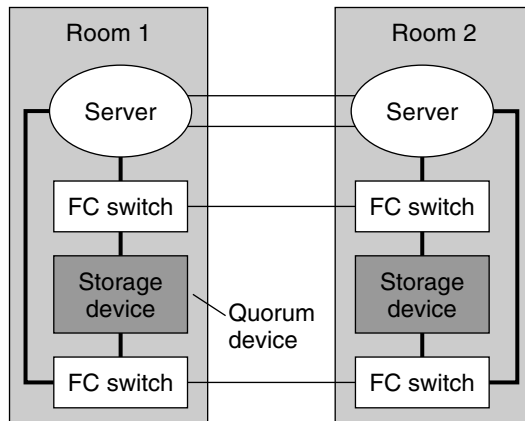
A two-room configuration with host-based data replication is defined as follows:

- Two separate rooms
- Both rooms with one node each and disk subsystems
- Data replicated across disk subsystems in these rooms
- At least one disk subsystem, attached to both hosts, used as a quorum device, located in one of the rooms

**If the room containing the quorum disk is lost, the system cannot recover automatically.** Recovery requires intervention from your Sun service provider.

Figure A-1 shows a sample two-room configuration with storage arrays in each room.





— Single-mode fiber

— Multimode fiber

FIGURE A-1 Example Two-Room Campus Cluster With Host-Based Data Replication (No Multipathing)

Figure 7-7 is similar to a standard noncampus configuration. The most obvious difference is that FC switches have been added to switch from multimode to single-mode fibers.

## SPARC: Using Storage-Based Data Replication

Figure A-2 shows a sample two-room configuration where data is replicated between two storage arrays. In this configuration, the primary storage array is contained in the first room, where it provides data to the nodes in both rooms. The primary storage array also provides the secondary storage array with replicated data. During normal cluster operation, the secondary storage array is not visible to the cluster. However, if the primary storage array becomes unavailable, the secondary storage array can be manually configured into the cluster by a Sun service provider.

---

**Note** – As shown in Figure A-2, the quorum device is on an unreplicated volume.

---

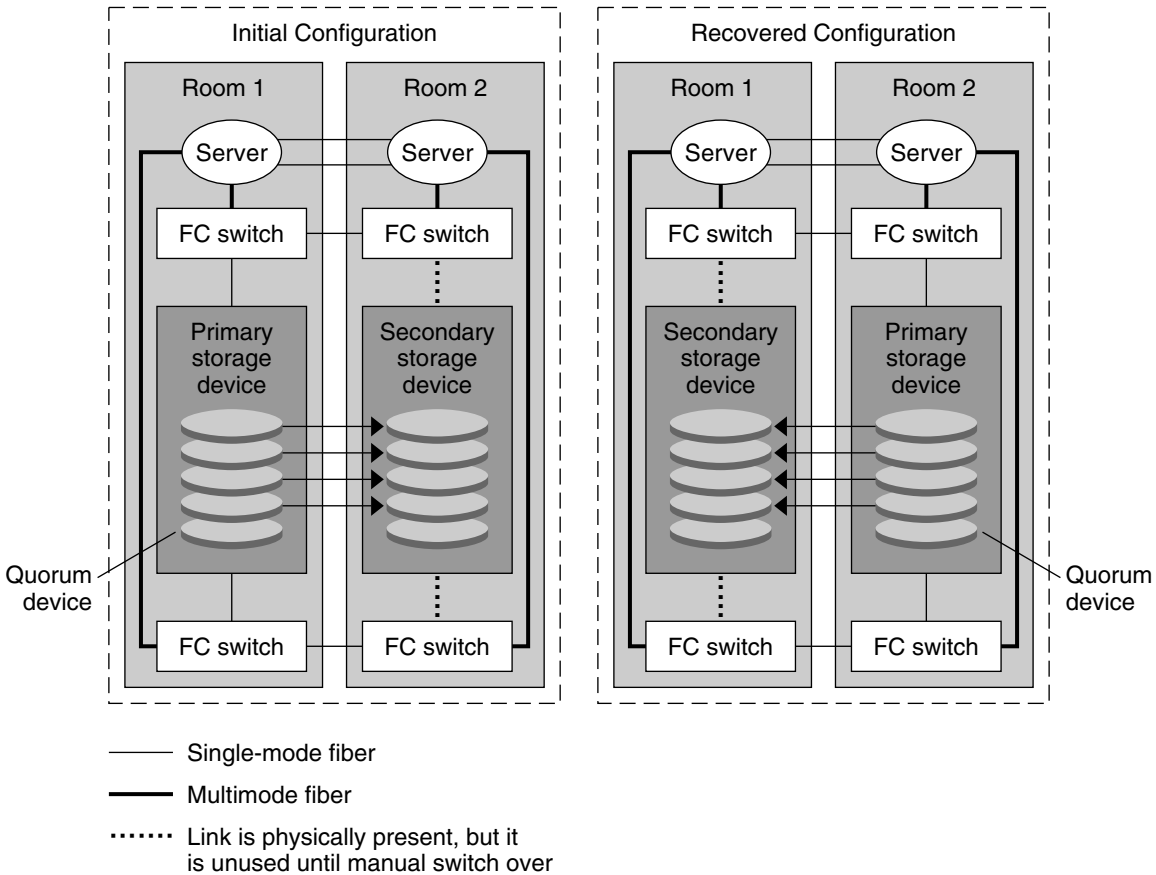


FIGURE A-2 SPARC: Example Two-Room Configuration With Storage-Based Data Replication

Storage-based data replication can be performed synchronously or asynchronously in the Sun Cluster environment, depending on the type of application that is used.

To ensure data integrity, use multipathing and the proper in-box RAID. The following list includes considerations for implementing a storage-based data replication, campus cluster configuration. These considerations are not specific to Sun Cluster.

- Node-to-node distance is limited by Sun Cluster's fibre channel and interconnect infrastructure. Contact your Sun service provider for more information about current limitations and supported technologies.
- Do not configure a replicated volume as a quorum device. Locate any quorum devices on an unreplicated volume.
- Ensure that only the primary copy of the data is visible to cluster nodes. Otherwise, the volume manager might try to access both primary and secondary copies of the data, which could result in data corruption because the secondary copy is read-only.

Refer to the documentation that came with your storage array for information about controlling the visibility of your data copies.

- Particular application-specific data might not be suitable for asynchronous data replication. Use your understanding of your application's behavior to determine how best to replicate application-specific data across the storage devices.

## **SPARC: Recovery Concerns When Using Storage-Based Data Replication**

As with all campus clusters, those that use storage-based data replication generally do not need intervention when they experience a single failure. However, if you lose the room that holds your primary storage device (as shown in [Figure A-2](#)), problems arise in a 2-node cluster. The remaining node cannot reserve the quorum device and cannot boot as a cluster member. In this situation, your cluster requires the following manual intervention:

1. Your Sun service provider must reconfigure the remaining node to boot as a cluster member.
2. You (or your Sun service provider) must configure an unreplicated volume of your secondary storage device as a quorum device.
3. You (or your Sun service provider) must reconfigure the remaining node to use the secondary storage device as primary storage. This reconfiguration might involve rebuilding volume manager volumes, restoring data, or changing application associations with storage volumes.





# Index

---

## A

- accessing node console, 33
- adapters
  - See* public network adapters
  - See* transport adapters
  - Sun Gigabit Ethernet, 59-60
- adding
  - public network adapters, 58
  - transport adapters, 50-52
  - transport cables, 50-52
  - transport junctions, 50-52

## B

- boot disks, replacing, 93-98

## C

- cables, *See* transport cables
- campus clustering, 71-87
  - and fibre channel, 81-82
  - configuration examples, 74-77, 99-101
  - data replication requirements, 72
  - differences from noncampus, 71-87
  - extended examples, 82-87
  - guidelines, 73-80
  - hardware installation and configuration, 81-82
  - interconnect technologies, 80
  - link technologies, 80-81
  - network requirements, 72
  - node compared to room, 73
  - private network technologies, 80

- public network technologies, 80-81
- quorum guidelines, 77-80
- quorum requirements, 72
- recovery with storage-based data replication, 103
- requirements, 71-73
- room, definition of, 73
- storage-based data replication, 101-103
- three-room examples, 82-85
- two-room, example without multipathing, 86
- two-room, multipathing example, 85
- VLANs, 48
- volume management, 73
- CD-ROMs, 19
- ce* Sun Ethernet driver, 60
- configuration examples (campus clustering)
  - three-room, 74-77
  - two-room, host-based data replication, 99-101
  - two-room, storage-based data replication, 101-103
- configuring, terminal concentrator, 26-28
- connecting, terminal concentrator, 25

## D

- data replication
  - host-based, 99-101
  - requirements, 72
  - storage-based, 101-103
- disk cables, moving
  - between host adapters, 95
  - between nodes, 96-97
- disk devices, recabling, 93-98
- DR, *See* dynamic reconfiguration; replacing disks

**drivers**

- ce Sun Ethernet, 60
  - Gigaswift Ethernet, 61
- dual-port HBAs, 66-69**
- Solaris Volume Manager, 68-69
  - Solaris Volume Manager for Sun Cluster, 69
  - supported configurations, 68
- dynamic reconfiguration, 18**

**E**

- /etc/path\_to\_inst file, 59-60, 61
- Ethernet switches, in the interconnect, 38
- Ethernet transport cables and junctions, 40-41
- example configurations (campus clustering)
  - three-room, 74-77
  - two-room, host-based replication, 99-101
  - two-room, storage-based replication, 101-103

**F**

- fibre channel and campus clustering, 81-82
- files**
- ce.conf file, 61
  - /etc/path\_to\_inst file, 59-60, 61
  - /kernel/drv/ge.conf file, 59-60

**G**

- GigaSwift Ethernet driver, jumbo frames, 61

**H**

- hardware installation**
- for campus clusters, 81-82
  - overall installation process, 16-17
  - overview, 15
- hardware RAID, 63-66**
- hardware redundancy, verifying, 89-92**
- help, 11**
- high availability**
- testing, 89-92

- testing device group availability, 89-90
- testing interconnect availability, 91-92
- testing IP multipathing availability, 92

**host adapters**

- dual-port configurations, 66-69
  - moving disk cables, 95
- host-based data replication, 99-101**

**I**

- InfiniBand transport cables and junctions, 40-41**
- installing**
- cluster hardware, 15
  - Ethernet transport cables, 40-41
  - Ethernet transport junctions, 40-41
  - InfiniBand transport cables, 40-41
  - PCI-SCI switches, 41-46
  - PCI-SCI transport cables, 41-46
  - public network hardware, 58
  - Solaris and cluster software, 17
  - Sun Fire Link switches, 46-47
  - Sun Fire Link transport cables, 46-47
  - terminal concentrator, 22
- integrated mirroring, 63-66**
- interconnect**
- configuration for campus clustering, 81-82
  - jumbo frames requirements, 38
  - SCI card placement, 38
  - speed requirements, 37-38
  - technologies for campus clustering, 80
  - testing redundancy, 91-92
- internal hardware disk mirroring, 63-66**
- IP multipathing, testing redundancy, 92**

**J**

- joining multicasts failed, 41**
- jumbo frames**
- and cluster interconnect requirements, 38
  - and public network requirements, 57
  - ce.conf file, 61
  - /etc/path\_to\_inst file, 61

**K**

/kernel/drv/ge.conf file, 59-60

**L**

link technologies, campus clustering, 80-81

local disks, 19

Local Mirroring, *See* storage-based replication

local mirroring, *See* storage-based replication

**M**

mirroring internal disks, 63-66

moving disk cables

    between host adapters, 95

    between nodes, 96-97

multihost disks, 19

multipathing

    example three-room campus cluster, 82-85

    example two-room campus cluster, 85

**N**

NAFO groups

    adding adapters, 59

    redundancy testing, 92

Network Adapter Failover groups, *See* NAFO groups

network requirements for campus clusters, 72

node access, through terminal concentrator, 33

node redundancy, testing, 89-90

nodes, moving disk cables, 96-97

**O**

Oracle Parallel Server, 7

Oracle Real Application Clusters, 72

**P**

PCI-SCI

    error messages, 41

    installing transport cables, 41-46

    interconnect hardware, 41-46

    transport junctions PCI-SCI, transport junctions  
    (Continued)

        installing, 41-46

        troubleshooting interconnect, 45

port configuration, correcting access error, 29

powering off, 18

powering on, 18

private network technologies, campus clustering, 80

Project Liberty, 66-69

public network

    hardware installation, 58

    jumbo frames requirements, 57

public network adapters

    adding, 58

    removing, 59

    replacing, 59

public network technologies, campus clustering, 80-81

**Q**

quorum devices

    and storage-based replication, 103

    campus cluster guidelines, 77-80

    campus cluster requirements, 72

**R**

raidctl command, 63-66

recabling

    boot disks, 93-98

    disk devices, 93-98

recovery, clusters with storage-based data replication, 103

redundancy

    testing interconnect redundancy, 91-92

    testing IP multipathing redundancy, 92

    testing node redundancy, 89-90

Remote Mirroring, *See* storage-based replication

remote replication, *See* storage-based replication

removable media, 19

removing

    public network adapters, 59

    transport adapters, 54-55

- transport cables, 54-55
- transport junctions, 54-55
- replacing
  - boot disks, 93-94
  - public network adapters, 59
  - transport adapters, 52-53
  - transport cables, 52-53
  - transport junctions, 52-53
- replication, storage-based, 101-103
- requirements, interconnect speed, 37-38
- room, definition of (campus clustering), 73
- room compared to node (campus clustering), 73

**S**

- SAN
  - general cluster requirements, 20
  - requirements in campus clusters, 72
- SCI cards, restriction on placement, 38
- setting port parameters, terminal concentrator, 28
- shutdown protocol
  - clustered environment, 18
  - nonclustered environment, 18
- SnapShot, *See* storage-based replication
- snapshots, *See* storage-based replication
- software installation, 17
- Solaris Volume Manager, dual-port HBAs, 68-69
- Solaris Volume Manager for Sun Cluster, dual-port HBAs, 69
- standard clusters, differences from campus clusters, 71-87
- storage-based replication, 101-103
  - and quorum devices, 103
  - limitations, 102
  - recovery, 103
- Sun Cluster software, updating, 97-98
- Sun Fire Link
  - installing transport cables, 46-47
  - installing transport junctions, 46-47
- Sun Gigabit Ethernet Adapter, 59-60
- switches
  - See* transport junctions
  - See also* transport junctions

**T**

- tapes, 19
- technical support, 11
- terminal concentrator
  - accessing node console, 33
  - configuring, 26-28
  - connecting, 25
  - correcting port configuration, 29
  - disconnect users, 34-35
  - establishing default route, 31-32
  - installing, 22
  - reset port, 34-35
  - setting port parameters, 28
- testing
  - high availability, 89-92
  - interconnect availability, 91-92
  - interconnect redundancy, 91-92
  - IP multipathing availability, 92
  - IP multipathing redundancy, 92
  - NAFO group redundancy, 92
  - node availability, 89-90
  - node redundancy, 89-90
- training, 13
- transport adapter firmware, upgrading, 55-56
- transport adapters
  - adding, 50-52
  - removing, 54-55
  - replacing, 52-53
- transport cables
  - adding, 50-52
  - Ethernet, installing, 40-41
  - PCI-SCI, installing, 41-46
  - removing, 54-55
  - replacing, 52-53
  - Sun Fire Link, installing, 46-47
- transport junctions
  - adding, 50-52
  - Ethernet, installing, 40-41
  - PCI-SCI, installing, 41-46
  - removing, 54-55
  - replacing, 52-53
  - Sun Fire Link, installing, 46-47
- troubleshooting, PCI-SCI interconnect, 45
- TrueCopy, *See* storage-based replication

**U**

updating, Sun Cluster software, 97-98  
upgrading, transport adapter firmware, 55-56

**V**

verifying, hardware redundancy, 89-92  
virtual local area networks, *See* VLANs  
VLANs  
    campus clusters, 48  
    guidelines, 47-48  
volume management with campus clustering, 73  
volumem, *See* storage-based replication

