EMC Host Connectivity Guide for Sun Solaris

P/N 300-000-607
REV A11

EMC Corporation
Corporate Headquarters:
Hopkinton, MA 01748-9103
1-508-435-1000
www.EMC.com

Copyright © 2007 EMC Corporation. All rights reserved.

Printed April, 2007

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

All other trademarks used herein are the property of their respective owners.

Contents

Chapter 2 Solaris SPARC and Symmetrix over Fibre Channel

Chapter 3 Solaris SPARC and Symmetrix over iSCSI

Chapter 4 Solaris SPARC and Symmetrix over SCSI

Chapter 5 Sun Cluster 2.x and High-Availability Environment

Chapter 6 **Sun Cluster 3.x**

Chapter 7 Solaris SPARC and CLARiiON

Chapter 8 VCS Cluster with CLARiiON

Chapter 9 Reassigning LUN Ownership with CLARiiON

PART 2 Solaris x86 Symmetrix/CLARiiON Connectivity

Chapter 10 Sun Cluster 3.x for x86

Chapter 11 Solaris x86 Symmetrix/CLARiiON over Fibre Channel

Chapter 12 Solaris x86 and Symmetrix over iSCSI

Figures

Tables

Preface

As part of an effort to improve and enhance the performance and capabilities of its product line, EMC from time to time releases revisions of its hardware and software. Therefore, some functions described in this document may not be supported by all revisions of the software or hardware currently in use. For the most up-to-date information on product features, refer to your product release notes.

If a product does not function properly or does not function as described in this document, please contact your EMC representative.

This guide describes the features and setup procedures for Sun Solaris host interfaces to EMC Symmetrix and CLARiiON systems over Fibre Channel and (Symmetrix only) SCSI.

Audience

This guide is intended for use by storage administrators, system programmers, or operators who are involved in acquiring, managing, or operating Symmetrix, CLARiiON, and host devices.

Readers of this guide are expected to be familiar with the following topics:

- Symmetrix or CLARiiON operation
- Sun Solaris operating environment
- For the most up-to-date support information, refer to the *EMC Support Matrix*, available through E-Lab Interoperability Navigator or Powerlink, at:

  http://elabnavigator.EMC.com
  http://Powerlink.EMC.com

Conventions used in this guide

EMC uses the following conventions for notes and cautions.

Note: A note presents information that is important, but not hazard-related.

CAUTION

A caution contains information essential to avoid damage to the system or equipment. The caution may apply to hardware or software.

Typographical conventions

EMC uses the following type style conventions in this guide:

Normal

Used in running (nonprocedural) text for:

- Names of interface elements (such as names of windows, dialog boxes, buttons, fields, and menus)
- Names of resources, attributes, pools, Boolean expressions, buttons, DQL statements, keywords, clauses, environment variables, filenames, functions, utilities
- URLs, pathnames, filenames, directory names, computer names, links, groups, service keys, file systems, notifications

Bold:

Used in running (nonprocedural) text for:

- Names of commands, daemons, options, programs, processes, services, applications, utilities, kernels, notifications, system call, man pages

Used in procedures for:

- Names of interface elements (such as names of windows, dialog boxes, buttons, fields, and menus)
- What user specifically selects, clicks, presses, or types

Italic:

Used in all text (including procedures) for:

- Full titles of publications referenced in text
- Emphasis (for example a new term)
- Variables

`Courier:`

Used for:

- System output, such as an error message or script
- URLs, complete paths, filenames, prompts, and syntax when shown outside of running text

`Courier bold:`

Used for:

- Specific user input (such as commands)

`Courier italic:`

Used in procedures for:

- Variables on command line
- User input variables

< >	Angle brackets enclose parameter or variable values supplied by the user
[ ]	Square brackets enclose optional values
|	Vertical bar indicates alternate selections - the bar means “or”
{ }	Braces indicate content that you must specify (that is, x or y or z)
...	Ellipses indicate nonessential information omitted from the example

Where to get help

EMC support, product, and licensing information can be obtained as follows.

Product information — For documentation, release notes, software updates, or for information about EMC products, licensing, and service, go to the EMC Powerlink website (registration required) at:

`http://Powerlink.EMC.com`

Technical support — For technical support, go to EMC Customer Service on Powerlink. To open a service request through Powerlink, you must have a valid support agreement. Please contact your EMC sales representative for details about obtaining a valid support agreement or to answer any questions about your account.

Your comments

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications. Please send your opinion of this guide to:

`techpub_comments@EMC.com`

PART 1

Solaris SPARC Symmetrix/CLARiiON Connectivity

Part 1 includes information specific to the Sun Solaris SPARC environment:

1

Solaris SPARC and Symmetrix Environment

This chapter provides information specific to Sun Solaris SPARC hosts connecting to Symmetrix systems.

Solaris SPARC and Symmetrix environment

This section lists some Symmetrix® support information specific to the Solaris environment.

Also refer to the appropriate chapter(s) in this list:

- Chapter 3, ”Solaris SPARC and Symmetrix over iSCSI”
- Chapter 4, ”Solaris SPARC and Symmetrix over SCSI”
- Chapter 5, ”Sun Cluster 2.x and High-Availability Environment”

Hardware connectivity

- Refer to the *EMC Support Matrix* or contact your EMC representative for the latest information on qualified hosts, host bus adapters, and connectivity equipment.

Solaris operating system

Refer to the *EMC Support Matrix* for required Solaris operating system versions.

Boot device support

Booting from the Symmetrix is available to Solaris hosts as described under "Boot Device Support" in the *EMC Support Matrix*.

Symmetrix configuration

The Symmetrix system is configured by an EMC Customer Engineer through the Symmetrix service processor.

Note: Refer to "Fibre Bit Settings" in the the *EMC Support Matrix* for recommended director bit settings.

Useful Solaris utilities and functions

This section lists Solaris functions and utilities you can use to define and manage Symmetrix devices. The use of these functions and utilities is optional. They are listed for reference only:

- **`format`** — The Solaris disk format utility. Allows you to format, partition, and label disk drives.
- **`newfs`** — Creates a file system.
- **`shutdown`** — Gracefully shuts down the system.

Note: **`shutdown`** is the preferred command for system shutdown. The **`halt`** command is not recommended.

System and error messages

Solaris logs system and error messages to a file called `/var/adm/messages` and also displays these messages at the system console.

Solstice DiskSuite

Solstice DiskSuite is a tool for disk and file management. This tool can be used to create and manage logical disks, mirrored or striped volumes, and file systems. DiskSuite supports large file systems, file system expansion and volume manager level intent logging for fast file system recovery.

Refer to the following documents for instructions on installing the Solstice DiskSuite software, creating metadevices, creating the diskset, and other related operations:

- *Solstice DiskSuite Installation and Product Notes*
- *Solstice DiskSuite User's Guide*
- *Solstice DiskSuite Reference Guide*

DiskSuite state database replicas

The *state database* stores all configuration and status information for DiskSuite objects. Without the state database DiskSuite is unable to access any devices and all data could be lost. Replicas of the database are created so that DiskSuite can compare copies to verify the current configuration and running state of all objects.

It is recommended that at least three replicas be created. If one replica is not available, the remaining two can still be compared to verify configuration and state information. Three replicas can be stored on a system boot disk, however, this creates a single point of failure at the boot disk. Additional replicas should be created on other system disks including Symmetrix devices to protect against the loss of the boot disk.

The following considerations apply when planning locations for state database replicas.

- Always create at least three replicas. DiskSuite will not function if it has only one state database.
- Replicas can be stored on any unused partition or on partitions that are also part of a metadevice or logging device with the exceptions of `root`, `swap`, `/usr` or file system.
- Replicas should be spread evenly across host controllers. In Symmetrix, replica devices should be behind separate even and odd host bus adapters.
- Store replicas on least two disks on each controller.
- Each disk should contain only one replica.

Sun ZFS (Zettabyte file system)

ZFS file system is a Sun product built into the Solaris 10 Operating System. It presents a pooled storge model that eliminates the concept of volumes as well as all of the related partition management, provisioning, and file system sizing matters. ZFS combines scalability anf flexibility while providing a simple command interface.

For more information on how to operate ZFS functionalities, refer to the Sun's *Solaris ZFS Administration Guide*, available at: http://docs.sun.com/app/docs/doc/819-5461.

CAUTION

EMC supports ZFS in Solaris 10 11/06 or later. The Snapshot and Clone features of ZFS are supported only through Sun Microsystems.)

VERITAS Volume Manager

VERITAS Volume Manager (VxVM) and VERITAS File System (VxFS) are tools for disk and file management. VxVM can be used to create logical disks, mirrored and striped volumes. VxFS supports large file systems, file system expansion and a journaling file system.

Refer to the following documents for instructions on installing VxVM and VxFS, as well as creating disk groups, mirror volumes, striped volumes, and other related operations:

- *VERITAS Volume Manager Installation Guide*
- *VERITAS Volume Manager User's Guide*
- *VERITAS Volume Manager System Administrator's Guide*
- *VERITAS Volume Manager Release Notes*
- *VERITAS File System Installation Guide*
- *VERITAS File System Administrator's Guide*

CAUTION

VERITAS Dynamic Multipathing (DMP) functionality requires enabling the Symmetrix director C-bit flag.

CAUTION

The VxVM 4.0 MP1 and MP2, the VxVM 4.1 MP1 and RP4, and the VxVM 5.0 support the Symmetrix SPC2 flag.

CAUTION

All VxVM revisions or their MPx combinations are not supported by the SPC2 flag flip (from disable to enable, or from enable to disable) except a combination of: VxVM 4.1 + MP1 + RP4.

Creating and mounting a file system

Volumes created and managed by Solstice DiskSuite or VxVM volume manager may be used as raw devices, with the standard UNIX file system (UFS), or with the VxFS journaled file system.

Intent logging

Intent logging records pending changes to the file system structure in an intent log. The intent log is replayed during system failure recovery to complete or abandon changes to the structure that were pending at the time of system failure. The file system can then be mounted without completing a full structural check (fsck). An intent logging system can significantly reduce recovery time following a system failure.

Intent logging may be preformed at the file system level or at the volume manager level. Intent logging at the file system level, generally known as a journaling file system (JFS), is usually more effective than volume manager intent logging.

VxFS and VxVM intent logging

The VERITAS File System (VxFS) includes a journaling file system that provides intent logging at the file system level. Pending changes to the file system, written to an intent log, are scanned during recovery from a system failure. Changes that were active at the time of failure are completed, and the file system is mounted without requiring a fsck of the entire file system. File system recovery is done in a few seconds; much faster than a standard UFS recovery that requires a complete fsck.

VxVM Dirty Region Logging (DRL) provides intent logging at the volume manager level to reduce the time required to resynchronize mirrored volumes after a system failure. DRL is applied to VxVM mirrored volumes only. Striped or concatenated volumes do not use intent logging, but may rely on VxFS for fast recovery.

SDS intent logging

Solstice DiskSuite (SDS) uses a standard UNIX file system (UFS) that does not provide intent logging at the file system level. However, the VERITAS File System (VxFS) can be used in the SDS environment to provide a journaling file system for SDS volumes.

At the volume manger level, SDS uses a method called the *UFS logging feature* to provide intent logging for all volume types (striped, concatenated, mirrored, and RAID). UFS logging is not a journaling file system. It uses the standard UFS and does intent logging at the volume manager level.

The following paragraphs outline the steps required to add, create and mount a standard UNIX file system or a VxFS journaled file system for volumes and raw devices.

CAUTION

You may wish to place the Symmetrix devices in the mount table. This requires editing /etc/vfstab. This file is syntax-sensitive, and if not edited properly can prevent the system from booting.

UFS on raw device

To create standard UNIX file systems under Solaris OS, log in as **root** and proceed as follows for each new device.

Create new file system

Once you have formatted, partitioned, and labeled each Symmetrix disk device, create a new file system for each Symmetrix disk. To do this, use the **newfs** command in a statement similar to the following:

```
newfs -v /dev/rdsk/c1t0d0s0
```

At the `Construct new file system?` prompt, type **y** and press ENTER.

The actions above created a new file system for the Symmetrix disk connected to SCSI controller 1, target ID 0, lun 0, partition 0.

Create mount directory

Once the file systems for each Symmetrix device are created, create a mount directory for each device. To do this, type a statement similar to the following for each Symmetrix device:

```
mkdir /fs/c1t0d0s0
```

where **/fs/c1t0d0s0** is the complete path for the new file system directory.

Note: The `/fs` directory must exist prior to creating the mount directories.

Mount the file system

To mount each file system, type a statement similar to the following:

```
mount /dev/dsk/c1t0d0s0 /fs/c1t0d0s0
```

VxFS on raw device

A VxFS journaling file system is created using the `mkfs` command with arguments provided for block size, log size, device name and size. To create a VxFS file system first determine the size in sectors of the volume. The size of the volume is displayed under the `Sector Count` field of the `prtvtoc` output.

Once you have formatted, partitioned, and labeled each Symmetrix disk device, create a new file system for each Symmetrix disk. To create VxVM journaling file systems under VxFS and Solaris OS, log in as **root** and proceed as follows for each new device.

Create new file system

1. To display sector count information, enter:

```
prtvtoc /dev/rdsk/c1t0d0s0
```

Note: The size of the disk is displayed under the **Sector Count** field. (Assume 4099000 for this example.)

2. To create the VxFS file system for the volume, enter:

```
mkfs -F vxfs -o bsize=4096,logsize=256 /dev/rdsk/c1t0d0s0 4099000
```

where:
bsize = block size in bytes (1k, 2k, 4k, or 8k - 1k default for file systems < 4 GB, 4k default for file systems > 4 GB
logsize = size of VxFS file system logging in blocks (256 blocks default, 32 to 1024 blocks)
4099000 = file system size in sectors (from `prtvtoc` command)

Create mount directory

Once the file systems for each Symmetrix device are created, create a mount directory for each device. To do this, type a statement similar to the following for each Symmetrix device:

```
mkdir /fs/c1t0d0s0
```

where **/fs/c1t0d0s0** is the complete path for the new file system directory.

Mount the file system

To mount each file system, type a statement similar to the following:

```
mount -F vxfs /dev/dsk/c1t0d0s0 /fs/c1t0d0s0
```

UFS on SDS device

To create a standard UNIX file system (UFS) under Solstice DiskSuite, log in as **root** and proceed as follows for each new device.

Create new file system

Once you have formatted, partitioned, and labeled each Symmetrix disk device, and created SDS volumes, create a new file system on each volume. To do this, use the `newfs` command in a statement similar to the following:

```
newfs -v /dev/md/rdsk/d0
```

At the `Construct new filesystem?` prompt, type **y** and press ENTER.

The actions above created a new file system for the Symmetrix disk defined as SDS metadevice `d0`.

Create mount directory

Once the file systems for each device are created, create a mount directory for each device. To do this, type a statement similar to the following for each device:

```
mkdir /fs/d0
```

where **/fs/d0** is the complete path for the new file system directory.

Note: You can assume that the `/fs` directory existed prior to creating the d0 directory.

Mount the file system

To mount each file system, type a statement similar to the following:

```
mount /dev/md/dsk/d0 /fs/d0
```

VxFS on SDS device

A VxFS journaling file system is created using the `mkfs` command with arguments provided for block size, log size, device name and size. To create a VxFS file system first determine the size in sectors of the volume. The size of the volume is displayed under the `Sector Count` field of the `prtvtoc` output.

To create VxFS journaling file systems under Solstice DiskSuite, log in as **root** and proceed as follows for each new device.

Create new file system

1. To display sector count information, enter:

   ```
   prtvtoc /dev/md/rdsk/d0
   ```

Note: The size of the disk is displayed under the Sector Count field. (Assume 4099000 for this example.)

2. To create the VxFS file system for the volume, enter:

```
mkfs -F vxfs -o bsize=1024 logsize=512 /dev/md/rdsk/d0 4099000
```

where:
bsize = block size in bytes (1k, 2k, 4k, or 8k - 1k default for file systems < 4 GB, 4k default for file systems > 4 GB
logsize = size of VxFS file system logging in blocks (256 blocks default, 32 to 1024 blocks)
4099000 = file system size in sectors (from `prtvtoc` command)

Create mount directory

Once the file systems for each device are created, create a mount directory for each device. To do this, type a statement similar to the following for each device:

```
mkdir /fs/d0
```

where **/fs/d0** is the complete path for the new file system directory.

Mount the file system

To mount each file system, type a statement similar to the following:

```
mount -F vxfs /dev/md/dsk/d0 /fs/d0
```

UFS on VxVM device

To create a standard UNIX file system (UFS) under VERITAS Volume Manager, log in as **root** and proceed as follows for each new device.

Create new file system

Once you have formatted, partitioned, and labeled each Symmetrix disk device, and created VxVM volumes, create a new file system on each volume. To do this, use the **newfs** command in a statement similar to the following:

```
newfs -v /dev/vx/rdsk/dskgrp/vol1
```

At the `Construct new filesystem?` prompt, type **y** and press ENTER.

The actions above created a new file system for the Symmetrix disk defined as `vol1` in the diskgroup named `dskgrp`.

Create mount directory

Once the file systems for each device are created, create a mount directory for each device. To do this, type a statement similar to the following for each device:

```
mkdir /fs/vol1
```

where **/fs/vol1** is the complete path for the new file system directory.

Mount the file system

To mount each file system, type a statement similar to the following:

```
mount -F vsfs /dev/vx/dsk/dskgrp/vol1 /fs/vol1
```

VxFS on VxVM device

A VxFS journaling file system is created using the **mkfs** command with arguments provided for block size, log size, device name and size. To create a VxFS file system first determine the size in sectors of the volume. The size of the volume is displayed under the `Sector Count` field of the **prtvtoc** output.

To create VxFS journaling file systems under VERITAS Volume Manager, log in as **root** and proceed as follows for each new device.

Create new file system

1. To display sector count information, enter:

   ```
   prtvtoc /dev/vx/rdsk/dskgrp/vol1
   ```

   Note: The size of the disk is displayed under the **Sector Count** field. (Assume 4099000 for this example.)

2. To create the VxFS file system for the volume, enter:

```
mkfs -F vxfs -o bsize=4096 logsize=512 /dev/vx/rdsk/dskgrp/vol1 4099000
```

where:

bsize = block size in bytes (1k, 2k, 4k, or 8k **-** 1k default for file systems < 4 GB, 4k default for file systems > 4 GB
logsize = size of VxFS file system logging in blocks (256 blocks default, 32 to 1024 blocks)
4099000 = file system size in sectors (from **prtvtoc** command)

Create mount directory

Once the file systems for each device are created, create a mount directory for each device. To do this, type a statement similar to the following for each device:

```
mkdir /fs/vol1
```

where **/fs/vol1** is the complete path for the new file system directory.

Mount the file system

To mount each file system, type a statement similar to the following:

```
mount -F vxfs /dev/vx/dsk/dskgrp/vol1 /fs/vol1
```

File system expansion

DiskSuite volume manager allows you to increase the available storage space of an existing volume by concatenating additional volumes to the metadevice using **metattach**. A file system can then be expanded to fill all or part of the additional space using the **growfs** command. The file system can remain mounted, but will be locked (lockfs) during the expansion. For detailed information on expanding a file system under DiskSuite, refer to the *Solstice DiskSuite User's Guide*.

VxVM allows *growing* of a mounted file system using **vxassist** commands. The following steps outline the procedure:

1. Log in as **root**.
2. To determine how large the volume can grow, enter:

   ```
   vxassist maxgrow vol
   ```

   where **vol** is the volume name.

   The result is similar to the following:

```
Volume vol can be extended by 12533760 to 16629760 (8120Mb)
```

3. To determine the size of the current volume, enter:

   ```
   vxprint -vt
   ```

   The result is similar to the following:

```
Disk group: cust
V  NAME          USETYPE    KSTATE    STATE     LENGTH    READPOL    PREFPLEX
v  vol           fsgen      ENABLED   ACTIVE    4096000   ROUND      -
```

The length unit is in sectors (1 sectors = 512 bytes). Therefore vol is approximately 2 GB.

4. To grow the volume `vol` to 4 GB, enter:

   ```
   vxassist growto cust-mirvol 4g
   vxprint -vt
   ```

   The result is similar to the following:

```
Disk group: cust
V NAME         USETYPE    KSTATE     STATE      LENGTH     READPOL    PREFPLEX
v cust-mirvol  fsgen      ENABLED    ACTIVE     8388608    ROUND      -
```

Obtaining files from the EMC FTP server

The latest device definition files, as well as the Inquiry utility (**inq**) are available on the EMC FTP server. You can access the server through EMC.com or through an FTP software package.

Using EMC.com

You can connect to the EMC home page at www.emc.com. You can also go directly to EMC's anonymous FTP server by entering the URL: **ftp://ftp.emc.com**:

1. Launch your web browser and type ftp://ftp.emc.com at the prompt.
2. Select **pub**, **symm3000**, **solaris**.
3. FTP the desired files to your host. Refer to the appropriate section(s):
   - "Obtaining device definition files" on page 36
   - "Running inquiry" on page 37

Using FTP software

To connect to EMC's anonymous FTP server:

1. At the host, log in as **root** and create the directory /usr/ftp_emc:

   ```
   mkdir /usr/ftp_emc
   ```

2. Change to the /usr/ftp_emc directory:

   ```
   cd /usr/ftp_emc
   ```

3. Connect to EMC's FTP server:

   ```
   ftp ftp.emc.com
   ```

4. At the FTP server login prompt, log in as **anonymous**.
5. At the password prompt, enter your e-mail address.

   You are now connected to the FTP server. To display a listing of FTP commands available to you, type **help** and press ENTER at the prompt.

6. FTP the desired files to your host. Refer to the appropriate section(s):
   - "Obtaining device definition files" on page 36
   - "Running inquiry" on page 37

Obtaining device definition files

If a configuration requires one or more drive definition files, you must obtain these files from EMC and copy them to your host before configuring the Symmetrix system in your host environment.

You or your system administrator can FTP these files to the host from EMC's anonymous FTP server, `ftp.emc.com`. Refer to "Obtaining files from the EMC FTP server" on page 35. Depending on your particular host, you may need to copy one or more files.

Transferring the device definition files

Note: EMC recommends that you copy these files to a `/usr/ftp_emc` directory.

Once you are in the desired directory on the EMC FTP server:

1. Confirm you are in the correct directory and note the names and number of files present:

   ```
   pwd
   ls
   ```

2. Confirm that your host's current directory is `ftp_emc`:

   ```
   lcd /usr/ftp_emc
   ```

3. Disable the interactive mode by typing **`prompt`** and pressing ENTER. This allows you to copy several files without intervention.
4. Copy all files in the directory to your host:

   ```
   mget *
   ```

5. At the prompt, confirm that all files copied to the directory on your host:

   ```
   !ls
   ```

6. Exit the FTP session:

   ```
   quit
   ```

Running inquiry

The Inquiry command (**inq**) displays several fields that can help you determine which Symmetrix volume is associated with a particular device as seen by the host.

You can find an executable copy of the **inq** command on EMC's anonymous FTP server, `ftp.emc.com`, in the `/pub/sym3000/inquiry/latest` directory. (Refer to "Obtaining files from the EMC FTP server" on page 35.)

Example The following figure shows a sample output of **inq** when run from the host console.

```
Inquiry utility, Version 4.91
Copyright (C) by EMC Corporation, all rights reserved.
-----------------------------------------------------------
DEVICE :VEND :PROD :REV :SER NUM :CAP :BLKSZ
----------------------------------------------------------
dev/rdsk/c0t2d0s2 :SEAGATE :ST34371W SUN4.2G:7462 :9719D318 :4192560 :512
dev/rdsk/c0t3d0s2 :SEAGATE :ST34371W SUN4.2G:7462 :9719E906 :4192560 :512
dev/rdsk/c10t0d0s2 :EMC :SYMMETRIX :5264 :14000280 :224576 :512
dev/rdsk/c10t0d1s2 :EMC :SYMMETRIX :5264 :14001280 :224576 :512
dev/rdsk/c10t0d2s2 :EMC :SYMMETRIX :5264 :14002280 :224576 :512
dev/rdsk/c10t0d3s2 :EMC :SYMMETRIX :5264 :14003280 :224576 :512
dev/rdsk/c10t0d4s2 :EMC :SYMMETRIX :5264 :14004280 :224576 :512
dev/rdsk/c10t0d5s2 :EMC :SYMMETRIX :5264 :14005280 :224576 :512
```

The output fields are as follows:

- `DEVICE` = UNIX device name (full pathname) for the SCSI device
- `VEND` = Vendor Information
- `PROD` = Product Name
- `REV` = Revision number — for a Symmetrix, this will be the microcode version
- `SER NUM` = Serial number, in the format `SSVVVDDP`, where:
  - `SS` = last two digits of the Symmetrix serial number
  - `VVV` = Logical Volume number
  - `DD` = Channel Director number
  - `P` = port on the channel director
- `CAP` = Size of the device in kilobytes
- `BLKSZ` = Size in bytes of each block

Symmetrix configuration

In the Solaris host environment, you can configure the Symmetrix disk devices into logical volumes. EMC recommends the following logical volume ratios as indicated in Table 1.

Table 1 Disk device configuration

		Symmetrix logical volume to physical splits											
		1:1	2:1	3:1	4:1	5:1	6:1	7:1	8:1	9:1	10:1	11:1	12:1 ---> 32:1 [a]
Physical devices	4 GB	x	x										
	9 GB	x	x	x	x								
	18 GB		x	x	x	x	x	x	x	x	x	x	x
	23 GB		x [b]	x	x	x	x	x	x	x	x	x	x
	36 GB			x [b]	x	x	x	x	x	x	x	x	x
	47 GB			x [b]	x [b]	x	x	x	x	x	x	x	x
	50 GB [c]				x	x	x	x	x	x	x	x	x
	73 GB [c, d]					x	x	x	x	x	x	x	x
	181 GB												x

a. With microcode revision level 5263, maximum is 8:1. With 5264, maximum is 16:1. With 5265, possible maximum is 32:1. EMC recommends a maximum of 8:1 for best performance.

b. These ratios are available only in microcode revison level 5265 or higher.

c. Supported by microcode revision level 5x66 or higher.

d. Does not support the RAID-S configuration on the Symmetrix system.

The EMC Customer Engineer should contact the EMC Configuration Specialist for updated online information. This information is necessary to configure the Symmetrix system to support the customer’s host environment.

2

Solaris SPARC and Symmetrix over Fibre Channel

This chapter provides information specific to Sun Solaris SPARC hosts connecting to Symmetrix systems over Fibre Channel.

Symmetrix/Solaris SPARC Fibre Channel environment

This section lists some Symmetrix/Fibre Channel support information specific to the Solaris SPARC environment.

Also refer to Chapter 1, "Solaris SPARC and Symmetrix Environment."

Software

The Fibre Channel adapter driver functions as a device driver layer below the standard *sd* or *ssd* Solaris SCSI adapter driver. The Fibre Channel interface is therefore transparent to the Solaris disk administration system.

Addressing

Sun uses SCSI-2 device access protocol in addressing Fibre Channel devices, up to 256 (1 to 255) LUNs per host bus adapter (HBA) port for the *sd* driver and up to 4096 (0 to 4095) LUNs per HBA port for the *ssd* driver.

Understanding persistent binding in a fabric environment

Matching addresses with their associated devices requires that each Fibre Channel director port be *bound* to a target number, regardless of changes in the physical locations of the Fibre Channel fabric. The Symmetrix fabric implementation uses a method called *persistent binding*, which is a map of target, LUN, driver instance, and Symmetrix port. The Fibre Channel HBA stores this information permanently in non-volatile storage.

The Fibre Channel HBA driver also implements the capability to bind the devices by individual LUN (Symmetrix device).

Note: Before implementing persistent binding, be sure you understand the effects.

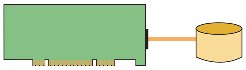
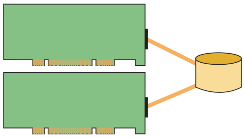
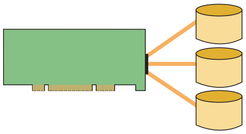
What happens without persistent binding

Without a persistent binding mechanism, the host cannot maintain persistent logical routing of the communication from a device address (`/dev/rdsk/cNtNdNsN`) across the fabric to a Symmetrix volume. If the physical configuration of the switch is changed (for example, the cable is swapped or the host is rebooted), the logical route becomes inconsistent, causing possible data corruption if the user application is modifying data through inconsistent logical routing of the communication from the driver entry point to a volume in a Symmetrix system across the fabric.

Binding models

There are three basic methods of binding implementation:

Table 1 Basic binding implementation models

Model	Configuration	Example
Straight	1 HBA port to 1 Symmetrix port	
Fan-out	*n* HBA ports to 1 Symmetrix port (28 to 1 maximum)	
Fan-in	1 HBA port to *n* Symmetrix ports (1 to 12 maximum)	

Refer to the *EMC Connectrix Enterprise Network System Planning Guide* for more information on persistent binding.

Host configuration with Emulex HBAs

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved HBAs. Sun SPARC-based servers support Emulex 2 GB/4 GB HBAs:

◆ LP9002L-E	(2 GB PCI adapter)
◆ LP9002DC-E	(2 GB PCI adapter)
◆ LP9802-E	(2 GB single port PCI-X adapter)
◆ LP10000-E	(2 GB single port PCI-X adapter)
◆ LP10000DC-E	(2 GB dual port PCI-X adapter)
◆ LP11000-E	(4 GB single port PCI-X adapter)
◆ LP11002-E	(2 GB dual port PCI-X adapter)
◆ LPe11000-E	(4 GB single port PCI Express adapter)
◆ LPe11002-E	(4 GB dual port PCI Express adapter)

There are two HBA drivers that can be used for Emulex HBAs:

- Emulex LightPulse Fibre Channel Adapter driver (lpfc)
  - Supports 2 GB HBAs
- Emulex-Sun LightPulse Fibre Channel Adapter driver (emlxs)
  - Supports 2 GB and 4 GB HBAs

CAUTION

EMC does not support FC-IP on the Emulex adapters.

EMC does not support the coexistence of the lpfc and emlxs drivers on the same host.

lpfc driver

To install one or more EMC-qualified Emulex host bus adapters into a Solaris host and configure the host for connection to the Symmetrix over Fibre Channel, follow the procedures in *EMC Fibre Channel with Emulex Host Bus Adapters in the Solaris Host Environment*.

You can obtain the document from the Emulex website, as follows:

1. Access `http://www.emulex.com`.
2. Click **drivers, software, and manuals** at the left side of the screen.
3. After **Select vendor:**, click **EMC**.
4. Click the link to your HBA model at the left side of the screen.
5. Under **Drivers for Solaris**, find the description of your HBA driver in the **Description** column. Then click the **Installation and Configuration** link in the associated **Online Manuals** column.

emlxs driver

The emlxs driver is a part of the Sun StorEdge SAN Foundation software (Sun SAN). The Sun SAN is embedded in the Solaris 10 Update 1 (01/06). If you intend to use Solaris 10 prior to S10-U1, there are two packages: SUNWemlxs and SUNWemlxu, that are required before installing required patch 120222-xx (refer to the *EMC Support Matrix* for suport revision). These packages are available on the Sun website:

`http::/www.sun.com/download/products.xml?id=42C4317d`

On Solaris 10, the Sun patch 120222-06 is a minimum version that has been qualified for Emulex PCI-X 4 GB adapters and PCI-E 4 GB adapters.

If you intend to use Solaris 8 or Solaris 9, you must follow the *Sun StorEdge SAN Foundation Software Installation Guide* which is provided by Sun on the Sun website:

`http://www.sun.com/documentation`

The Sun StorEdge SAN Foundation Software 4.4.7a (SAN 4.4.7a) is a minimum version that has been qualified for Emulex legacy 2 GB HBAs.

The Sun StorEdge SAN Foundation Software 4.4.9 (SAN 4.4.9) is a minimum version that has been qualified for Emulex PCI-X 4 GB HBAs.

To install/upgrade the Firmware and/or Fcode for an Emulex legacy adapter, follow the *FCA Utilities Reference Manual* documentation which is located on the Emulex website:

`http://www.emulex.com/ts/docoem/sun/10k.htm`

Host configuration with QLogic HBAs

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved HBAs.

Sun SPARC-based servers support QLogic 2 GB/4 GB HBAs:

QLA2340-E-SP	(2 GB single port PCI-X adapter)
QLA2342-E-SP	(2 GB dual port PCI-X adapter)
QLA2460-E-SP	(4 GB single port PCI-X adapter)
QLA2462-E-SP	(4 GB dual port PCI-X adapter)
QLE2460-E-SP	(4 GB single port PCI Express adapter)
QLE2462-E-SP	(4 GB dual port PCI Express adapter)

There are two HBA drivers that can be used for QLogic HBAs:

- QLA2x00 driver
  - Supports 2 GB HBAs
- qlc driver
  - Supports 2 GB and 4 GB HBAs

CAUTION

EMC does not support FC-IP on the QLogic HBAs.

QLA2x00 driver

To install one or more EMC-approved QLogic host bus adapters (HBAs) into a Solaris host and configure the host for connection to the Symmetrix over a Fibre Channel, follow the procedures in *EMC Fibre Channel with QLogic Host Bus Adapters in the Solaris Environment*.

You can obtain the document from the QLogic website, as follows:

1. Access `http://www.qlogic.com`.
2. Click **Downloads** at the left side of the screen.
3. Click the **EMC** link to the right of **OEM approved/recommended drivers and firmware**.

4. Find the description of your HBA and driver in the **Name** column of the table for your HBA model. Then click the **Readme** link in the associated **Description** column.

qlc driver

The **qlc** driver is a part of the Sun StorEdge SAN Foundation Software (Sun SAN). The Sun SAN is embedded in the Solaris 10. However, you have to install the recommended patch 119130-xx for the latest qualified qlc driver (see the *EMC Support Matrix* for the current patch 119130-xx revision approval).

On Solaris 10, the Sun patch 119130-16 is a minimum version that has been qualified for QLogic PCI-X 4 GB adapters and PCI-E 4 GB adapters.

If you intended to use qlc driver on Solaris 8 and/or Solaris 9, you must follow the *Sun StorEdge SAN Foundation Software Installation Guide* which is provided by Sun on the Sun website:

`http://www.sun.com/documentation`

The Sun StorEdge SAN Foundation Software 4.4.9 (SAN 4.4.9) is a minimum version that has been qualified for QLogic legacy 2 GB/4 GB HBAs.

To install/upgrade the Fcode for a QLogic legacy adapter, you can use the *SANsurfer FC HBA CLI for Solaris SPARC* utility which provided by QLogic on the QLogic website:

`http://support.qlogic.com/support/sun_page.asp`

CAUTION

EMC approves using the "SANsurfer FC HBA CLI" utility for downloading Fcode only.

Host configuration with Sun HBAs

CAUTION

EMC does not support FC-IP on the Sun HBAs.

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved Sun HBAs.

The Sun HBAs include Sun-branded QLogic adapters and Sun-branded Emulex adapters.

The following are Sun-branded QLogic HBAs:

SG-XPCI1FC-QF2	(2 GB single port PCI-X adapter)
SG-XPCI2FC-QF2	(2 GB dual port PCI-X adapter)
SG-XPCI1FC-QF4	(4 GB single port PCI-X adapter)
SG-XPCI2FC-QF4	(4 GB dual port PCI-X adapter)
SG-XPCIE1FC-QF4	(4 GB single port PCI Express adapter)
SG-XPCIE2FC-QF4	(4 GB dual port PCI Express adapter)

The following are Sun-branded Emulex HBAs:

SG-XPCI1FC-EM2	(2 GB single port PCI-X adapter)
SG-XPCI2FC-EM2	(2 GB dual port PCI-X adapter)
SG-XPCI1FC-EM4	(4 GB single port PCI-X adapter)
SG-XPCI2FC-EM4	(4 GB dual port PCI-X adapter)
SG-XPCIE1FC-EM4	(4 GB single port PCI Express adapter)
SG-XPCIE2FC-EM4	(4 GB dual port PCI Express adapter)

The qlc device driver is used for Sun-branded QLogic adapters, and the emlxs device driver is used for Sun-branded Emulex adapters. The qlc and emlxs drivers are part of the Sun StorEdge SAN Foundation Software. This package driver is also called a SAN driver.

EMC qualifies and supports Sun HBAs on:

- Solaris 8 and Solaris 9
  - The SAN 4.2 is a minimum version that has been qualified for Sun-branded QLogic 2 GB adapters.
  - The SAN 4.4.7a is a minimum version that has been qualified for Sun-branded Emulex 2 GB adapters.
  - The SAN 4.4.9 is a minimum version that has been qualified for Sun-branded QLogic PCI-X 4 GB adapters and Sun-branded Emulex PCI-X 4 GB adapters.
- Solaris 10

  The Sun StorEdge SAN Foundation Software is embedded in the Solaris 10.

  - Solaris 10 (03/05) is a minimum OS version that has been qualified for Sun-branded QLogic 2 GB adapters.
  - Solaris 10 Update 1 (01/06) is a minimum version that has been qualified for Sun-branded Emulex 2/4 GB adapters and Sun-branded Qlogic 4 GB adapters.

If you intend to use Sun-branded Emulex adapters on the Solaris 10 prior of S10-U1, there are two packages SUNWemlxs and SUNWemlxu that are required before installing required patch 120222-XX. (Refer to the *EMC Support Matrix* for the most up-to-date support version). These packages are available on the Sun website:

```
http://www.sun.com/download/products.xml?id=42c4317d
```

To install the EMC-qualified Sun HBAs into a Solaris host and to configure the host connection to the EMC storage array over Fibre Channel, follow the installation guide that came with your HBAs for specific instructions on setting up that particular hardware.

You also can obtain the document from the Sun website:

```
http://www.sun.com/products-n-solutions/hardware/docs/
   Network_Storage_Solutions/Adapters/index.html
```

Configuring MPxIO for Symmetrix devices

MPxIO is a feature of the Sun SAN application that allows I/Os to fail over from one path to another available path and that automatically resumes on the original path when the original path is repaired.

To enable MPxIO support for EMC® Symmetrix devices on a SPARC server running:

- Solaris 8 or Solaris 9
  - Set the following parameters to the file /kernel/drv/scsi_vhci.conf:

```
mpxio-disable="no";
device-type-scsi-option-list="EMC     SYMMETRIX", "Symmetrix-option";
symmetrix-option=0x1000000;
```

- Solaris 10
  - Set to the file /kernel/drv/fp.conf parameter:

    ```
    mpxio-disable="no";
    ```

  - Set the following parameters to the file /kernel/drv/scsi_vhci.conf:

```
device-type-scsi-option-list="EMC     SYMMETRIX", "Symmetrix-option";
symmetrix-option=0x1000000;
```

Note: After `device-type-scsi-options-list=`, there are five spaces between `EMC` and `SYMMETRIX`.

Ensure that the qlc driver configuration file `/kernel/drv/qlc.conf` does not contain the global setting:

```
mpxio-disable="yes"
```

Note: MPxIO functionality requires enabling the Symmetrix director C-bit flag.

Addressing Symmetrix devices

This section describes the methods of addressing Symmetrix devices over Fibre Channel.

Arbitrated loop addressing

The Fibre Channel arbitrated loop (FC-AL) topology defines a method of addressing ports, arbitrating for use of the loop, and establishing a connection between Fibre Channel NL_Ports (level FC-2) on HBAs in the host and Fibre Channel directors (via their adapter cards) in the Symmetrix system. Once loop communications are established between the two NL_Ports, device addressing proceeds in accordance with the SCSI-3 Fibre Channel Protocol (SCSI-3 FCP, level FC-4).

The Loop Initialization Process (LIP) assigns a physical address (AL_PA) to each NL_Port in the loop. Ports that have a previously acquired AL_PA are allowed to keep it. If the address is not available, another address may be assigned, or the port may be set to non-participating mode.

Note: The AL_PA is the low-order 8 bits of the 24-bit address. (The upper 16 bits are used for Fibre Channel fabric addressing only; in FC-AL addresses, these bits are **x'0000'**.)

After the loop initialization is complete, the Symmetrix port can participate in a logical connection using the hard-assigned or soft-assigned address as its unique AL_PA. If the Symmetrix port is in non-participating mode, it is effectively off line and cannot make a logical connection with any other port.

A host initiating I/O with the Symmetrix system uses the AL_PA to request an open loop between itself and the Symmetrix port. Once the arbitration process has established a logical connection between the Symmetrix system and the host, addressing specific logical devices is done through the SCSI-3 FCP.

Fabric addressing

Each port on a device attached to a fabric is assigned a unique 64-bit identifier called a World Wide Port Name (WWPN). These names are factory-set on the HBAs in the hosts, and are generated on the Fibre Channel directors in the Symmetrix system.

Note: For comparison to Ethernet terminology, an HBA is analogous to a NIC card, and a WWPN to a MAC address.

Note: The ANSI standard also defines a World Wide Node Name (WWNN), but this name has not been consistently defined by the industry.

When an N_Port (host server or storage device) connects to the fabric, a login process occurs between the N_Port and the F_Port on the fabric switch. During this process, the devices agree on such operating parameters as class of service, flow control rules, and fabric addressing. The N_Port’s fabric address is assigned by the switch and sent to the N_Port. This value becomes the Source ID (SID) on the N_Port's outbound frames and the Destination ID (DID) on the N_Port's inbound frames.

The physical address is a pair of numbers that identify the switch and port, in the format **s,p**, where **s** is a domain ID and **p** is a value associated to a physical port in the domain. The physical address of the N_Port can change when a link is moved from one switch port to another switch port. The WWPN of the N_Port, however, does not change. A Name Server in the switch maintains a table of all logged-in devices, so N_Ports can automatically adjust to changes in the fabric address by keying off the WWPN.

The highest level of login that occurs is the process login. This is used to establish connectivity between the upper-level protocols on the nodes. An example is the login process that occurs at the SCSI FCP level between the HBA and the Symmetrix system.

Migrating from SCSI to Fibre Channel

A Symmetrix SCSI director has four ports, and a Fibre Channel director has two, four, or eight. When replacing a SCSI director with a Fibre Channel director, you must follow certain procedures to assure that the hosts will know which devices are connected to which Symmetrix ports after the replacement.

EMC provides a utility that automates much of the migration process. The procedure can be summarized as follows:

1. Run the Symmetrix Inquiry utility (`inq`) to identify the configuration before changing the hardware.
2. Perform host-specific operations for the following host environments:
   - "Migrating in the VERITAS VxVM environment" on page 55
   - "Migrating in the Solstice DiskSuite environment" on page 56
3. Run EMC script *emc_s2f*, as described under "Running the EMC migration script" on page 56.

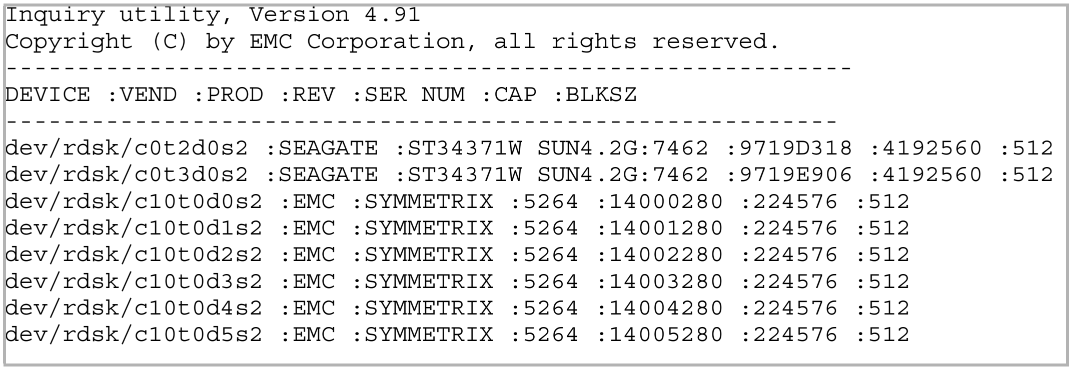
   Each script must be run before and after changing the hardware, as described in the appropriate sections. Run the "before" section as described.
4. Change the hardware.
5. Run the "after" parts of the script, followed by the host-specific steps (if applicable).

Running inquiry

You must identify the Symmetrix devices before making the hardware change. Run `inq` to display information you can use to determine which the Symmetrix volume is associated with a particular device as seen by the host.

You can find an executable copy of the `inq` command on EMC's anonymous FTP server, `ftp.emc.com`, in the `/pub/sym3000/inquiry/latest` directory. (Refer to "Obtaining files from the EMC FTP server" on page 35.)

Example The following figure shows a sample output of `inq` when run from the host console.



```
Inquiry utility, Version 4.91
Copyright (C) by EMC Corporation, all rights reserved.
-----------------------------------------------------------
DEVICE :VEND :PROD :REV :SER NUM :CAP :BLKSZ
----------------------------------------------------------
dev/rdsk/c0t2d0s2 :SEAGATE :ST34371W SUN4.2G:7462 :9719D318 :4192560 :512
dev/rdsk/c0t3d0s2 :SEAGATE :ST34371W SUN4.2G:7462 :9719E906 :4192560 :512
dev/rdsk/c10t0d0s2 :EMC :SYMMETRIX :5264 :14000280 :224576 :512
dev/rdsk/c10t0d1s2 :EMC :SYMMETRIX :5264 :14001280 :224576 :512
dev/rdsk/c10t0d2s2 :EMC :SYMMETRIX :5264 :14002280 :224576 :512
dev/rdsk/c10t0d3s2 :EMC :SYMMETRIX :5264 :14003280 :224576 :512
dev/rdsk/c10t0d4s2 :EMC :SYMMETRIX :5264 :14004280 :224576 :512
dev/rdsk/c10t0d5s2 :EMC :SYMMETRIX :5264 :14005280 :224576 :512
```

The output fields are as follows:

- `DEVICE` = UNIX device name (full pathname) for the SCSI device
- `VEND` = Vendor Information
- `PROD` = Product Name
- `REV` = Revision number — for a Symmetrix, this will be the microcode version
- `SER NUM` = Serial number, in the format `SSVVVDDP`, where:
  - `SS` = last two digits of the Symmetrix serial number
  - `VVV` = Logical Volume number
  - `DD` = Channel Director number
  - `P` = port on the channel director
- `CAP` = Size of the device in kilobytes
- `BLKSZ` = Size in bytes of each block

Migrating in the VERITAS VxVM environment

Note: Sliced Volumes with VERITAS Slice Tags will automatically be “rediscovered” at reboot. Simple Volumes (VERITAS on *s2*) will not.

1. Type **`umount -a`** and press ENTER to unmount all file systems.
2. Stop all volumes and “deport” the volume group:
   a. Type **`vxvol -g`** `<DiskGroup>` **`stopall`** and press ENTER.

   b. Type **`vxdg deport`** ***`<DiskGroup>`*** and press ENTER.

3. Run the EMC migration script. (Refer to "Running the EMC migration script" on page 56.) As described in that procedure, make the hardware changes at the appropriate time.
4. Import and recover the Disk Group:
   a. Type **vxdg import** ***<DiskGroup>*** and press ENTER.
   b. Type **vxrecover -g** ***<DiskGroup>*** and press ENTER.

Migrating in the Solstice DiskSuite environment

1. Type **copy /etc/opt/SUNWmd/md.tab** and press ENTER to back up the current DiskSuite configuration.
2. Record the output from `metastat` and `metadb -I`.
3. Make sure all 'sd' device links are recorded in `md.tab`. If they are not, add them and reboot.
4. Type **umount -a** and press ENTER to unmount all file systems.
5. Replace each replica currently configured per device:
   a. Type **metadb -d -f device** and press ENTER.
   b. Type **metadb -a -f -c x new_device** and press ENTER.
6. Edit `/etc/opt/SUNWmd/md.tab`, replacing each old device with its corresponding new device link.
7. Use `metaclear` to delete all volumes except the boot volume.
8. Use `metainit` to recreate all meta-devices.

Running the EMC migration script

The `emc_s2f` utility is used for most SCSI-to-Fibre Channel migration situations. An EMC shell script provides snapshots of the configuration before and after the director is replaced, so you can see where devices were reassigned.

Usage

The syntax of the utility is:

```
emc_s2f [-<option>] [-fp] [-all]
```

where:

<option> is one of these:

b — Specify when running `emc_s2f` before converting to Fibre Channel.

a — Specify when running `emc_s2f` after converting to Fibre Channel.

c — Specify to compare the "before" and "after" configurations.

-fp shows the full path of each device name.

-all (valid only if ***<option>*** is **c**) displays all information even if there is no change.

CAUTION

If the operating system's device name is not uniquely identified by its basename, you *must* use the -fp flag.

Limitations

Note the following limitations of `emc_s2f`:

- The comparison will not be accurate if the host is connected to multiple Symmetrix devices and the last two digits in the serial number of one Symmetrix are the same as the last two digits of the serial number of another Symmetrix.
- If multiple paths exist to the host before *and* after the migration, the "before" and "after" groups of devices will be displayed, but there will be no way to tell how the devices match each other.
- The Inquiry utility will not work on HP devices with the NIO driver. This driver does not accept the SCSI passthrough commands that are needed by **Inq**. Before running `emc_s2f` under these circumstances, be sure to create pseudo devices.

Procedure

Follow these steps to run the script:

1. Unmount the file systems.
2. Type **`emc_s2f -b`** and press ENTER to take a snapshot of the configuration before you change the hardware. The information is displayed, and written to a file named `emc_s2f.b`.
3. Replace the necessary hardware, then bring the Symmetrix back on line.
4. Type **`emc_s2f -a`** and press ENTER to take a snapshot of the new hardware configuration. The information is displayed, and written to a file named `emc_s2f.a`.

5. Type **`emc_s2f -c`** and press ENTER to compare the two files. The information is displayed, and written to a file named `emc_s2f.c`.

   Here is a sample output, from a Symmetrix system with a serial number ending in **65**:

```
65          002         B       c8t0d2          c9t8d2
                        A       c1t0d2          c4t8d2
65          040         B       c9t3d0
                        A       not_found
65          047         B       c9t3d7
                        A       c4t3d7
65          048         B       c9t4d0
                        A       no_change
```

   Before, dev #002 was seen through two ports; after, it is seen through two different ports.

   Before, dev #040 was seen as c9t3d0; after, the device is not visible to the host.

   Before, dev #047 was seen as c9t3d7; after, it is seen as c4t3d7.

   Before, dev #048 was seen as c9t4d0; after, there is no change. (This is shown only if the **-all** flag was used.)

3

Solaris SPARC and Symmetrix over iSCSI

This chapter contains Symmetrix Multi-Protocol Channel Director (MPCD) iSCSI connectivity implementation details for the Sun Solaris iSCSI software initiator kernel mode driver.

Hardware

Symmetrix iSCSI multiprotocol channel director (MPCD) is supported with Sun Gigabit Network Interface Cards (NIC) in the direct connect and the IP Switch environments.

Refer to the "iSCSI via Symmetrix Multi-Protocol Channel Director" section in Appendix A of the the *EMC Networked Storage Topology Guide* (available on http://Powerlink.EMC.com) for further information on the supported topologies.

Software

Sun iSCSI driver embedded in the Solaris 10 Update 1 or later. The iSCSI driver is included of two packages:

- SUNWiscsir - Sun iSCSI device driver
- SUNWiscsiu - Sun iSCSI management utilities

Addressing

Sun uses SCSI-2 device access protocol in addressing iSCSI devices, up to 256 (0 to 255) LUNs per network interface port.

Configuring Solaris iSCSI initiators

Refer to the Sun document *System Administration Guide* (available on http://docs.sun.com/app/docs/doc/819-2723?q=iscsi) to configure the Solaris iSCSI initiators.

Configuring Symmetrix iSCSI director

Refer to the section "Fibre Bit Settings" under "Symmetrix DMX Series" in the *EMC Support Matrix* for the recommended director bit setting for Sun servers.

Solaris iSCSI/Symmetrix case studies

The following are two basic case studies that incorporate information of the Symmetrix iSCSI MPCD and Solaris iSCSI host configurations.

Case study 1

Figure 1 show GigE Network adapters connecting directly to the iSCSI MPCD ports.

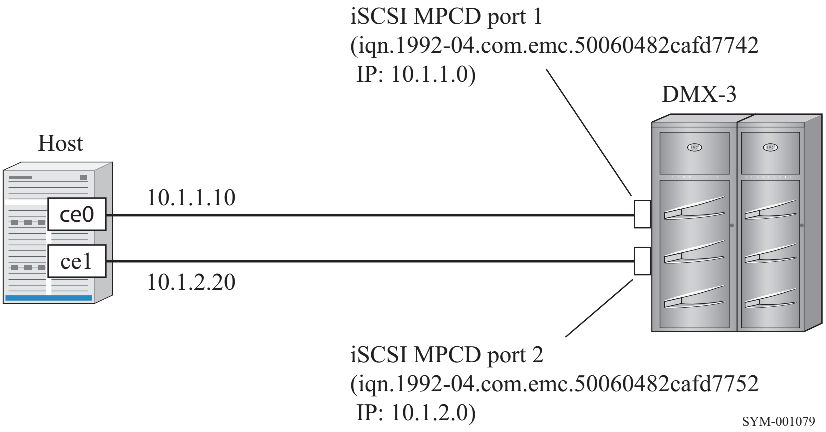


Figure 1 Connection directly to iSCSI MPCD ports

Case study 2

Figure 2 on page 62 shows GigE Network adapters connecting to the iSCSI MPCD ports via the IP Switch.

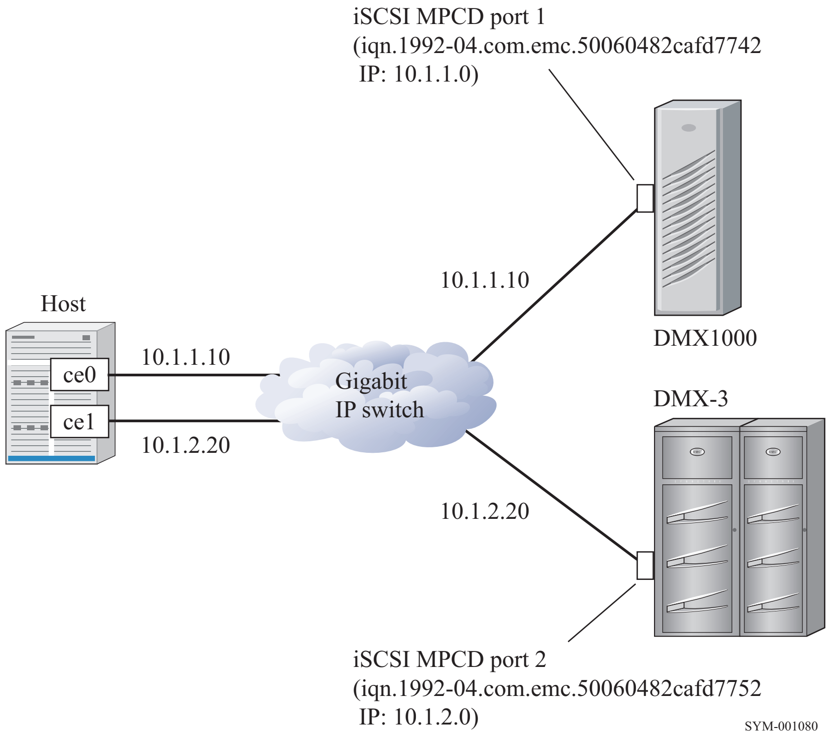


Figure 2 **Connection to iSCSI MPCD ports via IP switch**

Symmetrix configuration

"Case study 1" on page 61 and "Case study 2" on page 61 have the same iSCSI MPCD Channel Information settings.

1. Set "Primary IP Address" on the same subnet with the GigE Network adapters:

   Port 1: 10.1.1.0
   Port 2: 10.1.2.0

2. Set "Max Transmission":

   Port 1: 1500 (default)
   Port 2: 1500 (default)

3. Set “IP Mask” as same as the GigE Network adapters IP mask:

   Port 1: IP Mask = 255.255.255.0
   Port 2: IP Mask = 255.255.255.0

4. Set “IP DNS Group”:

   Port 1: NONE (default)
   Port 2: NONE (default)

5. Set “SNMP”:

   Port 1: YES (default)
   Port 2: YES (default)

6. Set “Default Gateway”:

   Port 1: 0.0.0.0
   Port 2: 0.0.0.0

7. Set “ISNS IP Address”:

   Port 1: 0.0.0.0
   Port 2: 0.0.0.0

Sun host configuration

“Case study 1” on page 61 and “Case study 2” on page 61 have the same host settings.

1. Enable network interface for each GigE Network adapter:

```
# ifconfig ce0 plumb
# ifconfig ce1 plumb
```

2. Set IP for each interface:

```
# ifconfig ce0 10.1.1.10 netmask 255.255.255.0 up
# ifconfig ce1 10.1.2.20 netmask 255.255.255.0 up
```

3. Add netmask value for the interfaces to the file /etc/inet/netmasks:

```
10.1.1.0 255.255.255.0
10.1.2.0 255.255.255.0
```

4. Add IP address of each interface to the file /etc/hosts:

```
10.1.1.10 iSCSI0
10.1.2.20 iSCSI1
```

5. Create host network file for each interface port:

   ```
   /etc/hostname.ce0  contains iSCSI0
   /etc/hostname.ce1  contains iSCSI1
   ```

6. You can use the static discovery method or SendTargets device discovery method:

   - Configure the static target discovery method:

     ```
     # iscsiadm add static-config
     iqn.1992-04.com.emc.50060482cafd7742,10.1.1.0:3260
     ```

     ```
     # iscsiadm add static-config
     iqn.1992-04.com.emc.50060482cafd7752,10.1.2.0:3260
     ```

   - Configure the SendTargets device discovery method:

     ```
     # iscsiadm add discovery-address 10.1.1.0:3260
     ```

     ```
     # iscsiadm add discovery-address 10.1.2.0:3260
     ```

7. Enable the iSCSI target discovery method

   - If you have configured the static discovery method, enable the static target discovery:

     ```
     # iscsiadm modify discovery –s enable
     ```

   - If you have configured the SendTargets discovery method, enable the SendTargets discovery:

     ```
     # iscsiadm modify discovery –t enable
     ```

CAUTION

You can only enable one discovery method at a time. If both SendTarget and Static discovery methods are enabled at the same time that may cause the host to PANIC.

8. Reboot the host with reconfigure for the changes to take effect:

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

9. If the host isn't detected to any iSCSI devices, use the following command to create iSCSI device nodes:

   ```
   # devfsadm –i iscsi
   ```

4

Solaris SPARC and Symmetrix over SCSI

This chapter provides information specific to Sun Solaris hosts connecting to Symmetrix systems over SCSI.

Sun Solaris SPARC/Symmetrix SCSI environment

For Symmetrix systems connected to Sun Solaris hosts over SCSI, note the following requirements and recommendations.

Symmetrix SCSI directors

EMC recommends using the Symmetrix SCSI directors shown in Table 2.

Table 2 Recommended Symmetrix SCSI director models

Symmetrix model	Recommended SCSI director(s)
5700/3700 54xx/34xx 53xx/33xx 8130/8230 8430/8530 8730/8830	Fast-Wide Differential SCSI director P/N 201-207-917 (60 MHz processor) Fast-Wide Differential SCSI director P/N 201-207-927 (66 MHz processor) Ultra-Wide Differential SCSI director P/N 201-277-917 (60 MHz processor) Ultra-Wide Differential SCSI director P/N 201-277-927 (66 MHz processor)
5500/3500 52xx/32xx 51xx/31xx	Fast-Wide Differential SCSI director P/N 200-881-903

Sun SCSI controllers

Refer to the *EMC Support Matrix* for supported SCSI controllers.

SPARCstations and SPARCservers

Perform this procedure prior to booting the operating system when running Solaris on SPARCstations and SPARCservers.

1. Power on the Sun server and wait for it to perform its self-tests.
2. When the system banner appears on the screen display, press BREAK if you are using an ASCII terminal or STOP-A if you are using a workstation terminal.
3. At the `OK>` prompt, type **`probe-scsi-all`** and press ENTER.

The host looks for all attached SCSI devices and displays information similar to the following (Units 0–3 are LUNs):

```
     Target 0
     Unit 0 Disk  EMC SYMMETRIX 50607801E153
        Copyright (c) 1994
        EMC Corp. All rights reserved.
Unit 1 Disk  EMC SYMMETRIX 50607801D153
        Copyright (c) 1994
        EMC Corp. All rights reserved.
Unit 2 Disk  EMC SYMMETRIX 50607801F160
         Copyright (c) 1994
        EMC Corp. All rights reserved.
Unit 3 Disk  EMC SYMMETRIX 506078020161
        Copyright (c) 1994
        EMC Corp. All rights reserved.
```

The disk value returned (`50607810E153`, for example) provides additional detail. The syntax of this number is `MMMMWWXXXYYZ`, where:

`MMMM` = Symmetrix microcode revision level

`WW` = Last two digits of the Symmetrix serial number

`XXX` = Symmetrix device number

`YY` = Symmetrix SCSI director number

`Z` = Symmetrix SCSI director port number

4. Make sure the target ID and LUNs listed agree with the device information configured in the Symmetrix unit.

Note: If this information does not agree, check the Symmetrix-host port connection and the AutoInstall configuration.

5. At the `OK>` prompt, type **`boot -r`** and press ENTER.

The host boots to a login prompt. The host also issues a warning message for each Symmetrix logical unit attached but not yet labeled.

UltraSPARC series

On certain SUN platforms such as the SUN UltraSPARC Series, the following message may appear when you run `probe-scsi-all`:

```
   This command may hang the system if a STOP-A or halt
command has been exectued. Please type reset-all to
reset the system before executing this command.
   Do you wish to continue? (Y/N) N
```

Before running the `reset-all` command, you must set the **auto-boot?** parameter to **false** to prevent automatic reboot of the system after reset:

```
sentenv auto-boot? false
```

Use the `printenv` command to verify.

Next, run the `reset-all` command:

```
reset-all
```

Then, run the `probe-scsi-all` command:

```
probe-scsi-all
```

Set the **auto-boot?** parameter back to **TRUE** if desired:

```
setenv auto-boot? true
```

Resume system operation:

```
go
```

Determining hardware mapping

Follow these steps to determine the target IDs available to the Symmetrix system:

If the host is not powered up:

1. Power on the Sun host system and wait for it to perform its self-tests.
2. At the Open Firmware prompt (`OK>`), type **`printenv`** and press ENTER.
3. Review the listing that appears and determine the value of **scsi_initiator_id**. This target ID is reserved for the system and may not be used by any other device.

If the host is powered up:

1. At the system prompt, type **`eeprom`** and press ENTER.
2. Review the listing that appears and determine the value of **scsi_initiator_id**. This target ID is reserved for the system and may not be used by any other device.

Note: *Do no*t use the **scsi_initiator_id** value as a target ID value for Symmetrix devices.

Recognizing LUNs

Symmetrix devices are addressed on the SCSI bus using SCSI target IDs and SCSI LUNs. Each target ID can have up to 16 LUNs associated with it. By default, Solaris only searches for target IDs on the SCSI bus. If LUN addressing is required in addition to target ID addressing, you will need to modify the file /kernel/drv/sd.conf to have the following definitions for each target ID which needs to support multiple LUNs.

```
name=”sd” class=”scsi”
   target=0 lun=0;
name=”sd” class=”scsi”
   target=0 lun=1;
name=”sd” class=”scsi”
   target=0 lun=2;
name=”sd” class=”scsi”
   target=0 lun=3;
name=”sd” class=”scsi”
   target=0 lun=4;
name=”sd” class=”scsi”
   target=0 lun=5;
name=”sd” class=”scsi”
   target=0 lun=6;
name=”sd” class=”scsi”
   target=0 lun=7;
```

This addition to `sd.conf` is necessary when using LUN addressing. Make sure that for target IDs that do not need LUN support, only the `target=x lun=0` line is specified. This decreases the time needed to reboot the system.

To enable Solaris to support the Symmetrix system, EMC provides Symmetrix device definition files. The device definition files are available on EMC’s FTP server, `ftp.emc.com`, in `/pub/symm3000/solaris`.

Note: “Obtaining device definition files” on page 36 contains instructions on how to transfer these files to your host.

Modifying the system specification file

Two parameters in the file `/etc/system` require modification when operating in a Solaris environment:

- **sd_io_time**
- **sd_max_throttle**

To modify these parameters:

1. Set **sd_io_time** to 120 seconds. This setting prevents the host from issuing warning messages while non-disruptive operations are performed on the Symmetrix system:

   ```
   set sd:sd_io_time = 0x78
   ```

2. Set **sd_max_throttle** to 20. This setting prevents the host from over-sending tag queuing commands which may cause **scsi cmd timeout** and **scsi bus reset**:

   ```
   set sd:sd_max_throttle = 20
   ```

   A maximum throttle setting of 20 means that each host device instance will have no more than 20 commands outstanding (incomplete IO's from the standpoint of the operating system) at any given time. The value of 20 was arrived at by testing the incremental gains of increasing queue depth, and it was found that a queue depth of 20 represents a point where negligible incremental performance gains will usually be reached. It does not make sense to additionally offload IO onto the stack, and thereby unnecessarily use up resources throughout the stack, for no performance gain. A balance should be found.

   In the case of meta devices (which have more physical devices on the back end and can thus physically process more IO's in parallel), it may be beneficial to increase the queue depth to 32. It is important to note that in Solaris the `sd_max_throttle/ssd_max_throttle` settings are global, so all devices including non-meta's will also be affected.

   The max throttle setting of 20 is suitable for many environments. However, in some situations this value can be further fine tuned for configuration-specific optimizations. Your local EMC performance expert can assist with fine tuning recommendations, if any.

Formatting, partitioning, and labeling

Use the **`format`** command to partition and label the new devices. The devices will appear under `/dev/dsk`.

Note: Internal Sun drives on SPARC 5, 10, 20, Ultra1, Ultra2, Ultra30, Ultra60 and Ultra450 are usually found on SCSI controller 0. Symmetrix drives typically start at SCSI controller 1 or SCSI controller 2. This may vary on SPARC 1000 and 2000, Ultra 3X00, 4X00, 5X00, 6X00, and 10000.

If you are running Solaris 2.3, follow the instructions below. If you are running Solaris 2.4 or higher, go to the the next section.

Solaris 2.3

To partition and label new devices:

1. At the system prompt, type **`format`** and press ENTER.

   The host searches for all disks. It generates a display placing all unlabeled disks at the beginning of the listing.

2. At the `Specify Disk` prompt, enter the number of the first EMC drive.

   The **Format** menu appears.

3. At the `format>` prompt, type the word **`type`** and press ENTER.

4. At the `Specify disk type (enter its number):` prompt, enter the disk type number from the list that appears on the screen, or enter **0** for Auto-configuration.

5. At the `format>` prompt, type **`l`** (lowercase L, for label) and press ENTER.

6. At the `Ready to label disk, continue?` prompt, type **`y`** and press ENTER.

7. At the `format>` prompt, type **`disk`** and press ENTER.

8. Repeat steps 3 through 7 above for all remaining unknown Symmetrix devices.

If you wish to use raw devices, there is no need to create file systems for each device. Otherwise, create new file systems for each device as described under "Creating and mounting a file system" on page 27.

Solaris 2.4 and later

To partition and label new devices:

1. At the `root` prompt, type **`format`** and press ENTER.

   The host searches for all disks. It generates a display placing all unlabeled disks at the beginning of the listing.

2. At the `Specify Disk` prompt, enter the number of the first EMC drive.

   The Format menu appears.

3. At the `format>` prompt, type **`label`** and press ENTER.
4. At the `Disk not labeled. Label it now?` prompt, type **`Y`** and press ENTER.
5. Type **`disk`** and press ENTER at the prompt to display a listing of the disks.
6. Repeat steps 2 through 5 for all Symmetrix disks.

If you wish to use raw devices, there is no need to create file systems for each device. Otherwise, create new file systems for each device as described under "Creating and mounting a file system" on page 27.

Adding devices online

Whenever devices are added online to the Symmetrix unit or device channel addresses are changed, you will need to perform the actions described below in order to introduce the new devices to the system.

CAUTION

Before adding new devices online, the device entries must be defined in the file `/kernel/drv/sd.conf` and the host must be rebooted for the changes to take effect.

To add new Symmetrix devices while online in the Solaris environment:

1. Confirm that the host system has been rebooted since the new devices were defined in `/kernel/drv/sd.conf`. If it has not, reboot it.

2. Use the online upgrade feature at the Symmetrix service processor to add/map new drives to the SCSI host channel.

3. Execute the following two utilities:

```
drvconfig
disks
```

4. Follow the instructions under "Formatting, partitioning, and labeling" on page 72 to introduce new devices to the host environment.

5

Sun Cluster 2.x and High-Availability Environment

This chapter discusses Symmetrix / Sun Cluster 2.x environment. Fundamental concepts and procedures related to Sun Cluster planning, setup, and administration are provided.

Sun Cluster 2.x overview

This section introduces Sun Cluster 2.x and briefly describes its salient features.

What is Sun Cluster 2.x?

Sun Cluster 2.x is high availability cluster software for Sun Ultra Enterprise systems. Sun Cluster allows combinations of different Sun host types, in clusters of two to four active nodes.

Sun Cluster 2.x combines the functionality that was present in Sun Solstice HA 1.3 and Sun PDB (Parallel Database) 1.2.

Protection for data services

Sun Cluster 2.x provides automated detection, monitoring, recovery, and failover of critical data services. A data service is a collection of resources such as network interfaces, databases such as Sybase or Oracle, disk groups, or other applications that need to be made highly available.

The data service is commonly assigned a Virtual IP address or an IP alias address to form an entity called a logical host. The logical host fails over between cluster nodes. Logical hosts are highly available entities that consist of a migrating IP alias address, volume manager shared disksets or diskgroups, and one or more user data applications such as NFS, Sybase, or Oracle.

Cascading configuration

Sun Cluster 2.x supports a pre-assigned cascading configuration. In this configuration, any logical host can fail over to any other host in a pre-assigned order. When the logical host is created, you specify the ordering of the nodes by which the logical hosts will fail over.

The shared disk architecture used in Sun Cluster 2.x parallel databases provide increased availability by allowing users to simultaneously access a single shared database through multiple cluster nodes. If one node fails, users can still access the database through another node.

Private heartbeat link requirement

Sun Cluster 2.x requires two private heartbeat interface links between each node for cluster control and heartbeat message exchange. Heartbeat links are typically implemented with Gigabit Ethernet, Fast Ethernet or Scalable Coherent Interface (SCI) connections between the two systems. These links are redundant, requiring only one for continued system operation.

Public network requirement

Sun Cluster 2.x requires at least one public network connection to a Local Area Network. Sun Cluster HA uses Virtual IP or IP alias addressing to establish a map between multiple logical host names and a single physical network interface. This enables one physical interface to respond to multiple logical host names. The physical interface on the host on which the logical interface is currently configured services packets destined for that logical host.

About logical host names

When all logical hosts are *mastered* by their respective, default master hosts, only one logical host name is associated with the physical network connection. When a takeover occurs, however, two logical host names are then associated with the single physical network connection.

You must assign a unique host name for each logical host on each public network.

Membership and fault monitors

Sun Cluster 2.x provides a membership monitor and a fault monitor. The membership monitor detects total failure of a system in the Sun Cluster configuration, while the fault monitor detects failures of individual services.

The principal function of the membership monitor is to make sure the servers are synchronized and to coordinate the configuration of the applications and services when the configuration state changes.

The fault monitor consists of a fault daemon and programs used to probe various parts of the data services. These probes are executed periodically by the fault daemon to ensure that services are working.

Table 3 summarizes Sun Cluster 2.x features.

Table 3 Sun Cluster 2.x features, by version and cluster type

Features/Capabilities	V2.0 HA	V2.0 PDB	V2.1 HA	V2.1 PDB	V2.2 HA	V2.2 PDB
Maximum number of nodes in cluster	4	2	4	2	4	4
Maximum number of active nodes	4	2	4	2	4	4
Symmetric configuration (two nodes)	yes	-	yes	-	yes	-
Asymmetric configuration (two nodes)	yes	-	yes	-	yes	-
Cascading configuration	yes	-	yes	-	yes	-
Sun Solaris 2.5.1 and greater	yes	yes	yes	yes	no	no
Sun Solaris 2.6 or 2.7	no	no	no	no	yes	yes
Sun Solstice DiskSuite (two nodes)	no	no	no	no	no	no
Sun Solstice DiskSuite (four nodes)	no	no	no	no	yes	no
Sun Enterprise Volume Manager (SEVM 2.4)	yes	no	yes	no	no	no
Sun Enterprise Volume Manager (SEVM 2.5)	no	no	no	no	yes	no
Cluster Volume Manager (CVM 2.2)	no	yes	no	yes	no	yes
VERITAS Volume Manager (VxVM 3.0.4)	no	no	no	no	yes	yes
VERITAS Journal File System (VxFS 2.3)	yes	-	yes	-	yes	-
Supports Sybase database v11.0	yes	-	yes	-	no	-
Supports Sybase database v11.5	no	-	no	-	yes	-
Supports Oracle database v7.3 and greater	yes	-	yes	-	no	-
Supports Oracle database v7.3.4 and 8.0.4	no	-	no	-	yes	-
Supports Oracle Parallel Server v7.3.x	no	yes	no	yes	no	yes
Supports Oracle Parallel Server v8.0.x	no	no	no	yes	no	yes
Supports Oracle Parallel Server v8.1.5	no	no	no	no	no	yes
Supports Informix database v7.2	yes	-	yes	-	no	-

Table 3 Sun Cluster 2.x features, by version and cluster type (continued)

Features/Capabilities	V2.0 HA	V2.0 PDB	V2.1 HA	V2.1 PDB	V2.2 HA	V2.2 PDB
Supports Informix database v7.23 and 7.30	no	-	no	-	yes	-
Supports NFS	yes	-	yes	-	yes	-
Supports IP interface failover	yes	yes	yes	yes	yes	yes
Supports heartbeat link via disk	no	no	no	no	no	no
Supports virtual IP or IP alias addressing	yes	-	yes	-	yes	-

Volume Managers deployed with Sun Cluster

Different volume managers are deployed with different releases of Sun Cluster, as shown in Table 4.

Table 4 Volume Managers deployed with specific releases of Sun Cluster

Sun Cluster releases	Volume Managers
Sun Cluster v2.2	• Sun StorEdge Volume Manager (SSVM) [a] • Cluster Volume Manager (CVM) [b] • Solstice DiskSuite (SDS) 4.2 • VERITAS Volume Manager (VxVM 3.0.4)
Sun Cluster v2.0/v2.1	• Sun Enterprise Volume Manager (SEVM) [b] • Cluster Volume Manager (CVM) [c]

a. Sun StorEdge Volume Manager (SSVM) is the same as VERITAS VxVM 2.5.

b. Sun Enterprise Volume Manager (SEVM) is the same as VERITAS VxVM v2.4/v2.5.

c. Cluster Volume Manager (CVM) is the same as VERITAS VxVM 2.2-1 with extentions for shared disk group management.

Solstice Disk Suite Volume Manager

Solstice Disk Suite (SDS) is one of the Sun logical volume manager tools that can used with SCv2.2. SDS supports large file systems, file system expansion, and volume manager-level intent logging (called UFS logging) for fast file system recovery.

Note: Make sure that you install all required Sun patches for SDS. Refer to the SDS release notes and check with Sun customer support to identify all required patches for the appropriate Solaris operating system and SDS version.

Consider the following when using SDS volume manager in the Sun Cluster and Symmetrix environments:

Note: For detailed information on the setup and management of SDS in the Sun environment, refer to the *Solstice DiskSuite 4.0 (or 4.1/4.2) Administrator's Guide*. For details on setting up SDS in the Sun Cluster environment, refer to the *Sun Cluster 2.0 (or 2.1/2.2) System Administration Guide*.

- Solstice DiskSuite 4.2 supports shared disks that are grouped into entities called *disk sets*.
- Sun Cluster 2.2, using SDS 4.2 or above, does not have a two-disk set, two-host limitation. More than two disk sets can be created, all of which can be shared by a maximum of four hosts.
- Disks added to a shared disk set are automatically formatted to contain slice s7, the private region used to store a database replica, and slice s0, the public region or user space. SDS automatically maintains the number and distribution of database replicas among the shared disks in the disk set.
- Sun Cluster controls the reservation and release of the disk set during failover. SDS uses SCSI reservation on a per-disk basis to prevent simultaneous access of the disk set by more than one host. SDS probes the disks once every second to determine if the host still owns the disks. The host will panic if SDS determines that the host has lost the disk set reservation. This can happen if another host forcibly reserves (takes over) the diskset.

VERITAS Volume Manager (VxVM, SEVM, SSVM)

VxVM is a volume manager that can be used in Sun Cluster configurations. Depending on the version of Sun Cluster, VxVM is referred to as VERITAS Volume Manager, Sun Enterprise Volume Manager (SEVM), or Sun StorEdge Volume Manager (SSVM).

Here are some basic considerations for using VxVM in a Sun Cluster and Symmetrix environment:

Note: For detailed information on the setup and management of VxVM in the Sun Cluster environment, refer to the *Sun Cluster System Administration Guide* (P/N 805-4238-10).

- A disk group defined with VxVM can be presented to two or more hosts. You do this by deporting the disk group from one host and then importing it to the other host. A disk group can be *owned* by only one host at a time in an HA configuration. Do not attempt to deport (or import) a disk group from (or to) more than one host simultaneously, unless you are running an OPS configuration.

- Sun Cluster failover scripts control disk group importing and deporting. Once the host has successfully imported a disk group, the volumes can be started and used as raw volumes or file systems (UFS or VxFS). These volumes or file systems can then be mounted. Sun Cluster automates all importing, deporting, starting, and mounting of all VERITAS volumes and file systems (UFS or VxFS).
- When configured for transparent NFS failover, major and minor device numbers must match on all hosts.

Note: Refer to the Sun Cluster documentation for information on checking and reconfiguring the major and minor numbers for VERITAS volumes.

- Initialize the *rootdg* using only internal disks. Use a small partition (about six cylinders) to initialize the rootdg disk group. Shared disks should never be added to the rootdg disk group, since the rootdg is local only to one host; the rootdg disk group cannot be shared with other hosts.
- The configuration for each disk group should be backed up to tape or a spare disk. If there is ever a split-brain failure in which two hosts simultaneously import the same diskgroup, the configuration stored in the private regions of all the disks in the disk group can become corrupted. The backup copy of the VERITAS configuration can be used to rebuild the entire VERITAS volume configuration. The data remains unaffected, and is accessible if all volume configurations and specifications that are present following the rebuild match the original configurations and specifications.
- Sun Cluster imports a disk group using the **-t** option to disable the AutoImport feature. The **-t** option sets the `noautoimport` flag on all disks in the disk group to prevent an already imported disk group from being automatically imported by another node during system boot up. If you manually import the shared disk group, you must use the **-t** option to prevent corruption of the VERITAS volume configuration database.

Sun Cluster 2.x and Symmetrix connections

This section describes the connections that are used within a Sun Cluster 2.x, and the types of connections used to incorporate Symmetrix into a Sun Cluster 2.x.

Sun Cluster connections

All cluster nodes are connected to two common networks (IP subnets), which make up the private network. The private network interfaces are used to monitor heartbeat messages and pass cluster control messages between cluster nodes.

Hosts communicates via one or more public networks, which are separate from the private heartbeat links. Multiple networks can be deployed in a primary and backup arrangement to support public IP interface failover.

Symmetrix connections within a Sun Cluster

When you integrate Symmetrix into a Sun Cluster 2.x environment, you must make sure that all of the volumes used by a given application can be accessed by each node in the cluster. This is important because the application must retain access to all of its volumes when the application fails over to a different node in the cluster.

Symmetrix can connect to host computers using the following types of connections:

- Fast-Wide Differential (FWD) SCSI connections
- Ultra-Wide Differential (UWD) SCSI connections
- Fibre Channel single-path connections
- Fibre Channel multi-path fabric and SCSI connections (for use with EMC PowerPath® or host-based mirroring)

You define disk resources on each cluster host the same way you would do so for a single-host system. The device definition, volume creation, and file system creation procedures are identical.

Symmetrix director bit settings

Refer to the *EMC Support Matrix* for information on bit settings.

Quorum device requirements

For two-node configurations, you must configure, at minimum, a 6-cylinder device for use as the quorum device. In addition, you must configure two 40-cylinder devices for the Cluster Configuration Database (CCD). Each 40-cylinder device must be mapped to a different Symmetrix director for high availability. Sun Cluster HA mirrors across these two devices.

For four-node configurations, you must configure two six-cylinder devices for use as the quorum device. Even though the terminal concentrator is used as a network partition voting or locking mechanism and not the quorum disk in a four-node cluster, the quorum disk should be assigned and mapped to two Symmetrix director ports in case the Sun Cluster implementation changes in the future.

Optionally, configure two 40-cylinder devices reserved for future use as CCD devices. Currently, the CCD is stored on the internal boot disk of each node. If Sun changes the implementation, the CCD devices will be ready for use. Each 40-cylinder device should be mapped to a different Symmetrix director port for high availability.

Symmetrix/Sun Cluster cabling guidelines

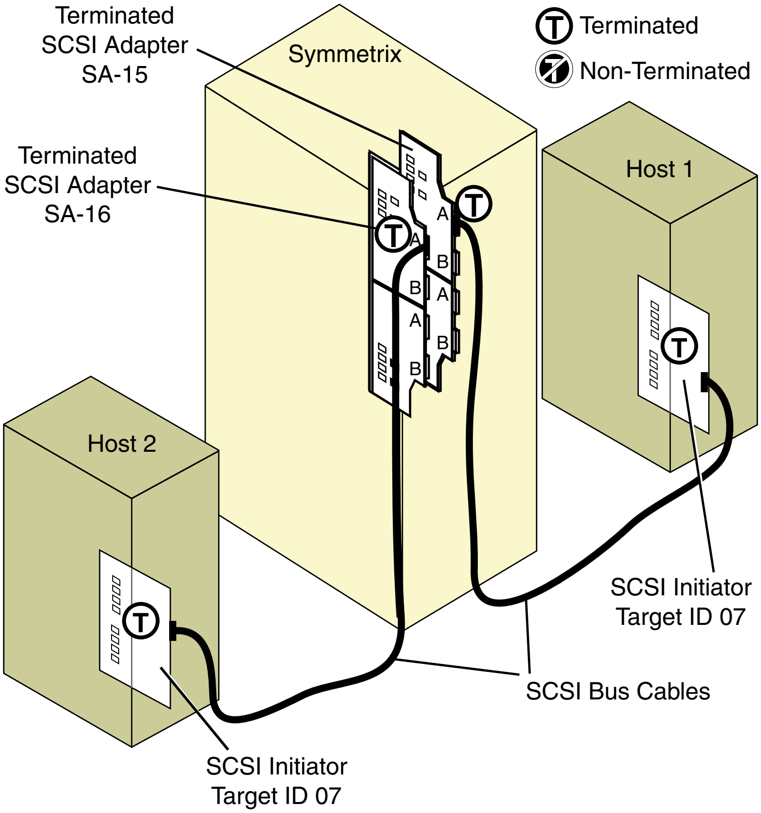
The Symmetrix ICDA architecture features multi-port switching that allows access to the same devices through multiple SCSI or Fibre channel buses. This capability allows access by multiple cluster nodes.

Each access path to the Symmetrix system is through a separate interface:

- Fibre Channel interface — Connection is to a Symmetrix Fibre Channel adapter port via a Fibre Channel cable.
- Single-initiator SCSI interface — Connection is to a Symmetrix FWD-SCSI or Ultra-SCSI adapter port with termination enabled, via a standard fast wide SCSI cable (C##M-68S model).

When connecting a cluster host to Symmetrix channel adapters, be sure to alternate between odd and even adapters. In an HA configuration, connecting the standby host to an opposite even or odd adapter helps to distribute the I/O load in a takeover situation and eliminates the SCSI/Fibre Channel path and internal Symmetrix bus as a potential single point of failure.

Figure 3 on page 84 shows a cabling example in a cluster environment.



Ensure that SCSI bus terminators are on both the host initiators and the Symmetrix SCSI adapters.

Figure 3 **Multipath, single-initiator, two-host cluster**

Failure fencing and quorum voting

This section briefly describes two key concepts of Sun Clusters: failure fencing and quorum voting. Failure fencing and quorum voting determine how the software keeps track of disk ownership and make sure that services continue following a failure.

After a node in a clustered system has failed and is no longer in the cluster, it must be prevented from writing to the multi-host disk devices. Data corruption can occur if a failed node is able to write to shared disk devices while surviving nodes are accessing those devices. Sun Cluster software uses the term *failure fencing* to refer to the mechanisms used for preventing failed nodes from accessing shared storage devices.

Depending on the cluster topology, failure fencing and quorum majority voting are handled differently. A brief discussion of the various cluster topologies and how Sun Cluster implements failure fencing and quorum voting follows.

Note: Refer to the *Sun Cluster Software Installation Guide* for detailed information.

Failure fencing in VxVM, SSVM, SEVM, CVM configurations

This section describes failure fencing configurations within the contexts of the following volume managers: VxVM, SSVM, SEVM, and CVM.

Two-node configurations

If, for any reason, one or more connections are lost, both nodes initiate a cluster reconfiguration process. In a two-node configuration, the quorum device determines which node remains in the cluster; the failed node cannot start its own cluster since it cannot reserve the quorum device. The SCSI-2 reservation mechanism is used in Sun Cluster to fence a failed node and prevent access to the shared storage devices.

Configurations with more than two nodes

The SCSI-2 reservation mechanism breaks down in configurations of more than two nodes. Since reservations are host-specific, a reservation from one node would effectively fence all other surviving nodes if a failure is detected in the cluster.

Instead of the quorum scheme, Sun Cluster uses a cluster-wide lock mechinism called a *nodelock*. One of the cluster members always holds this lock from the time the first node successfully forms a new cluster until the last node leaves the cluster. If the node holding the lock fails, the lock is automatically moved to another node.

The node lock uses a port on the terminal concentrator or the System Service Processor. The location of the node lock is determined during the Sun Cluster software installation process. The information is then stored in the Cluster Configuration Database.

Failure fencing in SDS configurations

In Sun Cluster HA configurations using Solstice Disk Suite as the volume manager, it is the volume manager that determines cluster quorum and failure fencing, regardless of the cluster topology.

No quorum device or nodelock mechinism is defined during the Sun Cluster software installation. There is no concept of shared disks in the HA environment.

At most, only one node can master (own) an SDS diskset at any one time. The solution to disk fencing relies on the SCSI-2 concept of disk reservation, which requires that a disk be reserved by exactly one node. The SCSI reservation is accomplished using multi-host "ioctls."

For SDS configurations, the Sun Cluster installation program does not prompt the administrator to supply a quorum device selection or a failure fencing partition management policy, as it does for configurations using VxVM, SEVM, SSVM or CVM.

Quorum majority voting

In the event of dual failures on the public and private interconnections (resulting in total node partitioning failure), the cluster nodes use majority quorum voting to determine the current state of the system. Both nodes and a single disk device take part in the voting. During the recovery process, both nodes attempt to reserve the quorum device. The node that wins the vote reconfigures the cluster to include itself and the storage devices. The node that loses the vote is not automatically made a member of the cluster; operator intervention is required to bring that node back into the cluster.

A quorum disk is assigned during cluster software installation and configuration. The quorum disk must be configured as a shared disk to be visible to all nodes, it must have the same device serial number as seen by all hosts, and it must not be used to store data. This requires that the Symmetrix Common Serial Number Bit be enabled to present the same device serial number to all cluster nodes.

About the terminal concentrator

The SCSI reservation mechanism used to lock the quorum disk does not work well in a four-node cluster. Instead, Sun Cluster implements the cluster-wide locking mechanism using the terminal concentrator (TC) port.

During the Cluster software installation, a TC port is specified for use as the cluster locking port. The node that successfully locks the terminal server port during node partition failure will stay up. Other nodes that fail to lock the TC port remove themselves from the cluster. Thus, for a four node Sun Cluster, a terminal server is required hardware.

About the Sun Cluster configuration database

Sun Cluster maintains multiple copies of the Sun Cluster configuration database (CCD) in case of failure. The implementation of CCD in Sun Cluster for two node and greater than two-node (up to four-node) clusters differs as discussed below.

Clusters with two nodes

Two shared disks must be assigned to be visible on both nodes. Sun Cluster HA creates a diskgroup called SC_DG and creates a mirrored volume across these two disks with a UFS file system. The CCD database is then put on this mirrored volume.

When both cluster nodes are up, the CCD daemons running on each node coordinate with each other and allow changes to the CCD database. When there is a node partition or one node is down, the remaining node is allowed to change the CCD database only if it has a quorum majority vote. This vote consists of the ownership (reservation) of the quorum disk and the access to the CCD disk, which is visible to both nodes since the disk is a shared disk.

During Symmetrix configuration, two 40-cylinder disks must be configured for use as CCD database disks. Each disk must be assigned to different Symmetrix channel director ports for host-level mirroring, which provides CCD access continuity and high availability in case of single-path failure.

Clusters with three or four nodes

In clusters of more than two nodes, the CCD is stored on the internal boot disk of each node; therefore, no shared CCD disk is required.

Note: A shared CCD disk should be assigned, however, just in case the Sun Cluster CCD implementation is changed in the future.

The CCD daemon running on each node maintains the consistency between both copies of the CCD database (one on each node). Changes to the CCD database require the quorum majority vote.

A quorum majority vote can take place only if greater than half of the number of nodes in the cluster are operational and the CCD database checksum is the same on all of those nodes.

Sun Cluster 2.x HA setup considerations

Before you install a Sun Cluster 2.x high availability (HA) cluster, you must plan effectively.

This section provides a brief description of the steps involved in planning.

Note: Refer to the *Sun Cluster Software Installation Guide* and the *Sun Cluster System Administration Guide* for detailed information.

Step 1: Select a cluster configuration

Choose a configuration with two to four nodes. Sun Cluster v2.0/2.1/2.2 uses VxVM (SEVM or SSVM), which permits a two-node configuration. Sun Cluster v2.2 uses SDS, which permits a four-node configuration.

Step 2: Plan the network connections

There must be at least one public and exactly two private network connections. Choose a hardware platform and public network configuration that includes two private heartbeat interfaces and two public network interfaces, for redundancy. The two public network interfaces will allow for IP network interface failover if an IP network adapter fails or if a network cable is disconnected.

Step 3: Choose logical host addresses and names

Logical hosts are data service entities that fail over between cluster nodes. A logical host consists of a virtual IP address (IP alias) and a logical host name. You define the logical host IP address and logical host name in the `/etc/hosts` file.

Note: For details on defining logical host addresses and names, see the *Sun Cluster System Administration Guide.*

A node can own (service) multiple logical hosts. The client accesses specific data services (for example, NFS or Oracle) using the logical host name assigned to the specific data service.

Step 4: Set up the hardware

Setting up your hardware consists of installing network connections, all local and multi-host disks, and any other optional hardware.

For an HA cluster with more than two nodes, you must set up an appropriate terminal concentrator, which is required for cluster-wide node locking.

Note: For details, refer to the hardware planning and installation guides for the HA servers you are using.

Sun Cluster 2.x parallel database setup considerations

Before you install a Sun Cluster 2.x parallel database cluster, you must plan effectively.

This section provides a brief description of the steps involved in planning.

Note: Refer to the *Sun Cluster Software Installation Guide* and the *Sun Cluster System Administration Guide* for detailed information.

Step 1: Select a cluster configuration

You must choose a configuration with either two nodes or more than two nodes (up to 4). In a parallel database configuration, Sun Cluster v2.0/2.1/2.2 uses VxVM 3.0.4 or Cluster Volume Manager (CVM). CVM is actually VxVM 2.2-1 with additional functionality for management of shared disk groups.

Step 2: Set up the hardware

Setting up your hardware configuration involves installing your network connections, all local and multi-host disks, and any other optional hardware. For clusters with more than two nodes, you must set up the Terminal Concentrator required for cluster-wide node locking.

Note: Refer to the hardware planning and installation guides for your servers for details on setting up your hardware.

Incorporating Symmetrix into Sun Cluster 2.x environment

This section provides basic information about incorporating the Symmetrix system into the Sun Cluster 2.x environment.

Introduction

Incorporating the Symmetrix system into the Sun Cluster 2.x environment requires the following steps:

1. Set these Symmetrix channel director bits so the Symmetrix system can present shared devices to Sun Cluster nodes:
   - Common Serial Number bit (C bit)
   - Sun Cluster bit (SCL bit)
2. Connect the multipath cables from the Symmetrix system to the Sun Enterprise Cluster nodes.
3. Incorporate the Symmetrix volumes into the Solaris / Sun Cluster cluster nodes.
4. Bring the Symmetrix volumes under CVM control.

Software requirements

SunCluster nodes must run Solaris 2.5.1, Sun Enterprise Cluster 2.0 or higher, Oracle7 Server version 7.3.2.2.2 or higher, and VxVM 3.0.4 or CVM 2.2.1. The minimum Symmetrix microcode levels are 5063.30.23 (Symm-3) or 5263.30.23 (Symm-4). Refer to Table 3 on page 78.

Symmetrix/Sun Cluster 2.x setup

Successful configuration of the Symmetrix system in the Sun Cluster 2.x environment requires the appropriate Symmetrix microcode and activation of the common serial number and Sun Cluster director settings.

Performing the following actions will present the appropriate serial numbers for Symmetrix devices to the cluster nodes. This is necessary for the configuration of the quorum device (as discussed earlier) and for proper functioning of the shared cluster devices.

CAUTION

The following procedure requires access to the Symmetrix service processor; therefore, the procedure must be performed by a qualified EMC service representative.

1. On the Symmetrix AutoInstall **Edit Director Information** screen, set the C bit for each Fibre Channel and SCSI director. The C bit is identified by the string `Common Serial Number` in the online help area.
2. Set the SCL bit for each Fibre Channel and SCSI director. The SCL bit is identified by the string `Enable Sunapee` in the online help area.
3. Run the EMC **`inq`** (inquiry) utility for the disks to ensure that the proper common serial number is returned from both hosts. To run the utility, type **`inq`** and press ENTER.

   Check the **Serial Number** field in the output for the common serial number. Do this for all nodes connected to the Symmetrix system. For devices that are shared by multiple nodes, the same serial number should appear on each node.

   Output example:

```
Inquiry utility, Version 7.04   (SIL Version 4.1.2)
Copyright (C) by EMC Corporation, all rights reserved.
For help type inq -h.

-------------------------------------------------------------------------------
DEVICE                :VEND   :PROD               :REV   :SER NUM   :CAP(kb)
-------------------------------------------------------------------------------
/dev/rdsk/c4t0d0s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46000000  :4418880
/dev/rdsk/c4t0d1s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46001000  :4418880
/dev/rdsk/c4t0d2s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46002000  :4418880
/dev/rdsk/c4t0d3s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46003000  :4418880
/dev/rdsk/c4t0d4s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46004000  :4418880
/dev/rdsk/c4t0d5s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46005000  :4418880
/dev/rdsk/c4t0d6s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46006000  :4418880
/dev/rdsk/c4t0d7s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46007000  :4418880
/dev/rdsk/c4t1d0s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46008000  :4418880
/dev/rdsk/c4t1d1s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :46009000  :4418880
/dev/rdsk/c4t1d2s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :4600A000  :4418880
/dev/rdsk/c4t1d3s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :4600B000  :4418880
/dev/rdsk/c4t1d4s2   :EMC   :SYMMETRIX-SUNAPE    :5265  :4600C000  :4418880
```

Defining the quorum device

A quorum device is a disk device that is visible to all nodes and is used for cluster voting. The quorum device is defined by default at Sun Cluster install time.

If you need to define a Symmetrix device as a quorum device after install time, use the following procedure:

1. Enter the following command through the cconsole and crlogin GUI interfaces from the ccp toolbar:

```
/opt/SUNWcluster/bin/scconf cluster_name -q host_1 host_2
```

   The **-q** option probes the list of devices attached to each host and lists the devices the two hosts share. The quorum device can then be selected from this list. Note that this command should be run from all nodes at the same time through the cconsole and crlogin GUI interfaces from the ccp toolbar.

   Output example:

```
Select quorum device for nodes 0 (host_1) and 1 (host_2).
Type the number corresponding to the desired selection.
For example: 1<CR>
1) DISK    :c2t2d0s2   :17100000
2) DISK    :c2t2d1s2   :17101000
3) DISK    :c2t2d2s2   :17102000
4) DISK    :c2t2d3s2   :17103000
5) DISK    :c2t2d4s2   :17104000
6) DISK    :c2t2d5s2   :17105000
7) DISK    :c2t2d6s2   :17106000
Quorum device:
```

2. Select the quorum device from the list.

Bringing Symmetrix devices under VxVM or CVM control

The following example illustrates methods for creating shared volumes for running Oracle Parallel Server (OPS) software on a Sun Cluster system.

1. Initialize all of the disks that are to be managed by CVM. You can use either the vxva tool (the graphical user interface, or GUI) or the command line interface (CLI).

For the CLI, the general format of the command is as follows:

```
/etc/vx/bin/vxdisksetup -i devname
```

where ***devname*** consists of the controller, target, and disk numbers; for example, **c4t0d0**.

Repeat this command for each disk to be placed under CVM control.

2. Bring up the Sun Cluster software on the first node. This machine becomes the cluster master. (The Sun Cluster software runs on only one node.) Issue the following command:

```
/opt/SUNWcluster/bin/scadmin startcluster node_name cluster_name
```

   After issuing this command, there should be only one master node in the cluster.

3. Create a shared disk group by issuing the following command:

```
vxdg -s init demo c4t0d0 c5t0d0
```

4. Check the results by issuing the `vxdisk` command; for example:

```
vxdisk list
```

   This command generates a detailed list of all disk groups with a separate entry for each disk.

Creating a mirrored, striped volume with VxVM or CVM

Use the following procedures to create a mirrored, striped volume using VxVM 3.0.4 or Cluster Volume Manager (CVM). This example creates a raw volume intended for use with Oracle Parallel Server.

To avoid a potential single point of failure, EMC recommends using VERITAS host-based mirroring for this configuration.

Note: In configurations using EMC PowerPath, host-based mirroring is not required; mirroring is done at the Symmetrix system level.

1. Create subdisks issuing a command similar to the following:

```
vxmake -g oracledg sd c4t0d0-01 dm_name=c4t0d0 len=500
```

   In this example, **oracledg** is the name of disk group, **c4t0d0-01** is the name of the subdisk, and **500** is the size of the volume in megabytes.

2. Create a striped plex issuing a command similar to the following (all on one line):

```
vxmake -g oracledg plex system-pl layout=stripe st_width=128 ncolumn=3
   sd=c4t0d0-01,c5t0d0-01,c5t0d0-01
```

   In this example, **`system-pl`** is the name of the plex; **`st_width`** is the width of the stripe; and **`c4t0d0-01`**, **`c5t0d0-01`**, and **`c6t0d0-01`** are the subdisks that make up the plex.

3. Create a volume from the striped plex by issuing a command similar to the following:

```
vxmake -U gen -g oracledg vol system-vol plex=system-pl
```

   In this example, **`oracledg`** is the name of disk group, **`system-vol`** is the name of volume, **`sys-tem-pl`** is the name of plex that makes up volume, and **`gen`** is the raw device.

4. Add a mirror to a volume by issuing the following command:

```
vxassist -g oracledg mirror system-vol layout=stripe
```

   Other values, such as the size of the mirror, default to the size defined in the previous command where the volume was defined.

Creating a new cluster-shared disk group

The **vxdg** utility manages Volume Manager disk groups. vxdg determines whether a disk group is cluster-shareable. The **-s** option, when used with vxdg, initializes or imports a disk group as shared.

Scenario 1

If you have already set up the cluster and now want to create a cluster-shared disk group, issue a command similar to the following:

```
vxdg -s init <disk_group_name> c4t0d0 c4t1d0 c5t0d0 c5t1d0
```

Scenario 2

If you set up the disk groups before running the cluster software, and you now want to import the disk groups into the cluster arrangement, issue a command similar to the following:

```
vxdg -s import <disk_group_name>
```

Note: The system cannot tell if a disk is shared. To protect the integrity of disks that are accessed by multiple systems, be sure to use the correct designation when adding a disk to a disk group. In addition, the entire configuration should be managed from only one node to avoid confusion and operational problems.

To display information about disk groups, issue the following command:

```
vxdg list
```

The output is similar to the following:

```
NAME          STATE              ID
rootdg        enabled            855087068.1025.lss6128
demo          enabled,shared     855959322.1418.lss6128
oracledg      enabled,shared     855252495.1247.lss6128
testdg2       enabled,shared     858177251.1571.lss6128
testdg3       enabled,shared     858267086.1690.lss6128
testdg4       enabled,shared     858288846.1785.lss6128
```

To display detailed information about disk groups, issue a command similar to the following:

```
vxdg list <disk_group_name>
```

Sun Cluster administration examples

This section provides examples of activities you might need to perform to administer Symmetrix in a Sun Cluster.

CAUTION

Before managing a Sun Cluster system, be sure to read the release notes and documentation that accompany the software. In addition, EMC strongly recommends that you complete a Sun Cluster training course before attempting any configuration procedures.

Note: For detailed guidelines on Sun Cluster administration, refer to the *Sun Cluster System Administration Guide.*

Introducing Symmetrix devices to Sun Cluster hosts

Use the format command to partition and label the new devices. The devices will appear under `/dev/dsk`.

To partition and label new devices:

1. At the `root` prompt, enter the following command:

   ```
   format
   ```

   The host searches for and displays all the disks.

2. At the `Specify Disk` prompt, enter the number of the first EMC drive.

   The **Format** menu appears.

3. At the `Format` prompt, enter the following command:

   ```
   label
   ```

4. At the `Disk not labeled. Label it now?` prompt, type **`y`** and press ENTER.

5. At the prompt to display a listing of the disks, type **`di`** and press ENTER.

6. Repeat steps 2 through 5 for all Symmetrix disks.

Adding devices online

Whenever devices are added online to the Symmetrix/Sun Cluster environment, or device channel addresses are changed, an EMC Customer Engineer performes these necessary steps:

1. Adds/maps new drives to the Fibre Channel/SCSI host channels using the online upgrade feature via the Symmetrix service processor.
2. Runs the `drvconfig` and `disks` host utilities.

Useful Sun Cluster commands and utilities

A Sun Cluster is managed and monitored via command line utilities. This section provides a brief description of some useful commands and utilities used in the Sun Cluster environment.

Note: Refer to the *Sun Cluster System Administration Guide* or man pages for a complete description of the command arguments.

Commands used with any configuration

The following common Sun Cluster administrative commands can be used with any Sun Cluster configuration.

scinstall

The **`scinstall`** command provides an interface that allows you to perform the initial configuration of the Sun Cluster systems. You must load Sun Cluster software onto each node and enter the information identically across all nodes.

scadmin startcluster

The **`scadmin startcluster`** command starts the Sun Cluster software framework on the node from which it is executed (the command cannot be run remotely). Use this command to start the first node of the cluster. The first node to join the cluster becomes the master. Issue this command only if no other nodes are currently in the cluster.

scadmin startnode/stopnode

Use the **`scadmin startnode`** command to add nodes to a cluster. In order to use this command, the cluster must already be operational on at least one node. The server must be a configured node of the cluster before it can join the cluster. Use the **`scadmin stopnode`** command to remove a node from an operational cluster. Stop all parallel applications before removing a node from the cluster. You can add and remove nodes simultaneously.

ccdadm

The **`ccdadm`** command launches a utility that provides administrative services to the CCD. Only a super user can issue this command. Refer to the `ccdadm` man page for functions, parameters, and restrictions.

scconf

The **`scconf`** command allows an administrator to view, create, or change the configuration of a cluster. Specifically, this command is used to configure or change the quorum device. Refer to the `scconf` man page for functions, parameters, and restrictions.

Commands used only with HA failover configurations

The following Sun Cluster commands are used only with HA failover configurations.

hastat

The **hastat** command displays the current state of the configuration. When run in default mode, the command displays the status of all components in the configuration and recent error messages from the messages file. This information is displayed once before the program exits.

haswitch

The **haswitch** command initiates a switch-over of a logical host to a destination server. In the following example, *logical_host_2* switches over to *destination_hostname*:

```
hahost1# haswitch <destination_hostname>
   <logical_host_2>
```

hafstab

The **hafstab** command allows you to edit copies of the `dfstab` and `vfstab` files. The program then performs a limited sanity check and distributes the files to all servers in the configuration.

hareg

Before a data service can provide services under Sun Cluster HA control, it must be registered using the **hareg** command. Data services can be registered and unregistered at any time.

Sun Cluster HA host states

Host system states can be viewed using either `hacheck` or `hastat`. These programs use the following terms to describe the various states of a host node:

- **Down** — The host is not currently functioning.
- **Reconfiguring** — The host is in the process of reconfiguring itself following a take-over or switch-over.
- **Stable** — The host is operating normally.
- **Unknown** — The fault probes cannot evaluate the current state of the host.

Event logging

The Sun Cluster software logs cluster events such as state changes, changes in heartbeat network status, HA event control requests, and agent messages. Separate logs are maintained for each Sun Cluster host.

Log messages are appended to the log file until the file size reaches its maximum limit. A new file is then created and the old file is saved. The messages are written to the `/var/adm/messages` file and are displayed on the console.

Depending on the specific cluster configuration, these files are located in the `/var/opt/SUNWcluster` directory, and some of its subdirectories.

6

Sun Cluster 3.x

This chapter discusses EMC Storage / Sun Cluster 3.x environment. Fundamental concepts and procedures related to Sun Cluster planning, setup, and administration are provided.

Sun Cluster 3.x overview

This section introduces Sun Cluster 3.x and briefly describes its important features and how they relate to EMC storage.

Note: Sun also refers to Sun Cluster as *Solaris Cluster.*

What is Sun Cluster 3.x?

Sun Cluster 3.x is a highly available and scalable cluster software framework that is tightly integrated with the Solaris Operating Environment. Sun Cluster 3.x is part of the SunPlex system that includes the Solaris OE, Sun Cluster 3.x, SPARC hardware and networking components. At the time of this writing EMC supports Sun Cluster 3.0 Update 3 and higher. Please refer to the *EMC Support Matrix* for EMC's latest support for Sun Cluster.

Sun Cluster 3.x enables the implementation of applications in either a failover or scalable topology or both. A failover configuration is one in which a set of resources and applications are automatically relocated to another server in the event that the primary node fails. For failover services, applications run on only a single server at any one time. In a scalable configuration, a set of resources / applications are spread across cluster servers and run concurrently on them. Service requests come into the cluster through a global network interface and are distributed to the cluster servers based one of several predefined algorithms. Sun Cluster 3.x can also be configured to run Oracle Parallel Server (OPS) or Real Application Cluster (RAC).

While Sun Cluster 3.x shares the same name as Sun Cluster 2.x, it is a completely new product written from the ground up.

All Sun Cluster 3.x documentation can be found at:

`http://docs.sun.com`

Hardware components

This section provides information on the hardware components.

Cluster nodes

A cluster node is a server that is running Solaris, Sun Cluster 3.x framework, and Sun Cluster 3.x Data Service software. Up to six (6) nodes are supported in a High Availability environment. Sun Cluster 3.x can be run on most Sun server families. Cluster nodes are connected to Symmetrix disks using both fiber channel and SCSI interfaces. Refer to the *EMC Support Matrix* for all relevant host bus adapters, drivers and switch versions. Nodes that are not physically attached to the storage, but participating in cluster membership, can gain access to storage through the cluster file system.

Cluster members communicate with each other through a mechanism called the Cluster Membership Monitor (CMM) over a set of physically independent networks called the cluster interconnect. The cluster interconnect is discussed later in this chapter.

In general, nodes in the cluster should have similar physical resources such as processors, memory and I/O capability to be able to sufficiently run the applications and resources that may failover to them. Additional server capacity may be required in an Active-Active topology. In such configurations, all servers are primaries for one set of resources and are secondaries in the event that another server in the cluster failed. In this case, the server may need additional system resources in order to run both sets of applications.

Storage

Both the Symmetrix 8000, DMX, and FC-4700/Cx Series families are supported with Sun Cluster 3.0 and higher. Minimum Enginuity™ and FLARE® code revisions exist for both Symmetrix and CLARiiON® families. Check the *EMC Support Matrix* for code revisions and other considerations. See Figure 4 on page 111 through Figure 7 on page 114 for supported storage topologies.

Cluster interconnect

The cluster interconnect is a set of private networks that are used to carry membership and data service communications between the nodes participating in the cluster. Redundant private networks are used to avoid a single point of failure in the event that one network component should fail. Up to six networks can be configured, and Sun Cluster 3.x will exploit the additional bandwidth when available. Some cluster topologies, such as Real Application Cluster, use the cluster interconnect extensively. For these configurations, high-speed interconnect technologies should be deployed. The cluster interconnect consists of Network Interface Cards (NICs), junctions (switches/hubs), and cables.

Software components for cluster servers

The following software packages are generally installed on cluster servers:

- Solaris Operating System
- Sun Cluster 3.x framework software
- Data service applications
- Volume Manager (Solaris Volume Manager, Solstice Disk Suite, or VERITAS Volume Manager)
- Multipathing software
  - EMC PowerPath or Sun StorEdge Traffic Manager (a/k/a MPxIO)

Supported software versions for Sun Cluster 3.x

Note: Refer to the *EMC Support Matrix* for the latest information regarding supported software versions. Also, refer to the SunSolve website for the latest Solaris and Sun Cluster patch levels.

The *EMC Support Matrix* provides the supported configurations and the minimum requirements of related software.

- Solaris 8, Solaris 9, or Solaris 10 with latest patches from SunSolve website.

  Be aware that the software package for Solaris 8 can only be used in Solaris 8 environment, the software package for Solaris 9 can only be used in Solaris 9 environment, and the software package for Solaris10 can only be used in Solaris 10 environment.

  Note: Refer to the SunSolve Home website for additional information.

- Sun Cluster 3.x, with the latest Sun Cluster Core Packages, if any.
- VERITAS Volume Manager (VxVM), Solaris Volume Manager (Solaris 9), and SDS (Solaris 8).

  Refer to the latest *EMC Support Matrix* for the availabilities of other supported volume managers.

- VERITAS DMP should *not* be disabled for Sun Cluster 3.x environments with EMC PowerPath.
- EMC PowerPath 4.3 or higher is supported in all configurations unless otherwise noted in the *EMC Support Matrix.* However, PowerPath 4.4 or higher is recommended.
- The *EMC Support Matrix* lists EMC-supported CLARiiON FLARE versions.
- Oracle UDLM package:
  - RAC 9i supports Solaris Volume Manager with Sun Cluster 3.1U3 or higher on OS Solaris 9 or higher.
  - OPS/RAC 9i supports VERITAS Volume Manager with Cluster Feature (CVM) on both Solaris 8/9 or higher.
  - The *EMC Support Matrix* lists supported OPS/RAC configurations.

Sun Cluster 3.x configuration examples

The diagrams on the following pages show several possible configurations for Symmetrix and CLARiiON systems in a Sun Cluster 3.x environment. Refer to the Sun Cluster 3.x Concepts manual for additional information.

Typical configurations will include two to six (6) nodes depending on the data services in use. Some or all of the nodes may be physically connected to the Symmetrix system. The current guidelines are as follows:

- Up to six (6) nodes in an HA configuration (non-OPS/RAC) can be configured. Any number of these nodes can be physically connected to the storage. While some nodes may not be physically connected to the storage, they have access to storage through the global namespace and cluster file system features in Sun Cluster 3.x. Refer to the *Sun Cluster 3.x Concepts* manual and the *Sun Cluster 3.x Systems Administration Guide* for more information on this functionality.
- Up to four nodes are possible for OPS/RAC in Sun Cluster 3.x. All nodes are physically connected to the EMC Storage system.

Figure 4 shows a typical two-node topology for either an HA or OPS configuration. One run of Symmetrix devices is presented to four FA channels and then to all HBAs on the cluster nodes. EMC PowerPath would be deployed for this configuration with Symmetrix RAID protection.

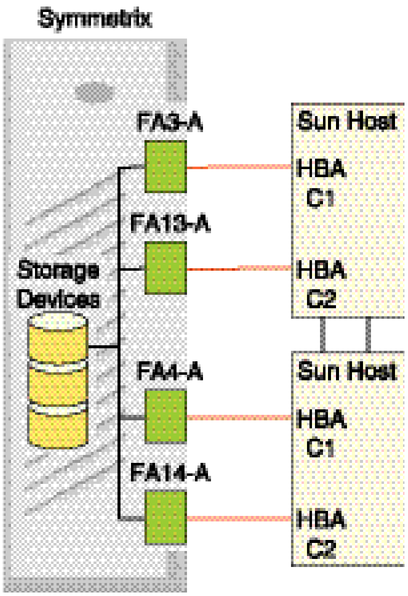


Figure 4 Typical two-node topology

Figure 5 shows a three-node HA configuration. Two nodes are physically attached to the Symmetrix system. The third node at the bottom has access to the storage through the global device namespace and cluster file system. EMC PowerPath would be deployed for this configuration with Symmetrix RAID protection.

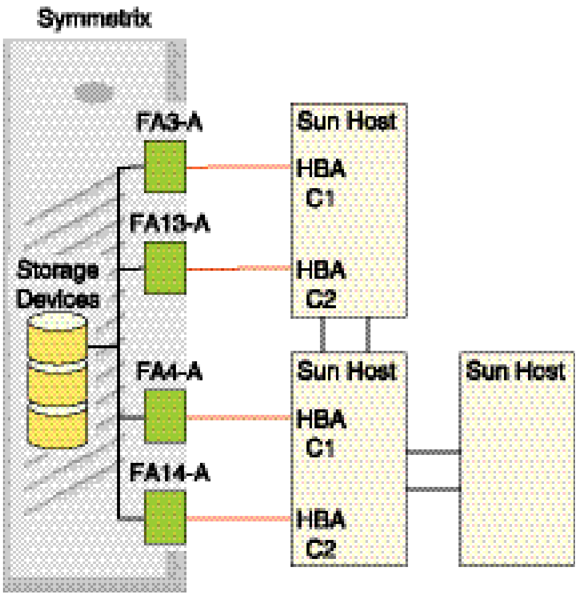


Figure 5 Three-node HA configuration

Figure 6 shows a four-node fully attached configuration. EMC PowerPath is deployed for this configuration with Symmetrix RAID protection.

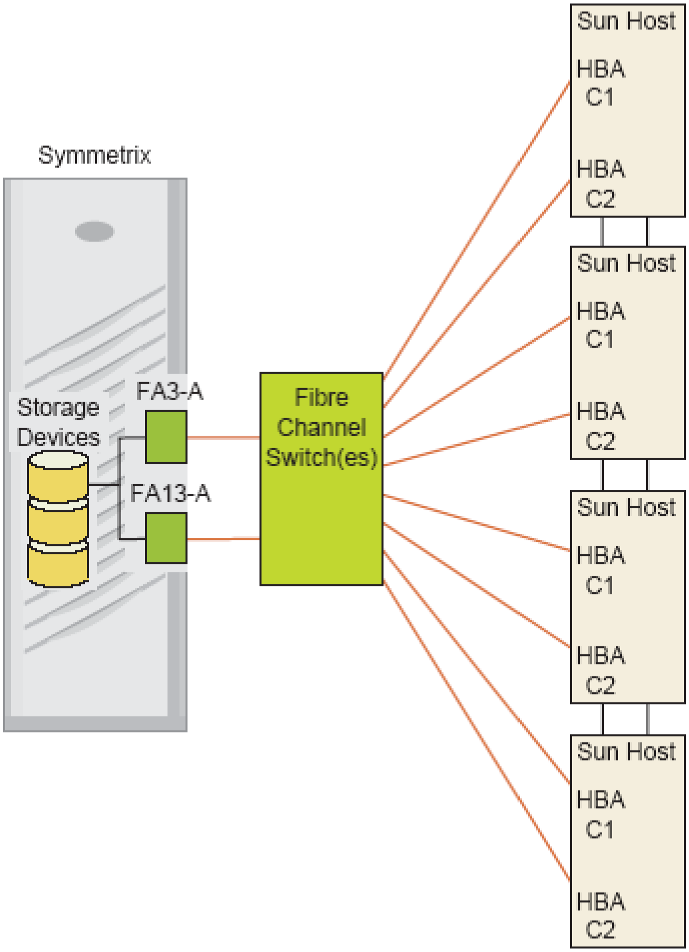


Figure 6 Four-node fully attached configuration

Figure 7 shows a two-node, host-based mirrored configuration. EMC PowerPath is *not* deployed in this configuration.

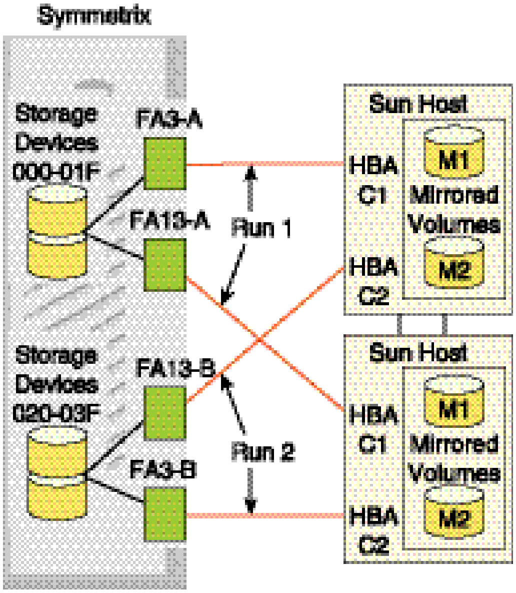


Figure 7 Two-node host-based mirrored configuration

Figure 8 on page 115 and Figure 9 on page 115 show possible configurations for an EMC CLARiiON storage system in a Sun Cluster 3.x environment. Refer to the *Sun Cluster 3.x Concepts* manual for additional information.

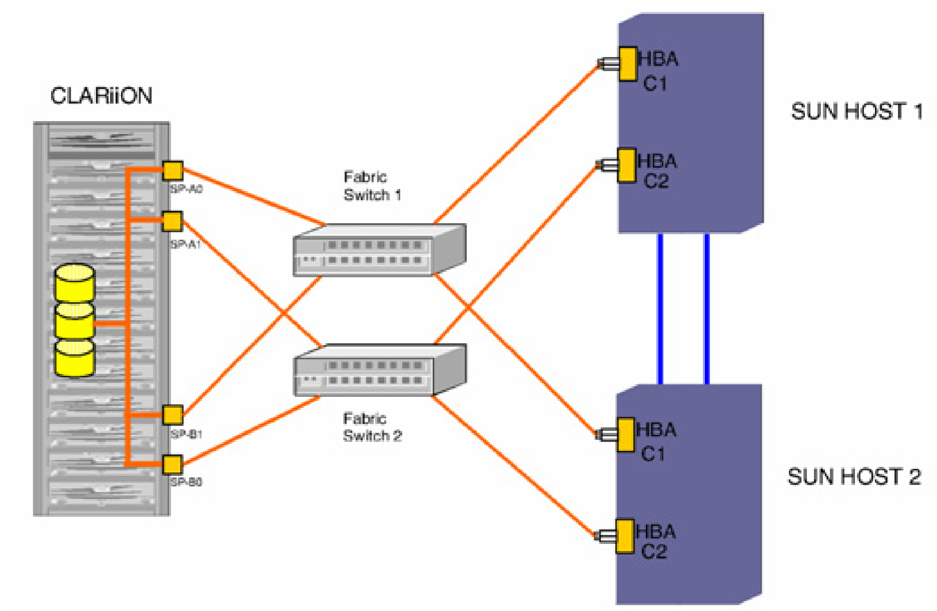


Figure 8 **Typical two-node fabric topology**

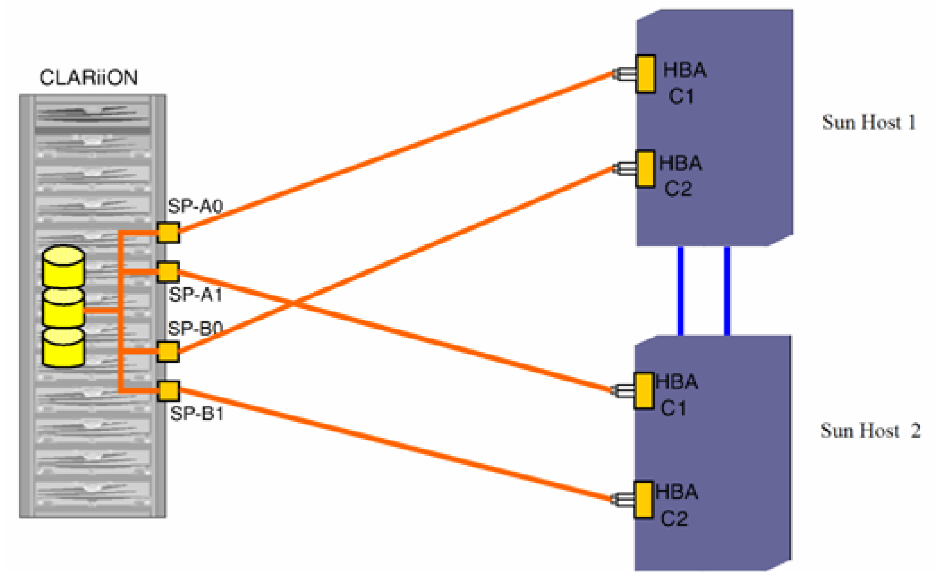


Figure 9 **Typical two-node direct-attach topology**

Key Sun Cluster 3.x concepts

This section provides information on key Sun Cluster 3.x concepts.

Cluster Membership Monitor (CMM)

The Cluster Membership Monitor (CMM) is a set of agents that use the private interconnects to communicate with the nodes that comprise a Sun Cluster. The CMM performs the following functions:

- Monitors changes in cluster membership when nodes join or leave the cluster
- Ensure that a faulty node leaves the cluster
- Ensure that a faulty node stays out of the cluster until it is repaired
- Protect the cluster from partitioning itself into multiple clusters (split-brain, amnesia)
- Verify full connectivity to all nodes in the cluster

Split-brain is a condition where in the event that all communication is lost between cluster members the cluster partitions into each node (or a subset of nodes) believes it is the only cluster node. In this condition, uncoordinated access to shared storage could result in severe data corruption.

Amnesia is a condition where a node starts with stale cluster configuration data. If a node fails and the cluster is reconfigured on the remaining nodes, then the configuration would be stale on the node that failed. If the failed node then attempts to join the cluster, it must be resynchronized with the current cluster configuration data.

Cluster Configuration Repository (CCR)

The Cluster Configuration Repository (CCR) is a cluster-wide database that stores cluster configuration and state information. Each node in the cluster maintains a copy of the CCR. The Sun Cluster framework maintains the integrity of the CCR by using a two-phase commit protocol where any update to the cluster configuration from one node must successfully complete on all nodes of the cluster.

Global devices

Sun Cluster 3.x uses a concept called global devices to present a cluster-common device view to all nodes in the cluster regardless of where specific devices are physically attached. The cluster automatically detects storage devices on system boot up and assigns unique IDs to each disk device. Devices are assigned global names, and are by default integrated into the cluster environment. This naming scheme allows for a common name to be used by all nodes in the cluster even the actual hardware path to the shared storage devices could vary from one node to the next. Sun Cluster 3.x essentially abstracts data location from data services. Data does not need to be attached to the server that hosts the services. This global namespace is held in the `/dev/global` directory on each node.

Device ID (DID)

To manage the global devices Sun Cluster 3.x uses a pseudo driver called the Device ID (DID). This driver searches for devices attached to the cluster node and automatically builds a list of unique disk devices and assigns a unique major and minor number that's consistent across the cluster. Device access is then performed using the unique device ID that was assigned by the DID instead of the traditional Solaris CxTxDx device ID.

An example of a device name using the DID naming scheme is:

```
/dev/did/dsk/d10s2
```

Global namespace

Global devices are enabled through a construct called the global namespace. The global namespace is implemented using a `/global/.devices/node@x` structure, where x is a node number within the cluster. This number can be 1 – 6, depending on the number of nodes making up the cluster. The global namespace includes the `/dev/global` directory hierarchy as well as the VxVM and SVM volume manager namespaces. All disk devices represented in `/dev/global` and all SVM metadevices and VxVM disk group volumes are symbolically linked to a `/global/.devices/node@x/dev` pseudo entry. Both the global namespace and standard volume manager namespace are available from any node in the cluster.

Table 5 list examples of local and global namespace mappings.

Table 5 Global namespace mappings

Object	Node path name	Global path
Disk	/dev/dsk/c1t0d10s2	/global/.devices/node@2/dev/dsk/c1t0d10s2
DID Name	/dev/did/dsk/d1s2	/global/.devices/node@2/dev/did/dsk/d1s2
SVM Diskset	/dev/md/test/dsk/d11	/global/.devices/node@2/dev/md/test/dsk/d11
VxVM Volume	/dev/vx/dsk/group/vol2	/global/.devices/node@2/dev/vx/dsk/group/vol2

Cluster file systems

The cluster file system feature of Sun Cluster 3.x is a proxy between the operating system kernel and the underlying file system. It allows for the access of file systems regardless of physical location within the cluster. Cluster file systems are dependent on global devices and can be accessed by any node in the cluster through a common name whether or not that node is physically connected to the storage device.

Cluster file systems enable the ability of mounting file systems on a node that is not physically attached to the storage system.

Cluster file systems are mounted using the **`mount -g`** command. They can also be mounted automatically with the `/etc/vfstab` file. See the section later in this chapter for an example of automatically mounting cluster file systems.

Sun Cluster 3.x data services

Sun Cluster 3.x includes agents to make applications highly available through the use of start, stop and monitoring scripts and programs. See the section on configuring the NFS Data Service later in this chapter for an example of Sun Cluster 3.x data service setup. Data services are installed after the Sun Cluster 3.x framework is installed.

Available data services for Sun Cluster 3.x are as follows:

- Sun Cluster HA for Apache
- Sun Cluster HA for Apache Tomcat
- Sun Cluster HA for BroadVision One-To-One Enterprise

- Sun Cluster HA for DHCP
- Sun Cluster HA for DNS
- Sun Cluster HA for MySQL
- Sun Cluster HA for NetBackup
- Sun Cluster HA for NFS
- Sun Cluster HA for Oracle E-Business Suite
- Sun Cluster HA for Oracle
- Sun Cluster Support for Oracle Parallel Server / Real Application Clusters
- Sun Cluster HA for SAP
- Sun Cluster HA for SAP liveCache
- Sun Cluster HA for SWIFTAlliance Access
- Sun Cluster HA for Samba
- Sun Cluster HA for Siebel
- Sun Cluster HA for Sun ONE Application Server
- Sun Cluster HA for Sun ONE Directory Server
- Sun Cluster HA for Sun ONE Message Queue
- Sun Cluster HA for Sun ONE Web Server
- Sun Cluster HA for Sybase ASE
- Sun Cluster HA for WebLogic Server
- Sun Cluster HA for WebSphere MQ
- Sun Cluster HA for WebSphere MQ Integrator

Resource groups

Resource groups are the logical constructs that Sun Cluster 3.x uses to group resources together so they can be managed and made highly available. A resource group is migrated from one node to another in the event of a cluster node failover or initiated switchover. Resource groups can contain data services, disk device groups, network interfaces or other resources. Dependencies can be defined for resource groups to assure that the group cannot be brought up unless all of it underlying resources are available.

Quorum and failure fencing

With any cluster it is important to protect the disk resource from the possibility of having uncoordinated cluster members from writing to shared storage devices and possibly corrupting data. Sun Cluster 3.x uses the CCM to protect the cluster from partitioning into multiple clusters in the event that the cluster interconnects fail. The specific types of failures were discussed above in the CMM section.

Sun Cluster uses a quorum vote algorithm where each node is assigned one vote. In order for a cluster to be operational it must have a majority of votes. If the cluster interconnects or node fails, the partition with a majority of votes will remain operational. This model works well with clusters with more than two nodes. In the case of a two node cluster where the vote majority is two and the partitioned nodes could not achieve a majority of votes. Sun Cluster 3.x solves this by assigning an external vote to a quorum device. The quorum device can be any disk device that is shared between two or more nodes. EMC Symmetrix/CLARiiON devices are commonly used for this purpose. It is recommended that an external quorum device be configured regardless of node count. An example of configuring a quorum device on Symmetrix is provided later in the chapter.

Important Sun Cluster 3.x utilities

This section provides information on important Sun Cluster 3.x utilities.

scinstall

The **`scinstall`** command performs Sun Cluster node initialization, installation, and upgrade tasks. It can be run as an interactive utility or by the command line.

The **`scinstall`** command installs and initializes a node as a new Sun Cluster member. It either establishes the first node in a new cluster or adds a node to an already-existing cluster. It can also be used to remove cluster configuration information and uninstall Sun Cluster software from a cluster node.

The upgrade form (-u) of **`scinstall`**, which has several modes and options, upgrades a Sun Cluster node. Always run this form of the scinstall command from the node being upgraded.

The print release form (-p) of **`scinstall`** prints release and package versioning information for the Sun Cluster software installed on the node from which the command is run.

Without options, the **`scinstall`** command attempts to run in interactive mode. Run all forms of the **`scinstall`** command other than the print release form (-p) as superuser.

The **`scinstall`** command is located in the Tools directory on the Sun Cluster CD-ROM. If the Sun Cluster CD-ROM is copied to a local disk, *cdrom-mnt-pt* is the path to the copied Sun Cluster CD-ROM image. The SUNWscu software package also includes a copy of the scinstall command.

scsetup

At post-install time, the **`scsetup`** utility performs initial setup tasks, such as configuring quorum devices and resetting *installmode.* Always run the **`scsetup`** utility just after the cluster is installed and all of the nodes have joined for the first time.

After *installmode* is disabled, scsetup provides a menu-driven front end to most ongoing cluster administration tasks.

You can execute **scsetup** from any node in the cluster. However, when installing a cluster for the first time, it is important to wait until all nodes have joined the cluster before running scsetup and resetting *installmode*.

The **scsetup** interactive utility can be used to configure cluster quorum devices.

scconf

The **scconf** utility manages the Sun Cluster software configuration. You can use **scconf** to add items to the configuration, to change properties of previously configured items, and to remove items from the configuration. **scconf** can be used to configure the cluster quorum device. In each of these three forms of the command, options are processed in the order in which they are typed on the command line. All updates associated with each option must complete successfully before the next option is considered.

The **scconf** command can only be run from an active cluster node. As long as the node is active in the cluster, it makes no difference which node is used to run the command. The results of running the command are always the same, regardless of the node used.

The `-p` option of scconf enables you to print a listing of the current configuration.

scdidadm

The **scdidadm** utility is used to administer the device identifier (DID) pseudo device driver.

The **scdidadm** utility performs the following primary operations:

- Creates driver configuration files
- Modifies entries in the file
- Loads the current configuration into the kernel
- Lists the mapping between device entries and did driver instance numbers

The startup script `/etc/init.d/bootcluster` uses the **scdidadm** utility to initialize the did driver. You can also use **scdidadm** to update or query the current device mapping between the devices present and the corresponding device identifiers and did driver instance numbers.

scgdevs

The **scgdevs** utility manages the global device namespace. The global device namespace is mounted under `/global` and consists of a set of logical links to physical devices. As `/dev/global` is visible to each node of the cluster, each physical device is visible across the cluster. This fact means that any disk, tape, or CD-ROM that is added to the global devices namespace can be accessed from any node in the cluster. The **scgdevs** command allows the administrator to attach new global devices (for example, tape drives, CD-ROM drives, and disk drives) to the global devices namespace without requiring a system reboot.

The **drvconfig** and **devlinks** commands must be executed prior to running **scgdevs**.

Alternatively, a reconfiguration reboot can be used to rebuild the global namespace and attach new global devices. **scgdevs** must be run from a node that is a current cluster member. If this script is run from a node that is not a cluster member, the script exits with an error code and leaves the system state unchanged.

scstat

The **scstat** utility displays the current state of Sun Cluster and its components. Only one instance of the **scstat** utility needs to run on any machine in the Sun Cluster configuration.

When run without any options, **scstat** displays the status for all components of the cluster. This display includes the following information:

- A list of cluster members
- The status of each cluster member
- The status of resource groups and resources
- The status of every path on the cluster interconnect
- The status of every disk device group
- The status of every quorum device

- The status of every Internet Protocol Network Multipathing group and public network adapter

scswitch

The **scswitch** utility is used to move resource groups or disk device groups from one node to another. It also evacuates all resource groups and disk device groups from a node by moving ownership elsewhere, brings resource groups or disk device groups offline and online, enables or disables resources, switches resource groups to or from an unmanaged state, or clears error flags on resource groups.

You can run the **scswitch** utility from any node in a Sun Cluster configuration. If a device group is offline, you can use **scswitch** to bring the device group online onto any host in the node list. However, after the device group is online, a switchover to a spare node is not permitted. Only one invocation of **scswitch** at a time is permitted.

Do not attempt to kill an **scswitch** operation that is already underway.

scshutdown

The **scshutdown** utility shuts down an entire cluster in an orderly fashion. Before starting the shutdown, **scshutdown** sends a warning message and then a final message asking for confirmation. Only run the **scshutdown** command from one node. The **scshutdown** performs the following actions when it shuts down a cluster:

- Changes all functioning resource groups on the cluster to an offline state. If any transitions fail, **scshutdown** does not complete and displays an error message.
- Unmounts all cluster file systems. If any unmounts fail, **scshutdown** does not complete and displays an error message.
- Shuts down all active device services. If any transition of a device fails, **scshutdown** does not complete and displays an error message.
- Runs `/usr/sbin/init 0` on all nodes and brings them to the **OK>** prompt.

For detailed instructions on how to use these Sun Cluster 3.z utilities refer to the *Sun Cluster 3.1 Reference Manual P/N 817–0522–10*

Configuring EMC Symmetrix with Sun Cluster 3.x

This section provides information on configuring EMC Symmetrix with Sun Cluster 3.x.

Symmetrix setup for Sun Cluster 3.x

The following settings are required for proper operation EMC Symmetrix within the Sun Cluster 3.x environment:

Sun Cluster 3.x uses SCSI-3 PGR (persistent group reservation) for storage devices that are accessible through more than two paths. Symmetrix systems support this functionality using the **PER** setting on the SymmWin **Edit Volumes** dialog. This flag must be set for all devices that will be presented to the Sun Cluster nodes.

Note: The **SCL** director flag must be OFF. The **SC3** director flag is not required for Sun Cluster 3.x.

Follow these steps to set up a Symmetrix system:

1. Configure the Symmetrix system for operation in the Sun Solaris operating system environment. Refer to the *EMC Support Matrix* for details. Verify that the Symmetrix system is running an appropriate version of the Enginuity operating environment.
2. Set the **C** (Common Serial Number) director flag for all FA/SA ports to be seen by the Sun Cluster 3.x nodes. This feature is accessed through the SymmWin **Edit Directors** screen.
3. Set the **PER** flag for all volumes that will be presented to the Sun Cluster 3.x nodes. This feature is accessed through the SymmWin **Edit Volumes** screen. The **PER** flag must also be set for data volumes and quorum devices. It is not needed for gatekeepers and VCM database volumes.

FA port sharing

Multiple Sun Clusters can share the same FA ports on a Symmetrix. In addition, Symmetrix FA ports can be shared between Sun Cluster 3.x nodes and non-clustered Solaris nodes. This feature is enabled through the use of the *EMC Solutions Enabler Symmetrix Device Masking CLI Product Guide.*

Configuring EMC CLARiiON with Sun Cluster 3.x

Setup and configuration of EMC CLARiiON in the Sun Cluster 3.x environment can be set up in fabric or loop mode depending on requirements. Multipathing software (EMC PowerPath or Sun's MPxIO) is required for HA configurations. You must configure a minimum of two paths to each CLARiiON device. Refer to the *EMC PowerPath for UNIX Release Notes* for details on driver and device configuration.

Sun Cluster 3 uses SCSI-3 PGR (Persistent Group Reservation) for storage devices that are accessible through more than two paths. This is either from a single node or multiple nodes. Figure 8 on page 115 shows an example of multiple nodes. EMC CLARiiON supports this functionality by deploying EMC PowerPath on all cluster nodes.

Installation guidelines

Verify that the CLARiiON storage system is running an appropriate version of FLARE firmware. Configure the CLARiiON storage system for operation in the Sun Solaris environment. For example, set the following settings on CLARiiON Array:

systemtype to 3
failovermode to 1
arraycommpath to 1
unitserialnumber to Array/LUN

You must set unitserialnumber = LUN on the CLARiiON array for Solaris 8 OS with Sun Cluster 3.x .

For Solaris 9 or higher, use the default setting (unitserialnumber = Array).

Sun Cluster 3.x servers

This section provides guidelines for a new installation. Refer to the *Sun Cluster 3.x Installation Guide* for additional details.

IMPORTANT

If you are using VxVM4.0 or 4.0MPx, you must also follow the procedure after step 7.

1. Install Sun Solaris software with latest patch set from SunSolve.
2. Make sure that there is at least 512 MB of available space on the local disk for the /globaldevices partition used for the global device namespace.
3. Install and configure HBA driver and /kernel/drv/sd.conf (for Emulex and QLogix drivers only). If using a third-party HBA, refer to the *EMC Support Matrix* for supported server and HBA combinations. Refer to the *Fibre Channel PCI and SBus HBA and Driver for Solaris Installation Guide* for details on driver and device configuration.
4. Install EMC PowerPath software and any required patches. Refer to the *PowerPath for UNIX Installation and Administration Guide* for details.
5. On all nodes, use the scinstall program to install the Sun Cluster 3.x framework software (included on the CD Distribution) and any Sun Cluster Core Packages patches.
6. After the cluster nodes have rebooted, use the /opt/cluster/bin/scsetup utility to reset the cluster installmode and configure a quorum device.
7. Install the latest Sun Cluster Patches from SunSolve website.
8. Install the volume manager and required patch(s) (if any) on all cluster nodes.

IMPORTANT

If using VxVM VxVM4.0 or 4.0MPx, you must use the following procedure.

Use this procedure to install VxVM version 4.0 or 4.0MPx, and PowerPath 4.3 or higher for Solaris.

a. If you are installing on a host connected only to a CLARiiON storage system, configure DMP such that all paths to the same device are grouped under the same DMP node. Run the following command after PowerPath has been installed:

```
vxddladm addjbod vid=DGC pagecode=0x83 offset=8
   length=16
```

This command needs to be run only once. You must reboot the system for it to take effect.

b. To verify that the configuration is correct, run the following commands and check the output:

```
# vxddladm listjbod VID PID Opcode Page Code Page Offset SNO length
   ================================================================
DGC ALL PIDs 18 131 8 16
```

In the example below, note the PATHS column now displays more than one path to each CLARiiON device (in this example there are four paths to each CLARiiON device).

```
# vxdmpadm getdmpnode enclosure=Disk
NAME STATE ENCLR-TYPE PATHS ENBL DSBL ENCLR-NAME
   ========================================================
c5t0d0s2 ENABLED Disk 4 4 0 Disk
c5t0d1s2 ENABLED Disk 4 4 0 Disk
c5t0d2s2 ENABLED Disk 4 4 0 Disk
c5t0d3s2 ENABLED Disk 4 4 0 Disk
c5t0d4s2 ENABLED Disk 4 4 0 Disk
c5t0d5s2 ENABLED Disk 4 4 0 Disk
```

Should PowerPath be removed from the host, DMP must be configured to its default state for CLARiiON devices by running the following command and rebooting: vxddladm rmjbod vid=DGC

9. Install any required cluster data services. Refer to the *Sun Cluster 3.x Data Services Installation Guide* for details.

Examples

This section provides examples which may be helpful.

Setting up a Sun Cluster 3.x quorum device on EMC storage

A quorum device can be configured either using the interactive **`scsetup/clsetup (SC3.2 only)`** utility or with the **`scconf/clquorum (SC3.2 only)`** command line utility. Examples of both procedures are given below.

Using the **`scconf`** command:

```
# /usr/cluster/bin/scconf -a -q globaldev=d12
```

where:

`d12` is the global device number.

A list of available global devices can be generated by using the `/usr/cluster/bin/scdidadm -L` command.

Or (for SC3.2 only):

```
# /usr/cluster/bin/clquorum add d12
```

Using the scsetup interactive utility:

1. Enter the scsetup utility:

   ```
   # /usr/cluster/bin/scsetup
   ```

   Or (for SC3.2 only):

   ```
   # /usr/cluster/bin/clsetup
   ```

   The main menu is displayed.

2. Select option 1 (Quorum) from the main menu.

   The Quorum Menu is displayed.

3. Select option 1 from the Quorum Menu **Add a quorum device**.

4. Follow the interactive instructions, and type in the global device number of the device to be used as the quorum device.

5. Verify that the quorum device has been added and is online with either of the following commands:

   ```
   # /usr/cluster/bin/scstat -q
   ```

Or (for SC3.2 only):

```
# /usr/cluster/bin/clquorum status
```

How to create a cluster file system

This procedure assumes that all components of the cluster are installed and configured (Solaris, Sun Cluster 3.x framework and VERITAS Volume Manager).

1. Create and register a VERITAS Volume Manager (VxVM) disk group.

   a. To create a VERITAS disk group:

   ```
   # vxdg init <dgname> c1t1d1,c1t1d2,c1t1d3... etc.
   ```

   where:

   <dgname> = name of the disk group to be created

   b. To register the disk group with the Sun Cluster 3.x framework:

   ```
   # /usr/cluster/bin/scsetup
   ```

2. Select **Device groups and volumes.**
3. Select **Register a VxVM disk group as a device group**.
4. Follow the prompts to register the newly created disk group.
5. Create your volumes, synchronize, and newfs them

   a. To create a VERITAS volume within the newly created disk group:

```
# vxassist -g <dgname> -U fsgen make <volname> <vol-size> <diskname> (or use vmsa)
```

   where:

   <dgname> = name of the disk group

   <volname> = name of the volume

   <volsize> = size of the volume

   <diskname> = name of the disk to use for the volume

   b. After volumes are created, the disk group needs to be synchronized. This is also the case when any volume changes are made to a disk group.

   ```
   # /usr/cluster/bin/scsetup
   ```

Or (for SC3.2 only):

```
# /usr/cluster/bin/clsetup
```

- Select **Device groups and volumes**.
- Select **Synchronize volume information for a VxVM device group**.

c. newfs your volumes:

for ufs:

```
newfs /dev/vx/rdsk/<dgname>/<volname>
```

for vxfs:

```
mkfs -F vxfs /dev/vx/rdsk/<dgname>/<volname>
```

6. Mount volumes:

```
mkdir /global/<mnt> ON ALL NODES5.
```

7. Add entries to vfstab ON ALL NODES THAT ARE DIRECTLY ATTACHED

for ufs:

```
/dev/vx/dsk/<dgname>/<volname>
/dev/vx/rdsk/<dgname>/<volname> /global/<mnt> ufs 2
yes global,logging
```

for vxfs:

```
/dev/vx/dsk/<dgname>/<volname>
/dev/vx/rdsk/<dgname>/<volname> /global/<mnt> vxfs
2 yes global,log6.
```

8. Mount the file system ON ONE NODE:

```
# mount /global/<mnt>
```

Configuring the Sun Cluster 3.x Data Service for Network File System (NFS)

Building upon the previous section the set up a Sun Cluster 3.x Cluster File System, the following steps can be used to setup HA-NFS for failover.

1. Install the Sun Cluster HA for NFS packages using the `/usr/cluster/bin/scinstall` utility

   Run the scinstall utility with no options, and select **Add support for new data services to this cluster node**. Follow prompts to load the data services packages from the data services cd.

Perform the installations on all cluster nodes that will possibly run the data service.

Installation of the Sun Cluster HA for NFS can be verified by running the following command:

```
# pkginfo -l SUNWscnfs
```

2. Register and Configure Sun Cluster HA for NFS

Verify that all of the cluster nodes are online:

```
# /usr/cluster/bin/scstat -n
```

Or (for SC3.2 only):

```
# /usr/cluster/bin/clnode status
```

Add the failover logical hostname/ip address to the `/etc/inet/hosts` file on ALL cluster nodes. The logical hostname is the name of the entity that will failover from one cluster node to another. An ip address needs to be associated with the logical host.

Create a Pathprefix directory. This directory is used to maintain administrative and status information for Sun Cluster HA for NFS. For example make the directory on ONE node as follows:

```
# mkdir -p /global/nfs
```

3. Create a failover resource group that will contain the NFS resources:

```
# scrgadm -a -g <nfs-rg> -y Pathprefix=/global/nfs -h <node1,...>
```

Or (for SC3.2 only):

```
# clresourcegroup create -n <node1...> -p PathPrefix=/global/nfs <nfs-rg>
```

where:

<nfs-rg> = name of the resource group
<node1,...> = list of cluster nodes that can run the NFS data service
For example:

```
# scrgadm -a -g nfs-res-group -y Pathprefix=/global/nfs node1,node2,node3
```

Or (for SC3.2 only):

```
# clresourcegroup create -n node1,node2,node3 -p PathPrefix=/global/nfs
   nfs-res-group
```

4. Configure name service mapping in the `/etc/nsswitch.conf` file on all cluster nodes to first check the local files before checking NIS or NIS+ for rpc lookups. Setting the hosts entry in `/etc/nsswitch` does not contact NIS/DNS before attempting to resolve names locally.

```
# hosts: cluster files [SUCCESS=return] nis# rpc: files nis
```

Note: Please also ensure that the ipnodes entry is of the following format:

ipnodes: files

5. Add the logical hostname resources to the failover resource group:

   ```
   # scrgadm -a -L -g <nfs-rg> -l <log-host-name>
   ```

   Or (for SC3.2 only):

```
# clreslogicalhostname create -g <nfs-rg> -h <log-host-name>
  <log-hostname-resource>
```

   where:

   <nfs-rg> = name of the resource group
   <log-host-name> = name of the logical hostname
   <log-hostname-resource> = name of the logical hostname resource

6. Create the administrative subdirectory below the Pathprefix directory created earlier. For example:

   ```
   # mkdir /global/nfs/SUNW.nfs
   ```

   In the directory created above, create a `dfstab.resource` file, and enter the share options for the NFS data service.

   ```
   # cd /global/nfs/SUNW.nfs
   # vi dfstab.nfs-res
   ```

   The format of this file is the same as the `/etc/dfs/dfstab` file, and a typical entry would look like:

   ```
   # share -F nfs -o ro -d <description> nsf/SUNW.nfs
   ```

7. Register the NFS resource type.

   For Sun Cluster HA for NFS, the resource type is SUNW.nfs:

   ```
   # scrgadm -a -t SUNW.nfs
   ```

Or (for SC3.2 only):

```
# clresourcetype register SUNW.nfs
```

8. Create the NFS resource in the failover resource group.

```
# scrgadm -a -j <r-nfs> -g <nfs-rg> -t SUNW.nfs
```

Or (for SC3.2 only):

```
# clresource create -g <nfs-rg> -t SUNW.nfs -p <r-nfs>
```

where:

<r-nfs> = any unique name for the resource
<nfs-rg> = name of the resource group
SUNW.nfs = name of the resource type

9. Enable the resources and switch the resource group into the online state:

```
# scswitch -Z -g <nfs-rg>
```

Or (for SC3.2 only):

```
# clresource group online -emM <nfs-rg>
```

Setting up Sun Cluster 3.x data service for OPS or RAC

The Sun Cluster 3.x data service for Oracle Parallel Server (OPS) or Real Application Cluster (RAC) enables these applications to run on Sun Cluster nodes and to be managed using Sun Cluster commands. It does not provide for automatic failover or monitoring. OPS/RAC has this functionality already built in. Unlike other Sun Cluster 3.x data services it is not registered to the Sun Cluster 3.x framework. This is also the case with the shared disk groups that are used by OPS/RAC, Solaris Volume Manager, and VERITAS Volume Manager are both supported with the Cluster feature. The Cluster features enables the ability to create shared disk groups. Shared disk groups are simultaneously imported on multiple cluster nodes. The Cluster feature requires a separate license in addition to the base VERITAS Volume Manager license.

OPS/RAC also can be used without a volume manager with Sun Cluster 3.x. In this configuration, redundancy is provided by the RAID support on the storage array.

General setup guidelines for configuring Sun Cluster 3.x OPS/RAC data service

Detailed configuration instructions can be found in the *Sun Cluster 3.1 Data Service for Oracle Parallel Server/Real Application Clusters Guide***.** The following steps assumes that the Sun Cluster 3.x framework, OPS/RAC and Volume Manager are installed. Refer to the installation guides for those products for specific installation procedures. This examples assume that VERITAS Volume Manager is being used.

1. Install the Sun Cluster Support for OPS/RAC packages from the Sun Cluster 3.x Data Services distribution cd.

   For Solaris 9:

   ```
   # cd /cdrom/suncluster_3_1/Sol_9/Packages
   ```

   On all cluster nodes that will be running the OPS/RAC data service install the packages:

```
# pkgadd -d . SUNWscucm SUNWscor SUNWudlm SUNWudlmr SUNWcvmr SUNWcvm
```

   Repeat these procedures on the other cluster nodes that will run the data service. Do not reboot the nodes until the Oracle Distributed Lock Manager (UDLM) has been installed and the shared memory settings have been set up in the `/etc/system` file on all nodes. Verify that licenses for both the Veritas Volume Manager and VERITAS Cluster Feature are installed on all nodes.

2. Install the Oracle UDLM

   If not already created, a database administrator group and Oracle user account need to be created.

   On each node, an example entry in to the `/etc/group` file for the dba group could look like the following:

   ```
   dba:*:600:root,oracle
   ```

   One each node, create an entry for the Oracle use ID in the `/etc/passwd` file. For example:

   ```
   # useradd -u 600 -g dba -d /oracle-home oracle
   ```

   The group/Oracle user ID should be the same on all nodes running the OPS/RAC data service.

   Install the ORCLudlm package on each of the nodes to run the OPS/RAC data service.

3. Update the `/etc/system` file to provide appropriate shared memory resource. These values depend on available resources of the server nodes.

   Note: This is an example of the `/etc/system` file parameter settings only:

   ```
   *SHARED MEMEORY SETTINGS FOR ORACLE
   set shmsys:shminfo_shmmax=4294967295
   set semsys:seminfo_semmap=8024
   set semsys:seminfo_semmni=8048
   set semsys:seminfo_semmns=8048
   set semsys:seminfo_semmsl=8048
   set semsys:seminfo_semmnu=8048
   set semsys:seminfo_semume=2048
   set shmsys:shminfo_shmmin=2048
   set shmsys:shmminfo_shmmni=2048
   set shmsys:shminfo_shmseg=2048
   set semsys:seminfo_semvmx=32767
   ```

   `set noexec_user_stack=1` (For Solaris 8 only. This parameter is enabled by default in later versions.)

   Shut down and reboot all the cluster nodes that will run the OPS/RAC data service.

4. Create a shared disk group for used with the Sun Cluster 3.x OPS/RAC data service.

   The disk devices need to initialized for use with the volume manager. The following is a example command:

   ```
   # /etc/vx/bin/vxdisksetup -i c2t0d25
   ```

   This command needs to run for each devices that will be used by the Veritas Volume Manager. The VERITAS **vxdiskadm** or VMSA utilities can also be used to initialize disk and create disk groups.

   Create a shared disk group:

```
# vxdg -s init <disk-group-name> c2t0d25 c2t0d26 c2t027 ...
```

   Use the following command to list disk groups:

   ```
   # vxdg list
   ```

   At this point, a shared disk group is created and can be used to store the database associated with the OPS/RAC application.

7

Solaris SPARC and CLARiiON

This chapter provides information specific to Sun Solaris hosts connecting to CLARiiON arrays.

Solaris SPARC/CLARiiON environment

This section lists some CLARiiON/Fibre Channel support information specific to the Solaris environment.

Host connectivity

Refer to the *EMC Support Matrix* or contact your EMC representative for the latest information on qualified hosts and host bus adapters.

Boot device support

Sun Solaris has been qualified for booting from CLARiiON FC4700 and CX-series arrays as described in the *EMC Support Matrix* and the appropriate HBA document:

- *EMC Fibre Channel with Emulex Host Bus Adapters in the Solaris Host Environment*, which is available on the Emulex website as described under "Installing and configuring the HBA" on page 147.
- *EMC Fibre Channel with QLogic Host Bus Adapters in the Solaris Host Environment*, which is available on the QLogic website as described under "Host configuration with QLogic HBAs" on page 149.

Logical devices

Solaris supports up to 255 LUNs per target. The CLARiiON FC4700 presents up to 223 LUNs, and the CX-series up to 256 LUNs.

The logical devices presented by the FC4700 are the same on each Storage Processor (SP). A logical unit (LU) reports itself Device Ready on one SP and Device Not Ready on the other SP.

CLARiiON configuration

This section contains information that will help with the installation, use, and management of the storage system. It also contains information that could adversely affect the performance of the storage system if any listed workaround or fix is not implemented.

Operating system

Although FC4500, FC4700, and CX-series storage systems support LUN expansion, the Solaris operating system cannot make use of the increased LUN capacity. When Solaris first labels an unlabeled disk, the geometry of the disk is included as part of the label. Solaris assumes that the disk geometry is fixed, and never updates this information. So even if the LUN is expanded, the Solaris operating system is unaware of any increased LUN capacity.

Online disk suite

Note: This section does not apply for hosts running Solaris 9.

When using SDS or ODS with a CLARiiON array running a version of core software that does not support the unitserialnumber feature, you may see an error message that incorrectly indicates that ODS has found multiple paths to the same device. ODS generates this message because it is looking for serial numbers to uniquely identify disk devices. The message is similar to the following:

```
server# metainit d14 2 1 c7t1d3s0 1 c7t1d4s0
metainit: server1: c7t1d4s0: overlaps with device in d14
```

To create a workaround for this issue, edit the file /kernel/drv/sd.conf by adding the following text:

```
# Start addition for CLARiiON storage
sd-config-list= "DGC     RAID 1", "clariion-data",
                "DGC     RAID 5", "clariion-data",
                "DGC     RAID 10","clariion-data";
clariion-data=1,0x8,0,0,0,0,0;
# End addition for CLARiiON storage
```

Note: Be sure to insert five spaces between **`DGC`** and the **`RAID`** type in the above text strings. The string must match the identification information returned in response to a SCSI **`Inq`** command.

Note: This workaround is invalid for SunCluster 2.2 and 3.x configurations.

Note: A server reboot is necessary for this workaround to take effect.

Sun ZFS (Zettabyte file system)

ZFS file system is a Sun product built into the Solaris 10 Operating System. It presents a pooled storge model that eliminates the concept of volumes as well as all of the related partition management, provisioning, and file system sizing matters. ZFS combines scalability anf flexibility while providing a simple command interface.

For more information on how to operate ZFS functionalities, refer to the Sun's *Solaris ZFS Administration Guide*, available at: http://docs.sun.com/app/docs/doc/819-5461.

CAUTION

EMC supports ZFS version 3 or higher *without* the Snapshot and Clone features. (ZFS v3 is built into the Solaris 10 11/06 Operating System.)

VERITAS Volume Manager

VERITAS Volume Manager (VxVM) is a tool for disk management, which you can use to create logical disks, mirrored and striped volumes. To use the Dynamic Multipathing (DMP) feature of VxVM with a CLARiiON storage system, you need the CLARiiON ASL, which is available from VERITAS.

For instructions on installing and removing VxVM, as well as creating disk groups, mirror volumes, striped volumes, and other related operations, refer to the following documents:

- *VERITAS Volume Manager Installation Guide*
- *VERITAS Volume Manager User's Guide*
- *VERITAS Volume Manager System Administrator's Guide*
- *VERITAS Volume Manager Release Notes*

The rest of this section provides important information for using Volume Manger and DMP.

Recommended Restore Demon parameters

Running the VxVM Restore Demon with a **check_all** restore policy and a frequent interval will result in poor I/O performance with the CX-Series or FC4700 storage systems. EMC recommends that you use the **check_disabled, interval=300** restore policy (VxVM default). If it is necessary to change the restore policy to **check_all**, a long interval, such as 60 minutes or more, is recommended.

Special NDU procedure

Note: Performing an NDU without a working path to each LUN through each SP might generate I/O failures.

When using DMP, you must modify the parameters of the Restore Demon during a Non-Disruptive Upgrade (NDU) on the CX-series or FC4700 storage system. Failing to do so can cause volumes to go off line during the NDU process. When the NDU completes, restore the parameters to their previous state.

You must also enable the NDU delay option on the storage system. You can do this using Navisphere® CLI or Navisphere Manager.

Using Navisphere CLI

Append **`-delay 120`** to the **`navicli ndu -install`**, **`navicli ndu -revert`**, or **`navicli uninstall`** command.

Using Navisphere Manager

In the **Software Installation** dialog box under **NDU Delay**, select **Enable Delay,** and enter a delay value of **120** seconds:

1. Enter the **`vxdmpadm stat restored`** command, and (for use in step 6) record the values for the daemon's interval and policy that the command returns.
2. Enter the **`vxdmpadm stop restore`** command.
3. Enter the following command:

```
vxdmpadm start restore policy=check_disabled interval=30
```

4. Perform the NDU. (Make sure the NDU delay is set to 120 seconds.)
5. Enter the **`vxdmpadm stop restore`** command.
6. Enter the following command:

```
vxdmpadm start restore policy=step_1_policy interval=step_1_interval
```

where:

`step_1_policy` is the value for the daemon policy that you recorded in step 1.

`step_1_interval` is the value of the daemon interval that you recorded in step 1.

Navisphere Manager device name mapping

The host device name displayed by Navisphere Manager for a given LUN might not always match the device name displayed by VxVM. For example, for a CX-Series or FC4700 storage system, Navisphere might display the Solaris device name `c2t1d0s2` for LUN0. When you enter the **`vxdisk list`** command at the host, the resulting list of disks may not include `c2t1d0s2`.

This behavior occurs because VxVM is managing multiple paths to that device. Each path is represented by a Solaris device name, and VxVM selects one of them to represent all the paths. Navisphere does the same, but it may not choose the same name as VxVM.

Resolving a device name mismatch

For any device name displayed by the **`vxdisk list`** command, enter **`vxdisk list`** ***`<devicename>`*** to display a list of all paths associated with that device. One of these paths will match the device name displayed by Navisphere Manager.

For example, if Navisphere displays c2t1d0s2 for LUN0, and the **vxdisk list** command displays c0t1d0s2 and c0t1d1s2, enter the following at the host:

```
vxdisk list c0t1d0s2
```

This displays paths c0t1d0s2, c1t1d0s2, and c2t1d0s2. The last path matches the name displayed by Navisphere Manager, so you know this VxVM device matches LUN0.

Usually, the path names vary from the vxdisk device name only by the controller number (the number following c in the device name). You can make an intelligent guess at the VxVM device associated with the device name displayed by Navisphere Manager by selecting one in which only the controller number is different.

format and vxdctl enable commands

When using DMP with a CLARiiON storage system, the **format** command might intermittently display a disk as being unformatted even though it is labeled. If this occurs, enter the disk number into the **format** command menu again, and it will return formatted.

Note: The **format** and **vxdctl enable** commands might take longer to execute when there is heavy I/O to the LUNs under DMP control.

Labeling new LUNs

When you bind a new LUN on the CX-series or FC4700 storage system, you must use the format command to label it before VxVM can use it. However, the format command displays each path to a disk as if it were a separate disk. If you have many LUNs with many paths each, the list of devices displayed by format will be long, and it can take some time to locate and label the newly bound LUN.

You can simplify the procedure as follows: When you run **vxdiskconfig** or reboot the host to make a newly bound LUN visible, a message appears on the system console identifying each unlabeled device. Also, the format command lists any unlabeled disks at the top of its output, just above **Available Disk Selections**.

FC5400/5500/560x/570x series disk processor enclosures

This release was tested on single-server and multi-server configurations, and on configurations incorporating Fibre Channel hubs. If you use hubs in a multi-server configuration and each server has dual HBAs, you can use only one port on each SP. If you connect two servers, each with dual HBAs, to a storage system *without* using Fibre Channel hubs, you can use both ports on each SP.

FC3400/3500 series storage systems

This release was tested on a dual-server configuration with a maximum configuration of three FC3500 series storage systems configured in a dual loop configuration using Fibre Channel copper cables.

If you are experiencing SCSI command time-outs under heavy I/O load, you may need to add the following entry to the file `/etc/system`:

```
set sd:sd_io_time=120
```

FC5000 series disk-array enclosures (JBOD configuration)

Simultaneous access to both ports of any disk in a dual-loop configuration is unsupported.

Host configuration with Emulex HBAs

This section describes the procedures required to install one or more EMC-approved Emulex host bus adapters (HBAs) into a Solaris host and configure the host for connection to a CLARiiON array over Fibre Channel.

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved HBAs.

There are two HBA drivers that can be used for Emulex HBAs: Emulex LightPulse Fibre Channel Adapter driver (lpfc) and Emulex-Sun LightPulse Fibre Channel Adapter driver (emlxs).

lpfc driver

Using the Emulex adapter with the Solaris operating system requires HBA I/O driver software. The driver functions as the host adapter driver in the host's Common SCSI Architecture (CSA), which is a layer below the Solaris SCSI Target Driver (sd), to present the EMC Fibre Channel devices to the operating system as if they were standard SCSI devices.

An Emulex HBA is identified in the Solaris host by the `lpfcX` (PCI HBA) or `lpfsX` (SBus HBA), where *X* is the driver instance number of the HBA. This information appears in console messages (execute `dmesg`), and can be viewed in the file `/var/adm/messages`.

The instance number of the `lpfc/lpfs` can exceed the number of the adapters. An administrator can determine the mapping between the physical card with the driver instance *X* by disconnecting the cable from the HBA and watching the console, which displays a message similar to the following `lpfc` example:

```
NOTICE:lpfcX: …WWPN:10:00:00:00:c9:YY:YY:YY WWNN:10:00:00:00:c9:YY:YY:YY
```

where **`lpfcX`** is the interface of the specific Emulex adapter.

Note: All EMC-approved Emulex HBAs use the same driver in a SPARC host.

Installing and configuring the HBA

For the steps to install HBAs, cables, and GBICs, and to install and configure the HBA driver, refer to *EMC Fibre Channel with Emulex Host Bus Adapters in the Solaris Host Environment*.

You can obtain the document from the Emulex website, as follows:

1. Access http://www.emulex.com.
2. Click **drivers, software, and manuals** at the left side of the screen.
3. Click **EMC** at the upper center of the next screen.
4. Click the link to your HBA at the left side of the screen.
5. Under **Drivers for Solaris**, find the description of your HBA driver in the **Description** column. Then click the **Installation and Configuration** link in the associated **Online Manuals** column.

emlxs driver

The emlxs driver is a part of the Sun StorEdge SAN Foundation Software. If you choose to use the emlxs driver, you must follow the *Sun StorEdge SAN Foundation Software Installation Guide* which is provided by Sun on the Sun website:

```
http://www.sun.com/documentation
```

The Sun StorEdge SAN Foundation Software 4.4.7a (SAN 4.4.7a) is a minimum version that has been qualified for Emulex legacy HBAs.

To install/upgrade the Firmware and Fcode for an Emulex legacy adapter, follow the *FCA Utilities Reference Manual* documentation which is located on the Emulex website:

```
http://www.emulex.com/ts/docoem/sun/10k.htm
```

Note: EMC does not support the coexistence of the lpfc and emlxs drivers on the same host.

Configuring a boot device on the storage array

Solaris hosts have been qualified for booting from EMC storage array devices interfaced through Fibre Channel as described in the *EMC Support Matrix*.

For the procedure to configure a storage array device as a boot device, refer to *EMC Fibre Channel with Emulex Host Bus Adapters in the*

Solaris Host Environment, which is available on the Emulex website as described under "Installing and configuring the HBA".

Migrating existing Emulex HBAs from lpfc to emlxs

To migrate the existing Emulex HBAs from lpfc to emlxs , refer to *Solaris lpfc to emlxs (SFS) Migration Guide*. This document is provided by Emulex and is located on the Emulex website:

`http://www.emulex.com/ts/docoem/sun/10k.htm`

Host configuration with QLogic HBAs

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved HBAs.

To install one or more EMC-approved QLogic host bus adapters (HBAs) into a Solaris host, configure the host for connection to a CLARiiON array, and configure a CLARiiON device as a host boot device, follow the procedures in *EMC Fibre Channel with QLogic Host Bus Adapters in the Solaris Environment*.

You can obtain the document from the QLogic website, as follows:

1. Access `http://www.qlogic.com`.
2. Click **Downloads** at the left side of the screen.
3. Click the **EMC** link to the right of **OEM approved/recommended drivers and firmware**.
4. Find the description of your HBA and driver in the **Name** column of the table for your HBA model. Then click the **Readme** link in the associated **Description** column.

Host configuration with Sun HBAs

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved HBAs.

EMC qualifies and supports Sun-branded QLogic HBAs and Sun-branded Emulex HBAs.

- The Sun StorEdge SAN Foundation Software 4.2 (SAN 4.2) is the minimum requirement for Sun-branded QLogic adapters.
- The Sun StorEdge SAN Foundation Software 4.4.7a (SAN 4.4.7a) is the minimum requirement for Sun-branded Emulex adapters.

To install the EMC-qualified Sun HBAs into a Solaris host and to configure the host connection to the EMC storage array over Fibre Channel, follow the installation guide that came with your HBAs for specific instructions on setting up that particular hardware.

You also can obtain the document from the Sun website:

```
http://www.sun.com/documentation
```

Making LUNs available to Solaris

This section describes how to specify Solaris disk names for LUNs and describes the following tasks that you must perform to make LUNs in the server's Storage Group available to Solaris.

Specifying Solaris disk names for LUNs

For Solaris, `diskname` has the format

```
cDtSdLsP
```

where:

- *D* is the number of the HBA in the server (controller number). Solaris assigns these numbers.

  For example, the number for an HBA in slot 1 is **0**; for an HBA in slot 2, it is **1**. Under some conditions, Solaris may assign other numbers to HBAs.

- *S* is the target ID (0 through 125) of the SP connected to HBA c*D*.
- *L* is the LUN number (0 through 255).
- *P* is the partition number on the target.

For example, if the HBA is 1, the target ID of the SP is 0, and the LUN number is 2, you would format the disk by entering the following command:

```
format c1t0d2
```

What next? Continue to the next section to label and partition the LUNs.

Partitioning and labeling LUNs

This section describes how to use the `format` command to partition and label LUNs. The version of the `format` command that ships with Solaris has an auto-configure option that configures LUNs (disks) for you.

1. At the root prompt, enter the following command:

   ```
   format diskname
   ```

where ***diskname*** is in the form explained under "Specifying Solaris disk names for LUNs". If you use the `format` command with no arguments, it displays all disks.

2. From the **Format** menu, select **Available Drive Types**.
3. At the `Specify disk type` prompt, enter **`0`** to select the auto-configure option.
4. If you want to partition the disk, use the `format` command again.

   See the Solaris documentation for information on the `format` command.
5. If you have not labeled the disk, label it using the `label` command.

   Refer to the Solaris documentation for information on the `label` command.

What next? Continue to the next section to make file systems on the LUN partitions.

Making file systems on partitions

Use the `newfs` command to make file systems on all newly created partitions. You must specify a character device name to indicate the partition where you want to create the file system.

For example, to make a file system on partition 0 on the first disk (LUN), enter the following command:

```
newfs /dev/rdsk/c1t0d0s0
```

What next? Continue to the next section to mount the file systems.

Mounting file systems

This procedure is the same as for any LUN used with a Sun operating system. Use the `mount` command to mount the file systems that you have created on your storage system. You must specify the block device name of the partition where you created the file system, and the mount point directory.

For example, to mount the file system partition **0** of the first disk configured, at the mount point `/temp`, enter the following command:

```
mount /dev/dsk/c1t0d0s0 /temp
```

What next?

- If you have more than one server connected to the storage system and you run the FirstWatch failover software or the VERITAS Cluster Server (VCS) software, proceed to Chapter 8, ”VCS Cluster with CLARiiON.” Otherwise, the server is ready to use the file systems.
- If you want to reassign ownership of LUNs, refer to Chapter 9, ”Reassigning LUN Ownership with CLARiiON.”

Verifying the server can see its LUNs

Use Navisphere Manager to verify that the server can see its LUNs.

1. For each storage system connected to the server, use Navisphere Manager to verify the server can see its LUNs.
   a. Select the storage system for management.
   b. Click the **Hosts** tab.
   c. Double-click the icon for the Solaris server.
   d. Double-click the **LUNs** icon.
   e. Look for an icon for each LUN that the server should have.
2. If an icon for each LUN exists, the persistent bindings are set correctly.
3. If the icon for any LUN is missing, make sure that the LUN belongs to a Storage Group connected to the server as follows:
   a. Double-click the **Storage Groups** icon.
   b. Right-click the icon for the Storage Group that should contain the missing LUN, and click **Propertie**s.
   c. In the **Storage Group Properties** dialog box, look for the missing LUN in the **LUNs in Storage Group** list.
   d. If the LUN is not listed, click the **Select LUNs** button.
   e. In the **Modify Storage Group** dialog box, look for the missing LUN in the **Select LUNS for Storage Group** list.
   f. If the LUN is not listed, click **Show LUNs in Other Storage Groups**.

- If the LUN is listed, then move the LUN from its current Storage Group into the Storage Group for the server by clicking the right arrow and then **OK**.
- If the LUN is still not listed, then the persistent bindings value in the `fcaw.conf` or `lpfc.conf` file is incorrect or the LUN is missing from the `sd.conf` file.

8

VCS Cluster with CLARiiON

This chapter provides an overview of the steps needed to install VERITAS Cluster Server (VCS) with a CLARiiON storage system. For complete installation instructions, refer to the *VERITAS Cluster Server Installation Guide*.

Setting up a VCS configuration

VERITAS Cluster Server (VCS) and Sun Cluster configurations should be set up to map the HBAs in the same order across all hosts within the cluster. Any LUNs, in these configurations, that require failover between the hosts in the cluster, must be accessible by each host. For this to happen, each host must be connected to the Storage Group to which the LUNs belong.

You can use VERITAS Cluster Server for either switched fabric (requires Access Logix™) or direct-attach configurations.

1. Make all the physical connections from the hosts to the switch (if using) and to the storage system.
2. If you are using a switch, configure the switch with single-initiator zoning. Zone the switch so that each of the HBAs are zoned with each of the SPs.
3. On all the servers you want to include in the cluster, make sure that the topology variable is set as follows:
   - For Fabric configurations:
     - PCI – In the `/kernel/drv/lpfc.conf` file, set the **topology** value to **2** (point-to-point).
     - SBus – In the `/kernel/drv/fcaw.conf` file, set the **enable n-port** value to **1**.
   - For FC-AL configurations:
     - PCI – In the `/kernel/drv/lpfc.conf` file, set the **topology** value to **4** (loop).
     - SBus – In the `/kernel/drv/fcaw.conf` file, set the **enable n-port** value to **0**.
4. Using Navisphere Manager, create RAID groups and LUNs.
5. Create a Storage Group and assign the desired number of LUNs to it.

   Make sure that the state of the Storage Group is set to **Sharable**. Select all the nodes of the VCS cluster that you want to include in the Storage Group.
6. In an ATF environment, run `atf_configure` and then perform a reconfiguration reboot using the command `reboot -- -r`.

7. When the hosts have rebooted, run the `format` command to verify that the hosts can see all the LUNs.

What next? Proceed to "Completing the VCS installation" on page 158.

Completing the VCS installation

Create file systems for the High Availability (HA) Service Groups. These file systems can be either VERITAS File Systems (VxFS) or standard UNIX file systems (UFS).

1. If you are using VERITAS Volume Manager (VxVM), use the appropriate VxVM utility to create Disk Groups and Volumes.
2. Create file systems as follows, where ***`raw_device`*** is a LUN with no file systems.
   - To create VxFS file systems, use the following command

     ```
     mkfs -F vxfs raw_device
     ```

   - To create UFS file systems, use the following command:

     ```
     mkfs raw_device
     ```

3. Using the InstallVCS utility, install VCS software on all the servers, and then create any Service Groups.
   - To Install VCS, refer to the "Installation Procedures" chapter in the *VCS Installation Guide* .
   - To configure VCS, refer to the "VCS Configuration Language" chapter in the *VCS User Guide.*
4. Start VCS cluster on each host using the **`hastart`** command.
5. Use the **`hagrp -state`** command to verify that the Service Groups are on line on their respective primary masters.
6. Use the **`hastatus`** command to verify cluster status.

Verifying the major and minor number configurations

Block devices that provide NFS service must have the same major and minor numbers on each system. Solaris uses major and minor numbers to identify the logical partition or disk slice. NFS also uses these numbers to identify the exported file system. Verify the major and minor numbers to make sure that the NFS identity for the file system is the same after the file is exported.

1. Display the major and minor numbers for each block device that support NFS services by entering the following command:

   Note: For Volume Manager volumes, import the associated shared disk group on each system, and then enter the command.

   ```
   # ls -lL <block_device_location>
   ```

   Example: `# ls -lL /dev/dsk/c1t2d0s3`

   The output may look similar to the following, where the major numbers (32,36), and minor numbers (134, 62) do not match.

   On System A:

   ```
   crw-r----- 1 root sys 32,134 Dec 3 11:50
      /dev/dsk/c1t1d0s3
   ```

   On System B:

   ```
   crw-r----- 1 root sys 36, 62 Dec 3 11:55
      /dev/dsk/c1t1d0s3
   ```

2. If the major numbers (32, 36) match, go to step 3.

   If they *do not* match:

   a. Place the VCS command directory in your path. For example:

   ```
   # export PATH=$PATH:opt/VRTSvcs/bin
   ```

   b. If the block device is a volume, identify on each system the two major numbers that the VERITAS Volume Manager uses:

   ```
   # grep vx /etc/name_to_major
   ```

   Output on System A would be similat to the following:

   ```
   vxio 32
   vxspec 33
   ```

Output on System B would be similar to the following:

```
vxio 36
vxspec 37
```

c. Type the following command on System B to change the major number (36/37) to match that of System A (32/33):

For disk partitions:

```
# haremajor -sd <major_number>
```

For volumes:

```
# haremajor -vx <major_number1> <major_number2>
```

The variable <*major_number*> represents the numbers from System A.

For example, for disk partitions, enter the following command:

```
# haremajor -sd 32
```

For volumes, enter the following:

```
# haremajor -vx 32 33
```

If this command fails, you receive a report similar to the following:

```
Error: Preexisting major number 32
These are available numbers on this system: 128...
Check /etc/name_to_major on all systems for
available numbers.
```

d. If you receive this report, type the following command on System A to change the major number (32/33) to match that of System B (36/37):

For disk partitions:

```
# haremajor -sd 36
```

For volumes:

```
# haremajor -vx 36 37
```

If the command fails again, you will receive a report similar to the following:

```
Error: Preexisting major number 36
These are available numbers on this node: 126...
Check /etc/name_to_major on all systems for
available numbers.
```

e. If you receive the second report, choose the larger of the two major numbers (in this example, 128, reported in the first report), and use this number in the **haremajor** command to reconcile the major numbers. Type the following command on both systems:

   For disk partitions:

   ```
   # haremajor -sd 128
   ```

   For volumes:

   ```
   # haremajor -vx 128 129
   ```

f. Reboot each system on which **haremajor** was successful.

3. If the minor numbers match on this partition, verify the major and minor numbers of your next partition. If the minor numbers *do not* match, do one of the following:

   - If the block device is a volume, go to step 4.
   - If the block device is a disk partition, go to step 5.

4. If the block device on which the minor number does not match is a volume, consult the man page `vxdg(1M)` for instructions on reconciling the Volume Manager minor numbers, with specific reference to the **reminor** option.

   You should now be able to fail over the Service Groups from one host to the other without interrupting I/O.

5. If the block device on which the minor number does not match is a disk partition, complete the following steps:

   a. Type the following command on both systems using the name of your block device:

   ```
   # ls -1 /dev/dsk/c1t1d0s3
   ```

   Output on System A and System B would be similar to the following:

   ```
   lrwxrwxrwx 1 root root 83 Dec 3 11:50 \
   /dev/dsk/c1t1d0s3 -> ../../ \
   devices/sbus@1f,0/QLGC,isp@0,10000/sd@1,0:d,raw
   ```

   where `sbus@1f,0/QLGC,isp@0,10000/sd@1,0` is the device name.

b. Type the following command on both systems to determine the instance numbers used by the SCSI driver:

```
# grep sd /etc/path_to_inst | sort -n -k 2,2
```

Output on System A and System B would be similar to the following:

```
"/sbus@1f,0/QLGC,isp@0,10000/sd@0,0" 0 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@1,0" 1 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@2,0" 2 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@3,0" 3 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@4,0" 4 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@5,0" 5 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@6,0" 6 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@8,0" 7 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@9,0" 8 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@a,0" 9 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@b,0" 10 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@c,0" 11 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@d,0" 12 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@e,0" 13 "sd"
"/sbus@1f,0/QLGC,isp@0,10000/sd@f,0" 14 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@0,0" 15 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@1,0" 16 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@2,0" 17 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@3,0" 18 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@4,0" 19 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@5,0" 20 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@6,0" 21 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@8,0" 22 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@9,0" 23 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@a,0" 24 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@b,0" 25 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@c,0" 26 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@d,0" 27 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@e,0" 28 "sd"
"/sbus@1f,0/SUNW,fas@e,8800000/sd@f,0" 29 "sd"
```

c. Locate the device names on each system and identify the instance numbers. In the example above a device name is **/sbus@1f,0/QLGC,isp@0,10000/sd@1,0**, and the associated instance number is **1**.

d. Compare the device names and associated instance numbers from one system to the other. They should be the same.

- If, for example, a device name on System A has an instance number, but the same device name on System B has no instance number edit the `/etc/path_to_inst` file for System B to add the instance number to match System A.
- If the device name on one system has a different instance number than the same device name on the second system, edit the `/etc/path_to_inst` file on both systems. Change the instance number of both device names to an unused number that is greater than the highest number used by all other devices.

e. Reboot each system on which you modified `/etc/path_to_inst` using the following command:

```
reboot -- -r
```

9

Reassigning LUN Ownership with CLARiiON

This chapter describes methods for reassigning LUN ownership and updating disk names after reassigning LUNs.

Methods for reassigning LUN ownership

You can reassign LUN ownership from one SP to another SP by either of the following:

- Manually using Navisphere Manager or `trespass_array`.
- Automatically using the Application-Transparent Failover (ATF) software (optional for direct connect; required for switch connection).

When you use Navisphere Manager to reassign LUN ownership, you change the default SP owner of the LUN. Neither trespass_array nor ATF change the default SP owner of the LUN. When you change ownership of LUNs using Navisphere Manager, the reassignment of the LUN to the other SP does not take effect until you power the storage system off and then on again. For this reason, EMC recommends that you use Navisphere Manager to reassign LUN ownership in non-failure situations only when:

- You add a second SP and you want to assign LUNs to the new SP.
- You add LUNs and you want to balance your LUNs between two SPs.

When an SP or a component (cable or adapter) in the path to an SP fails, the process of reassigning LUNs from the failed SP to the working SP is called a failover. If ATF is running and a failure occurs, ATF will automatically execute a failover. If ATF is not installed, and the working SP owns at least one LUN, you can fail over LUNs to the working SP using `trespass_array`. When you replace the failed SP or component, manually restore the LUN that failed over to its original SP using either the `atf_restore_all` command or another `trespass_array` command.

Failover using ATF

Note: ATF is required if your server has two HBAs connected to a storage system.

ATF (Application Transparent Failover) automatically reassigns LUNs when a failure occurs. This lets applications continue to run with minimal interruption after an FC-AL or fabric route fails. ATF is the easiest form of failover.

Note: If your system is booted with a path removed, Navisphere ATF will fail over the affected LUNs and allow I/O to take place to these LUNs. However, when the path is restored, Navisphere ATF cannot to restore these LUNs until you reboot the system.

Failover using trespass_array and rescan_array

Using `/usr/bin/trespass_array` and `/usr/bin/rescan_array`**,** you can manually reassign LUNs. When using `trespass_array`, you must allocate an unshared LUN to each SP.

Note: To install the `trespass_array` and `rescan_array` utilities, refer to either of the following manuals that ships with your HBA and HBA driver:

- *Fibre Channel Sbus HBA and Driver for Solaris Installation Guide*
- *Fibre Channel PCI HBA and Driver for Solaris Installation Guide*

trespass_array

The `trespass_array` utility does not require you to turn off storage-system power. Typically, you use `trespass_array` if a path to a LUN through an SP fails. For example, if an adapter is connected to two SPs, and one of the SPs fails, you can use `trespass_array` to transfer control of LUNs from the failed SP to the working SP.

For you to transfer control of LUNs to an SP using `trespass_array`, that SP must already own at least one LUN. The transferred LUNs belong to the new SP until you transfer them back, which you can do with another `trespass_array`. To create entries in `/dev` for the transferred LUNs, you must run the `rescan_array` utility (see the next section), or reboot the server using the `init 6` or `shutdown -y -i6 -g0` command.

The format for the `trespass_array` command is:

```
trespass_array device option
```

where

device is the Solaris name of the disk that you want trespassed.

option is one of the following:

-A	reassigns all LUNs. This is the default.
-H	prints help information for the **trespass_array** command.
-L *n*	reassigns LUN ***n*** only, where ***n*** is the LUN number.
-O	transfers control of all the original LUNs of this SP.

For example, to transfer control of only LUN 3 to device c1t0d0s0, you would issue the following command:

```
trespass_array c1t0d0s0 -L 3
```

rescan_array

Use the `rescan_array` utility after `trespass_array` to size the new devices without having to reboot the operating system.

The `rescan_array` command has no arguments; the format is simply `rescan_array`.

When you use `trespass_array`, the physical addresses of the LUNs change to the FC-AL address ID of the SP that now owns the LUNs. You use `rescan_array` to configure the new physical addresses.

The `rescan_array` utility restarts the disk driver, which rescans for devices. Since LUNs are named in the order that they are found, disk names might change after running `trespass_array` and `rescan_array`.

- **Direct or hub connection** — With Solaris 2.3 or higher, disk device filenames are derived from the controller number, FC-AL address ID, and LUN number of the device. Since a trespassed disk will have a different FC-AL address ID, the disk device filename will always be different after running `trespass_array` and `rescan_array`.
- **Switch connection** — Disk device filenames are derived from the controller number, SP target ID, and LUN number of the device. Since a trespassed disk will have a different SP target ID, the disk device filename will always be different after running `trespass_array` and `rescan_array`.

Updating disk names after reassigning LUNs

For Solaris, a change in LUN ownership using `trespass_array` and `rescan_array` affects disk names according to the type of SP failure that occurs; that is, if an SP fails or if you change SP ownership manually.

Note: If ATF reassigns the LUNs, disk names do not change.

Updating disk names if an SP fails

If an SP fails, you cannot access the LUNs on the failed SP. At this point, disk names remain unchanged. However, the next time you boot the server, you must use the **`boot -r`** command, which changes the disk names to reflect the new target ID (the **t***S* portion of the disk name). In a dual-adapter configuration, the disk names also reflect the new adapter (the **c***D* portion of the disk name).

Updating disk names if you manually change LUN ownership

Manually changing LUN ownership affects disk names. You must power down and reboot the server after manually changing LUN ownership:

1. Shut down the server's operating system.
2. Power off and power on the storage system.
3. Shut down the server.
4. Reboot the server using the **`boot -r`** command.

Now the disk names reflect the new target (the **t***S* portion of the disk name). In a dual-adapter configuration, the disk names reflect the new adapter (the **c***D* portion of the disk name).

Examples — Manually reassigning LUNs

When you have more than one route from the server to a LUN, you can reassign LUN ownership from one SP to the other SP. This section describes a sample unshared direct single-server configuration, and a sample unshared direct dual-server configuration, and then uses these configurations to describe how you can either reassign the default SP owner using Navisphere Manager, or fail over the LUN using `trespass_array`.

Unshared direct single server configuration

This example starts with a Sun Solaris server and one storage system with 10 unbound disk modules. The disk modules were bound into LUNs and the LUNs were made available to Solaris.

The sample system is illustrated in Table 6 and Figure 10.

Table 6 Unshared direct single-server configuration before assigning LUN 0

Adapter	SP and FC-AL address ID	LUNs	Solaris name
Controller 1 in slot 1	SP A FC_AL address ID 6	LUN 0 RAID 5	`/dev/dsk/c1t6d0s<0-7>`
Controller 2 in slot 2	SP B FC_AL address ID 0	LUN 1 RAID 5	`/dev/dsk/c2t0d1s<0-7>`

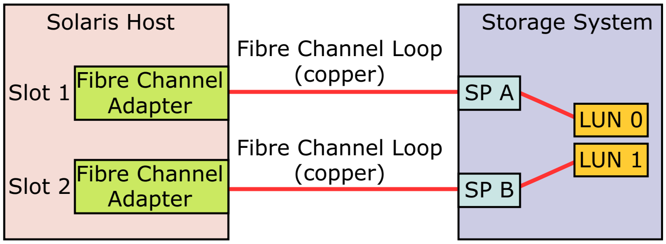


Figure 10 Unshared direct single-server configuration before assigning LUN 0

The procedures that follow describe how to reassign LUN 0 from SP A to SP B. After reassigning the LUN, the configuration will appear as described in Table 7 and Figure 11.

Table 7 Unshared direct single-server configuration after assigning LUN 0

Adapter	SP and FC-AL address ID	LUNs	Solaris name
Controller 1 in slot 1	SP A FC_AL address ID 6		
Controller 2 in slot 2	SP B FC_AL address ID 0	LUN 0 RAID 5	`/dev/dsk/c1t6d0s<0-7>`
		LUN 1 RAID 5	`/dev/dsk/c2t0d1s<0-7>`

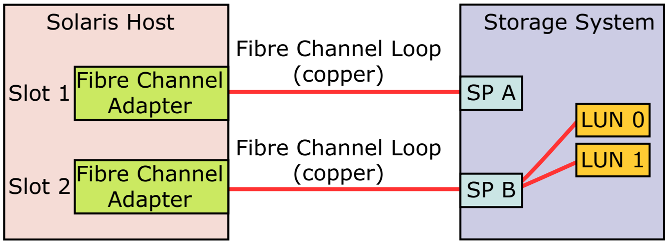


Figure 11 Unshared direct single-server configuration after assigning LUN 0

Reassigning LUN ownership using Navisphere Manager

If you have added LUNs, you might want to use Navisphere Manager to balance the load and reassign the default SP ownership for some LUNs. The new configuration will take effect only after you turn the power to the storage system off and on.

This procedure is not intended to temporarily reassign LUNs when a problem occurs. If a path or SP fails, use either ATF to automatically fail over LUNs or `trespass_array` to manually fail over LUNs.

This section describes how to reassign the default ownership of a LUN from one SP to another SP using the unshared direct configuration. The example reassigns LUN 0, `/dev/dsk/c1t6d0s0` mounted at `/mount1a`, from SP A to SP B.

1. Unmount the file system on the LUN to be reassigned.

   To unmount the file system, use the umount command. For example, if the mount point is /mount1a, you would enter the command:

   ```
   umount /mount1a
   ```

   Note: If you cannot unmount the file system, run fsck on your file system.

2. Change the SP ownership as described in the Navisphere Manager manual.

3. Halt the server by entering:

   ```
   shutdown -y -i0 -g0
   ```

4. Power the storage system off and on.

5. At the ok> prompt, reboot the server by entering:

   ```
   boot -r
   ```

   Rebooting enables the operating system to recognize that a device exists.

6. If necessary, run fsck on your file system. You may not need to run fsck if you were able to unmount the file system in step 1.

7. Mount and use the file system on the new LUN.

In step 1, we used the example mount point /mount1a. Although the new LUN has a different name, mount the partition on /mount1a so that users can still use the pathnames to which they are accustomed.

LUN 0 is reassigned to SP B.

Failing over a LUN using trespass_array

This section describes how to fail over a LUN using trespass_array. In this example, SP A fails and we reassign RAID 5 LUN /dev/dsk/c1t6d0s0 mounted at /mount1a from SP A to SP B.

Note: To install the **trespass_array** and **rescan_array** utilities, refer to either of the following manuals that ships with your HBA and HBA driver:

- *Fibre Channel Sbus HBA and Driver for Solaris Installation Guide*
- *Fibre Channel PCI HBA and Driver for Solaris Installation Guid*

1. Unmount the file system on the LUN whose ownership you are reassigning.

   To unmount the file system, use the `umount` command. For example, if the mount point is `/mount1a`, you enter the command:

   ```
   umount /mount1a
   ```

2. Issue the **`trespass_array`** command as follows to reassign all LUNs from SP A to SP B:

   ```
   trespass_array c1t6d0s0 -A
   ```

   Or issue the following command to reassign a specific LUN (in this instance LUN 0) to SP B:

   ```
   trespass_array c1t6d0s0 -L 0
   ```

3. Issue the **`rescan_array`** command as follows:

   ```
   rescan_array
   ```

   If you do not issue the **`rescan_array`** command, you must reboot the server for Solaris to recognize the reassigned LUN.

   The LUN that you just reassigned can now be remounted. For example:

   ```
   mount /dev/dsk/c2t0d0s0 /mount1a
   ```

4. Run **`fsck`** on the file system if **`rescan_array`** fails and displays a message to run **`fsck`**. This will occur if you could not unmount the file system in step 1.

5. Mount and use the file system on the new LUN.

Step 1 used the example mount point `/mount1a`. Although the new LUN has a different name, use the `mount` command with the new FC-AL address ID, so that users can still use the pathnames to which they are accustomed.

LUN 0 is reassigned to SP B.

Sample unshared direct dual-server configuration

The unshared direct dual-server configuration has two servers, each with an adapter connected to an SP through a Fibre Channel cable. Each server independently uses its own LUNs in the storage system. For detailed configuration information, see the hardware reference or installation manual that ships with the storage system.

The following example starts with two Sun Solaris servers and one storage system with 10 unbound disk modules that are bound into LUNs and are available to the operating system.

Table 8 and Figure 12 illustrate the example.

Table 8 Unshared direct dual-server configuration before assigning LUN 0

Servers	SP and FC-AL address ID	LUNs	Solaris name
Solaris host 1	SP A FC_AL address ID 6	LUN 0 RAID 5	`/dev/dsk/c1t6d0s<0-7>`
Solaris host 2	SP B FC_AL address ID 0	LUN 1 RAID 5	`/dev/dsk/c2t0d1s<0-7>`

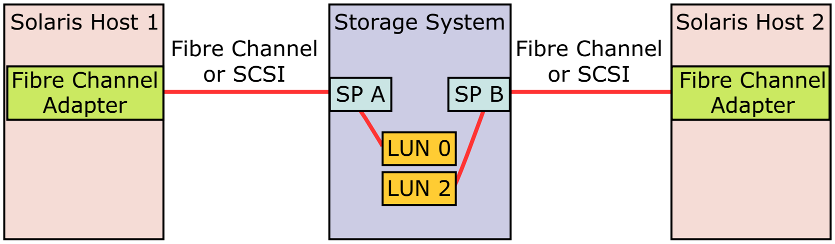


Figure 12 Unshared direct dual-server configuration before assigning LUN 0

The procedures that follow describe how to reassign LUN 0 from SP A to SP B. After reassigning LUN 0, the configuration will appear as shown in Table 9 and Figure 13.

Table 9 Unshared direct dual-server configuration after assigning LUN 0

Servers	SP and FC-AL address ID	LUNs	Solaris name
Solaris host 1	SP A FC_AL address ID 6		
Solaris host 2	SP B FC_AL address ID 0	LUN 0 RAID 5	`/dev/dsk/c1t6d0s<0-7>`
		LUN 1 RAID 5	`/dev/dsk/c2t0d1s<0-7>`

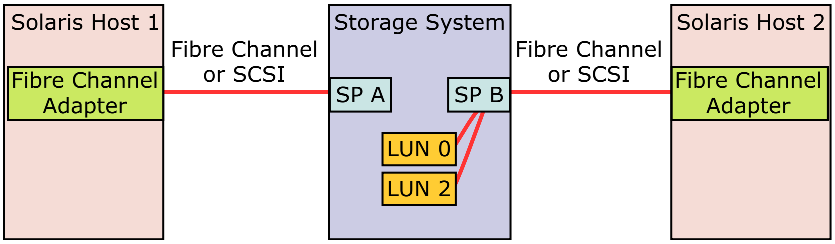


Figure 13 Unshared direct dual-server configuration after assigning LUN 0

Reassigning LUN ownership using Navisphere Manager

If you have added LUNs, you might want to use Navisphere Manager to balance the load and to reassign the default SP ownership for some LUNs. The new configuration will take effect only after you turn the power to the storage system off and on.

This procedure is not intended to temporarily reassign LUNs when a problem occurs. If a path or SP fails, use `trespass_array` to manually fail over LUNs.

From the sample unshared direct dual-server configuration, we reassign the Solaris host1 RAID 5 LUN `/dev/dsk/c1t6d0s0` mounted at `/mount1a` from SP A to SP B, which is connected to server Solaris host2. Before being reassigned, the LUN is owned by SP A at

FC-AL address ID 6, and after it is reassigned, it is owned by SP B at FC-AL address ID 0.

1. On Solaris host1, unmount the mount points for one LUN. In this example we transfer ownership of LUN 0, which has partition device entries `/dev/dsk/c1t6d0s0`. We unmount file systems on these partitions.

   Unmount the file system using the `umount` command. For example, if the mount point is `/mount1a`, the command is

   ```
   umount /mount1a
   ```

   Note: If you cannot unmount the file system, run `fsck` on your file system.

2. Change the SP ownership as described in the Navisphere Manager manual.
3. Halt both servers by entering on each:

   ```
   shutdown -y -i0 -g0
   ```

4. Power the storage system off and on.
5. At the `ok>` prompt, reboot each server by entering

   ```
   boot -r
   ```

   Rebooting enables the operating system to recognize that a device exists.

6. Mount and use the file systems on controller1 on Solaris host2.

Step 1 used the example mount point `/mount1a`. Although the new LUN has a different name, mount the partition on `/mount1a` on the new server so that users can still use the pathnames to which they are accustomed.

Failing over a LUN using trespass_array

This section describes how to reassign a LUN using trespass_array. From the sample unshared direct dual-server configuration, we reassign the Solaris host1 RAID 5 LUN /dev/dsk/c1t6d0s0 mounted at /mount1a from SP A to SP B, which is connected to server Solaris host2. Before being reassigned, the LUN is owned by SP A at FC-AL address ID 6, and after it is reassigned, it is owned by SP B at FC-AL address ID 0.

1. On Solaris host1, unmount the file systems on all LUNs that you are reassigning.

   To unmount a file system, use the umount command. For example, if the mount point is /mount1a, you would enter the command:

   ```
   umount /mount1a
   ```

2. On Solaris host2, issue the **trespass_array** command as follows to reassign all LUNs from SP A to SP B:

   ```
   trespass_array c1t6d0s0 -A
   ```

   Or issue the following command to reassign a specific LUN (in this instance LUN 0) to SP B:

   ```
   trespass_array c1t6d0s0 -L 0
   ```

3. On Solaris host2, issue the rescan_array command:

   ```
   rescan_array
   ```

   If you do not issue the rescan_array command, you must reboot the server for Solaris to recognize the reassigned LUN.

4. On Solaris host2, run fsck on the file systems if necessary.

5. On Solaris host2, mount and use the file systems.

Step 1 used the example mount point /mount1a. Although the new LUN has a different name, mount the partition on /mount1a on the new server so that users can still use the pathnames to which they are accustomed.

PART 2

Solaris x86 Symmetrix/CLARiiON Connectivity

Part 2 includes information specific to the Sun Solaris x86 environment:

10

Sun Cluster 3.x for x86

This chapter discusses EMC Storage / Sun Cluster 3.x for x86 environment. Fundamental concepts and procedures related to Sun Cluster planning, setup, and administration are provided.

Sun Cluster 3.x for x86 overview

This section introduces Sun Cluster 3.x and briefly describes its important features and how they relate to EMC storage.

Note: Sun also refers to x86 as *Opteron*.

What is Sun Cluster 3.x?

Sun Cluster 3.x is a highly available and scalable cluster software framework that is tightly integrated with the Solaris Operating Environment. Sun Cluster 3.x is part of the SunPlex system that includes the Solaris OE, Sun Cluster 3.x, Opteron hardware and networking components. At the time of this writing EMC supports Sun Cluster 3.1 Update 4 (8/05) for x86. Please refer to the *EMC Support Matrix* for EMC's latest support for Sun Cluster.

Sun Cluster 3.x enables the implementation of applications in either a failover or scalable topology or both. A failover configuration is one in which a set of resources and applications are automatically relocated to another server in the event that the primary node fails. For failover services, applications run on only a single server at any one time. In a scalable configuration, a set of resources/applications are spread across cluster servers and run concurrently on them. Service requests come into the cluster through a global network interface and are distributed to the cluster servers based one of several predefined algorithms. Sun Cluster 3.x can also be configured to run Real Application Cluster (RAC).

All Sun Cluster 3.x documentation can be found at:

`http://docs.sun.com`

Hardware components

This section provides information on the hardware components.

Cluster nodes

A cluster node is a server that is running Solaris, Sun Cluster 3.x framework, and Sun Cluster 3.x Data Service software. Up to six (6) nodes are supported in a High Availability environment. Sun Cluster 3.x can be run on most Sun server families. Cluster nodes are connected to Symmetrix disks using both fiber channel and SCSI interfaces. Refer to the *EMC Support Matrix* for all relevant host bus adapters, drivers and switch versions. Nodes that are not physically attached to the storage, but participating in cluster membership, can gain access to storage through the cluster file system.

Cluster members communicate with each other through a mechanism called the Cluster Membership Monitor (CMM) over a set of physically independent networks called the cluster interconnect. The cluster interconnect is discussed later in this chapter.

In general, nodes in the cluster should have similar physical resources such as processors, memory and I/O capability to be able to sufficiently run the applications and resources that may failover to them. Additional server capacity may be required in an Active-Active topology. In such configurations, all servers are primaries for one set of resources and are secondaries in the event that another server in the cluster failed. In this case, the server may need additional system resources in order to run both sets of applications.

Storage

Both the Symmetrix DMX-3 and Cx/Cx-3 Series families are supported with Sun Cluster 3.1, Update 4. Minimum Enginuity and FLARE code revisions exist for both Symmetrix and CLARiiON families. Check the *EMC Support Matrix* for code revisions and other considerations. See Figure 14 on page 186 through Figure 17 on page 189 for supported storage topologies.

Cluster interconnect

The cluster interconnect is a set of private networks that are used to carry membership and data service communications between the nodes participating in the cluster. Redundant private networks are used to avoid a single point of failure in the event that one network component should fail. Up to six networks can be configured, and Sun Cluster 3.x will exploit the additional bandwidth when available. Some cluster topologies, such as Real Application Cluster, use the cluster interconnect extensively. For these configurations, high-speed interconnect technologies should be deployed. The cluster interconnect consists of Network Interface Cards (NICs), junctions (switches/hubs), and cables.

Software components for cluster servers

The following software packages are generally installed on cluster servers:

- Solaris Operating System
- Sun Cluster 3.1 8/05 x86 framework software
- Data service applications
- Volume Manager (Solaris Volume Manager)
- Multipathing software
  - EMC PowerPath or Sun StorEdge Traffic Manager (a/k/a MPxIO)

Supported software versions for Sun Cluster 3.x

Note: Refer to the *EMC Support Matrix* for the latest information regarding supported software versions. Also, refer to the SunSolve website for the latest Solaris and Sun Cluster patch levels

The *EMC Support Matrix* provides the supported configurations and the minimum requirements of related software.

- Sun Cluster 3.x, with the latest Sun Cluster Core Packages, if any.
- VERITAS DMP should *not* be disabled for Sun Cluster 3.x environments with EMC PowerPath.
- EMC PowerPath 5.0 x86 is supported in CLARiiON configurations unless otherwise noted in the *EMC Support Matrix*.
- The *EMC Support Matrix* lists EMC-supported CLARiiON FLARE versions.
- RAC 10gR2 supports Solaris Volume Manager with Sun Cluster 3.1U4 or higher on OS Solaris 10 x86 or higher.
- The *EMC Support Matrix* lists supported RAC configurations.

Sun Cluster 3.x configuration examples

The diagrams on the following pages show several possible configurations for Symmetrix and CLARiiON systems in a Sun Cluster 3.x environment. Refer to the *Sun Cluster 3.x Concepts* manual for additional information.

Typical configurations will include two to four (4) nodes depending on the data services in use. Some or all of the nodes may be physically connected to the Symmetrix system. The current guidelines are:

- Up to six (6) nodes in an HA configuration (non-RAC) can be configured. Any number of these nodes can be physically connected to the storage. While some nodes may not be physically connected to the storage, they have access to storage through the global namespace and cluster file system features in Sun Cluster 3.x. Refer to the *Sun Cluster 3.x Concepts* manual and the *Sun Cluster 3.x Systems Administration Guide* for more information on this functionality.
- Up to four nodes are possible for RAC in Sun Cluster 3.x. All nodes are physically connected to the EMC Storage system.

Figure 14 shows a typical two-node topology for either an HA or RAC configuration. One run of Symmetrix devices is presented to four FA channels and then to all HBAs on the cluster nodes. Sun MPxIO would be deployed for this configuration with Symmetrix RAID protection.

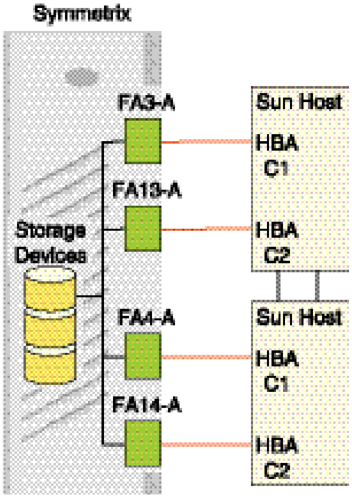


Figure 14 Typical two-node topology

Figure 15 shows a three-node HA configuration. Two nodes are physically attached to the Symmetrix system. The third node at the bottom has access to the storage through the global device namespace and cluster file system. Sun MPxIO would be deployed for this configuration with Symmetrix RAID protection.

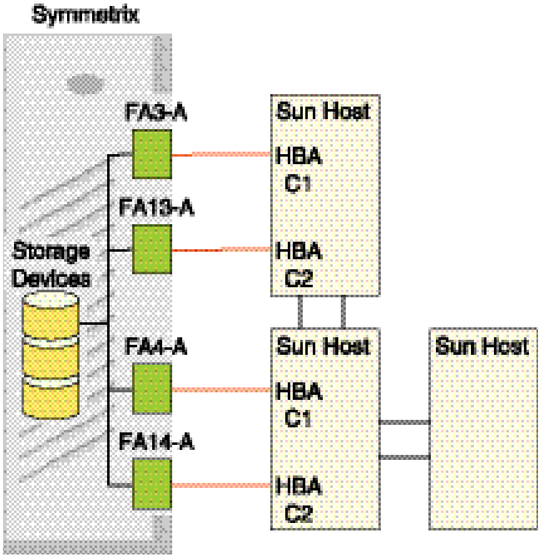


Figure 15 Three-node HA configuration

Figure 16 shows a four-node fully attached configuration. Sun MPxIO is deployed for this configuration with Symmetrix RAID protection.

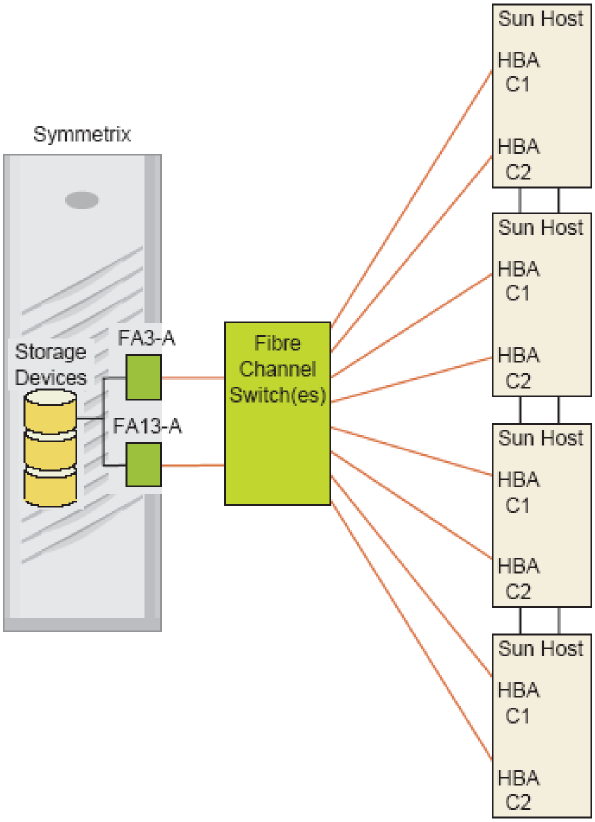


Figure 16 Four-node fully attached configuration

Figure 17 shows a two-node, host-based mirrored configuration. Sun MPxIO is *not* deployed in this configuration.

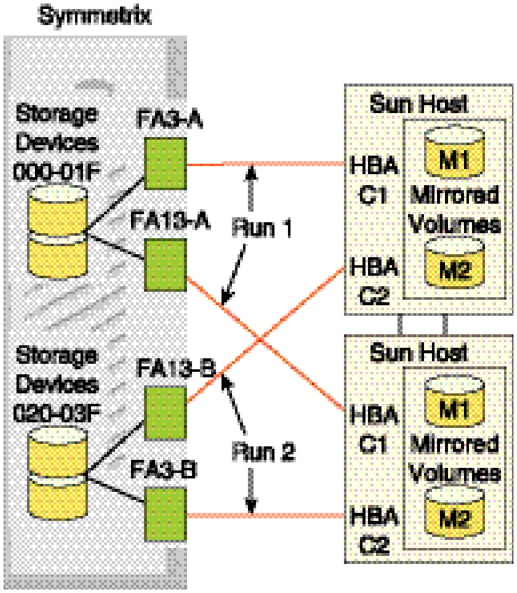


Figure 17 **Two-node host-based mirrored configuration**

Figure 18 on page 190 and Figure 19 on page 190 show possible configurations for an EMC CLARiiON storage system in a Sun Cluster 3.x environment. Refer to the *Sun Cluster 3.x Concepts* manual for additional information.

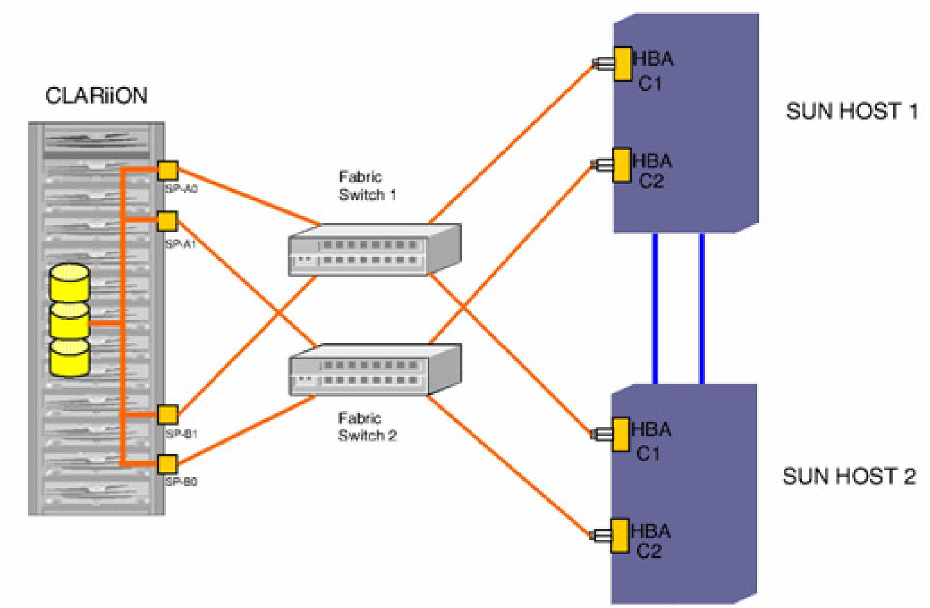


Figure 18 Typical two-node fabric topology

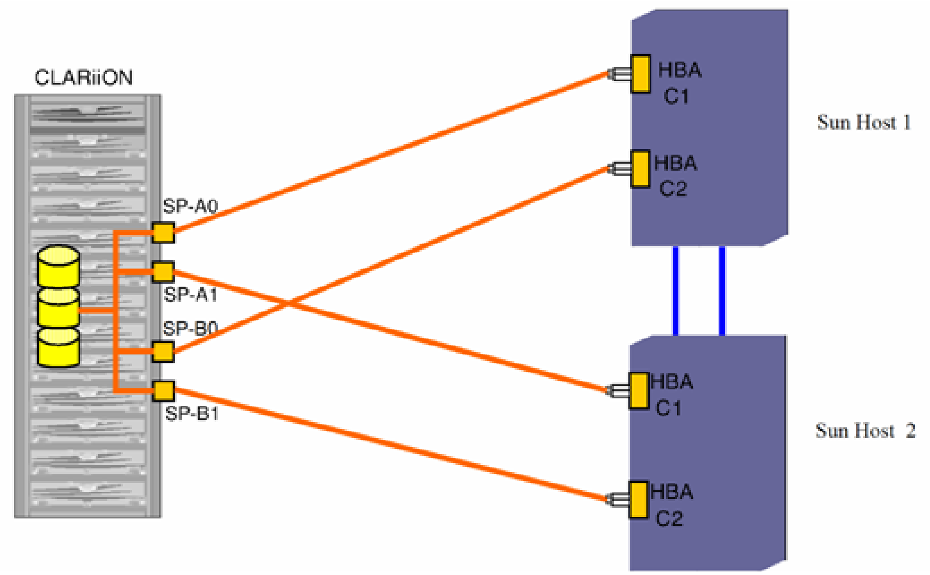


Figure 19 Typical two-node direct-attach topology

Key Sun Cluster 3.x concepts

This section provides information on key Sun Cluster 3.x concepts.

Cluster Membership Monitor (CMM)

The Cluster Membership Monitor (CMM) is a set of agents that use the private interconnects to communicate with the nodes that comprise a Sun Cluster. The CMM performs the following functions:

- Monitors changes in cluster membership when nodes join or leave the cluster
- Ensure that a faulty node leaves the cluster
- Ensure that a faulty node stays out of the cluster until it is repaired
- Protect the cluster from partitioning itself into multiple clusters (split-brain, amnesia)
- Verify full connectivity to all nodes in the cluster

Split-brain is a condition where in the event that all communication is lost between cluster members the cluster partitions into each node (or a subset of nodes) believes it is the only cluster node. In this condition, uncoordinated access to shared storage could result in severe data corruption.

Amnesia is a condition where a node starts with stale cluster configuration data. If a node fails and the cluster is reconfigured on the remaining nodes, then the configuration would be stale on the node that failed. If the failed node then attempts to join the cluster, it must be resynchronized with the current cluster configuration data.

Cluster Configuration Repository (CCR)

The Cluster Configuration Repository (CCR) is a cluster-wide database that stores cluster configuration and state information. Each node in the cluster maintains a copy of the CCR. The Sun Cluster framework maintains the integrity of the CCR by using a two-phase commit protocol where any update to the cluster configuration from one node must successfully complete on all nodes of the cluster.

Global devices

Sun Cluster 3.x uses a concept called global devices to present a cluster-common device view to all nodes in the cluster regardless of where specific devices are physically attached. The cluster automatically detects storage devices on system boot up and assigns unique IDs to each disk device. Devices are assigned global names, and are by default integrated into the cluster environment. This naming scheme allows for a common name to be used by all nodes in the cluster even the actual hardware path to the shared storage devices could vary from one node to the next. Sun Cluster 3.x essentially abstracts data location from data services. Data does not need to be attached to the server that hosts the services. This global namespace is held in the `/dev/global` directory on each node.

Device ID (DID)

To manage the global devices Sun Cluster 3.x uses a pseudo driver called the Device ID (DID). This driver searches for devices attached to the cluster node and automatically builds a list of unique disk devices and assigns a unique major and minor number that's consistent across the cluster. Device access is then performed using the unique device ID that was assigned by the DID instead of the traditional Solaris CxTxDx device ID.

An example of a device name using the DID naming scheme is:

```
/dev/did/dsk/d10s2
```

Global namespace

Global devices are enabled through a construct called the global namespace. The global namespace is implemented using a `/global/.devices/node@x` structure, where x is a node number within the cluster. This number can be 1 – 6, depending on the number of nodes making up the cluster. The global namespace includes the `/dev/global` directory hierarchy as well as the VxVM and SVM volume manager namespaces. All disk devices represented in `/dev/global` and all SVM metadevices and VxVM disk group volumes are symbolically linked to a `/global/.devices/node@x/dev` pseudo entry. Both the global namespace and standard volume manager namespace are available from any node in the cluster.

Table 10 list examples of local and global namespace mappings.

Table 10 Global namespace mappings

Object	Node path name	Global path
Disk	/dev/dsk/c1t0d10s2	/global/.devices/node@2/dev/dsk/c1t0d10s2
DID Name	/dev/did/dsk/d1s2	/global/.devices/node@2/dev/did/dsk/d1s2
SVM Diskset	/dev/md/test/dsk/d11	/global/.devices/node@2/dev/md/test/dsk/d11
VxVM Volume	/dev/vx/dsk/group/vol2	/global/.devices/node@2/dev/vx/dsk/group/vol2

Cluster file systems

The cluster file system feature of Sun Cluster 3.x is a proxy between the operating system kernel and the underlying file system. It allows for the access of file systems regardless of physical location within the cluster. Cluster file systems are dependent on global devices and can be accessed by any node in the cluster through a common name whether or not that node is physically connected to the storage device.

Cluster file systems enable the ability of mounting file systems on a node that is not physically attached to the storage system.

Cluster file systems are mounted using the **`mount -g`** command. They can also be mounted automatically with the `/etc/vfstab` file. See the section later in this chapter for an example of automatically mounting cluster file systems.

Sun Cluster 3.x data services

Sun Cluster 3.x includes agents to make applications highly available through the use of start, stop and monitoring scripts and programs. See the section on configuring the NFS Data Service later in this chapter for an example of Sun Cluster 3.x data service setup. Data services are installed after the Sun Cluster 3.x framework is installed.

Available data services for Sun Cluster 3.x are as follows:

- Sun Cluster HA for Apache
- Sun Cluster HA for Apache Tomcat
- Sun Cluster HA for BroadVision One-To-One Enterprise

- Sun Cluster HA for DHCP
- Sun Cluster HA for DNS
- Sun Cluster HA for MySQL
- Sun Cluster HA for NetBackup
- Sun Cluster HA for NFS
- Sun Cluster HA for Oracle E-Business Suite
- Sun Cluster HA for Oracle
- Sun Cluster Support for Oracle Parallel Server / Real Application Clusters
- Sun Cluster HA for SAP
- Sun Cluster HA for SAP liveCache
- Sun Cluster HA for SWIFTAlliance Access
- Sun Cluster HA for Samba
- Sun Cluster HA for Siebel
- Sun Cluster HA for Sun ONE Application Server
- Sun Cluster HA for Sun ONE Directory Server
- Sun Cluster HA for Sun ONE Message Queue
- Sun Cluster HA for Sun ONE Web Server
- Sun Cluster HA for Sybase ASE
- Sun Cluster HA for WebLogic Server
- Sun Cluster HA for WebSphere MQ
- Sun Cluster HA for WebSphere MQ Integrator

Resource groups

Resource groups are the logical constructs that Sun Cluster 3.x uses to group resources together so they can be managed and made highly available. A resource group is migrated from one node to another in the event of a cluster node failover or initiated switchover. Resource groups can contain data services, disk device groups, network interfaces or other resources. Dependencies can be defined for resource groups to assure that the group cannot be brought up unless all of it underlying resources are available.

Quorum and failure fencing

With any cluster it is important to protect the disk resource from the possibility of having uncoordinated cluster members from writing to shared storage devices and possibly corrupting data. Sun Cluster 3.x uses the CCM to protect the cluster from partitioning into multiple clusters in the event that the cluster interconnects fail. The specific types of failures were discussed above in the CMM section.

Sun Cluster uses a quorum vote algorithm where each node is assigned one vote. In order for a cluster to be operational it must have a majority of votes. If the cluster interconnects or node fails, the partition with a majority of votes will remain operational. This model works well with clusters with more than two nodes. In the case of a two node cluster where the vote majority is two and the partitioned nodes could not achieve a majority of votes. Sun Cluster 3.x solves this by assigning an external vote to a quorum device. The quorum device can be any disk device that is shared between two or more nodes. EMC Symmetrix/CLARiiON devices are commonly used for this purpose. It is recommended that an external quorum device be configured regardless of node count. An example of configuring a quorum device on Symmetrix is provided later in the chapter.

Important Sun Cluster 3.x utilities

This section provides information on important Sun Cluster 3.x utilities.

scinstall

The **`scinstall`** command performs Sun Cluster node initialization, installation, and upgrade tasks. It can be run as an interactive utility or by the command line.

The **`scinstall`** command installs and initializes a node as a new Sun Cluster member. It either establishes the first node in a new cluster or adds a node to an already-existing cluster. It can also be used to remove cluster configuration information and uninstall Sun Cluster software from a cluster node.

The upgrade form (-u) of **`scinstall`**, which has several modes and options, upgrades a Sun Cluster node. Always run this form of the scinstall command from the node being upgraded.

The print release form (-p) of **`scinstall`** prints release and package versioning information for the Sun Cluster software installed on the node from which the command is run.

Without options, the **`scinstall`** command attempts to run in interactive mode. Run all forms of the **`scinstall`** command other than the print release form (-p) as superuser.

The **`scinstall`** command is located in the Tools directory on the Sun Cluster CD-ROM. If the Sun Cluster CD-ROM is copied to a local disk, *cdrom-mnt-pt* is the path to the copied Sun Cluster CD-ROM image. The SUNWscu software package also includes a copy of the scinstall command.

scsetup

At post-install time, the **`scsetup`** utility performs initial setup tasks, such as configuring quorum devices and resetting *installmode*. Always run the **`scsetup`** utility just after the cluster is installed and all of the nodes have joined for the first time.

After *installmode* is disabled, scsetup provides a menu-driven front end to most ongoing cluster administration tasks.

You can execute **scsetup** from any node in the cluster. However, when installing a cluster for the first time, it is important to wait until all nodes have joined the cluster before running scsetup and resetting *installmode*.

The **scsetup** interactive utility can be used to configure cluster quorum devices.

scconf

The **scconf** utility manages the Sun Cluster software configuration. You can use **scconf** to add items to the configuration, to change properties of previously configured items, and to remove items from the configuration. **scconf** can be used to configure the cluster quorum device. In each of these three forms of the command, options are processed in the order in which they are typed on the command line. All updates associated with each option must complete successfully before the next option is considered.

The **scconf** command can only be run from an active cluster node. As long as the node is active in the cluster, it makes no difference which node is used to run the command. The results of running the command are always the same, regardless of the node used.

The `-p` option of scconf enables you to print a listing of the current configuration.

scdidadm

The **scdidadm** utility is used to administer the device identifier (DID) pseudo device driver.

The **scdidadm** utility performs the following primary operations:

- Creates driver configuration files
- Modifies entries in the file
- Loads the current configuration into the kernel
- Lists the mapping between device entries and did driver instance numbers

The startup script `/etc/init.d/bootcluster` uses the **scdidadm** utility to initialize the did driver. You can also use **scdidadm** to update or query the current device mapping between the devices present and the corresponding device identifiers and did driver instance numbers.

scgdevs

The **scgdevs** utility manages the global device namespace. The global device namespace is mounted under `/global` and consists of a set of logical links to physical devices. As `/dev/global` is visible to each node of the cluster, each physical device is visible across the cluster. This fact means that any disk, tape, or CD-ROM that is added to the global devices namespace can be accessed from any node in the cluster. The **scgdevs** command allows the administrator to attach new global devices (for example, tape drives, CD-ROM drives, and disk drives) to the global devices namespace without requiring a system reboot.

The **drvconfig** and **devlinks** commands must be executed prior to running **scgdevs**.

Alternatively, a reconfiguration reboot can be used to rebuild the global namespace and attach new global devices. **scgdevs** must be run from a node that is a current cluster member. If this script is run from a node that is not a cluster member, the script exits with an error code and leaves the system state unchanged.

scstat

The **scstat** utility displays the current state of Sun Cluster and its components. Only one instance of the **scstat** utility needs to run on any machine in the Sun Cluster configuration.

When run without any options, **scstat** displays the status for all components of the cluster. This display includes the following information:

- A list of cluster members
- The status of each cluster member
- The status of resource groups and resources
- The status of every path on the cluster interconnect
- The status of every disk device group
- The status of every quorum device

- The status of every Internet Protocol Network Multipathing group and public network adapter

scswitch

The **scswitch** utility is used to move resource groups or disk device groups from one node to another. It also evacuates all resource groups and disk device groups from a node by moving ownership elsewhere, brings resource groups or disk device groups offline and online, enables or disables resources, switches resource groups to or from an unmanaged state, or clears error flags on resource groups.

You can run the **scswitch** utility from any node in a Sun Cluster configuration. If a device group is offline, you can use **scswitch** to bring the device group online onto any host in the node list. However, after the device group is online, a switchover to a spare node is not permitted. Only one invocation of **scswitch** at a time is permitted.

Do not attempt to kill an **scswitch** operation that is already underway.

scshutdown

The **scshutdown** utility shuts down an entire cluster in an orderly fashion. Before starting the shutdown, **scshutdown** sends a warning message and then a final message asking for confirmation. Only run the **scshutdown** command from one node. The **scshutdown** performs the following actions when it shuts down a cluster:

- Changes all functioning resource groups on the cluster to an offline state. If any transitions fail, **scshutdown** does not complete and displays an error message.
- Unmounts all cluster file systems. If any unmounts fail, **scshutdown** does not complete and displays an error message.
- Shuts down all active device services. If any transition of a device fails, **scshutdown** does not complete and displays an error message.
- Runs `/usr/sbin/init 0` on all nodes and brings them to the **OK>** prompt.

For detailed instructions on how to use these Sun Cluster 3.z utilities refer to the *Sun Cluster 3.1 Reference Manual P/N 817–0522–10*

Configuring EMC Symmetrix with Sun Cluster 3.x

This section provides information on configuring EMC Symmetrix with Sun Cluster 3.x.

Symmetrix setup for Sun Cluster 3.x

The following settings are required for proper operation EMC Symmetrix within the Sun Cluster 3.x environment:

Sun Cluster 3.x uses SCSI-3 PGR (persistent group reservation) for storage devices that are accessible through more than two paths. Symmetrix systems support this functionality using the **PER** setting on the SymmWin **Edit Volumes** dialog. This flag must be set for all devices that will be presented to the Sun Cluster nodes.

Note: The **SCL** director flag must be OFF. The **SC3** director flag is not required for Sun Cluster 3.x.

Follow these steps to set up a Symmetrix system:

1. Configure the Symmetrix system for operation in the Sun Solaris operating system environment. Refer to the *EMC Support Matrix* for details. Verify that the Symmetrix system is running an appropriate version of the Enginuity operating environment.
2. Set the **C** (Common Serial Number) director flag for all FA / SA ports to be seen by the Sun Cluster 3.x nodes. This feature is accessed through the SymmWin **Edit Directors** screen.
3. Set the **PER** flag for all volumes that will be presented to the Sun Cluster 3.x nodes. This feature is accessed through the SymmWin **Edit Volumes** screen. The **PER** flag must also be set for data volumes and quorum devices. It is not needed for gatekeepers and VCM database volumes.

FA port sharing

Multiple Sun Clusters can share the same FA ports on a Symmetrix. In addition, Symmetrix FA ports can be shared between Sun Cluster 3.x nodes and non-clustered Solaris nodes. This feature is enabled through the use of the *EMC Solutions Enabler Symmetrix Device Masking CLI Product Guide.*

Configuring EMC CLARiiON with Sun Cluster 3.x

Setup and configuration of EMC CLARiiON in the Sun Cluster 3.x environment can be set up in fabric or loop mode depending on requirements. Multipathing software (EMC PowerPath or Sun's MPxIO) is required for HA configurations. You must configure a minimum of two paths to each CLARiiON device. Refer to the *EMC PowerPath for UNIX Release Notes* for details on driver and device configuration.

Sun Cluster 3 uses SCSI-3 PGR (Persistent Group Reservation) for storage devices that are accessible through more than two paths. This is either from a single node or multiple nodes. Figure 18 on page 190 shows an example of multiple nodes. EMC CLARiiON supports this functionality by deploying EMC PowerPath on all cluster nodes.

Installation guidelines

Verify that the CLARiiON storage system is running an appropriate version of FLARE firmware. Configure the CLARiiON storage system for operation in the Sun Solaris environment. For example, set the following settings on CLARiiON Array:

systemtype to 3
failovermode to 1
arraycommpath to 1
unitserialnumber to Array

Sun Cluster 3.x servers

This section provides guidelines for a new installation. Refer to the *Sun Cluster 3.x Installation Guide* for additional details.

1. Install Sun Solaris software with latest patch set from SunSolve.
2. Make sure that there is at least 512 MB of available space on the local disk for the /globaldevices partition used for the global device namespace.
3. Install and configure HBA driver and /kernel/drv/sd.conf (for Emulex and QLogix drivers only). If using a third-party HBA, refer to the *EMC Support Matrix* for supported server and HBA

combinations. Refer to the *Fibre Channel PCI and SBus HBA and Driver for Solaris Installation Guide* for details on driver and device configuration.

4. Install EMC PowerPath software and any required patches. Refer to the *PowerPath for UNIX Installation and Administration Guide* for details.

5. On all nodes, use the scinstall program to install the Sun Cluster 3.x framework software (included on the CD Distribution) and any Sun Cluster Core Packages patches.

6. After the cluster nodes have rebooted, use the /opt/cluster/bin/scsetup utility to reset the cluster installmode and configure a quorum device.

7. Install the latest Sun Cluster Patches from SunSolve website.

8. Install the volume manager and required patch(s) (if any) on all cluster nodes.

9. Install any required cluster data services. Refer to the *Sun Cluster 3.x Data Services Installation Guide* for details.

Examples

This section provides examples which may be helpful.

Setting up a Sun Cluster 3.x quorum device on EMC storage

A quorum device can be configured either using the interactive **scsetup/clsetup (SC3.2 only)** utility or with the **scconf/clquorum (SC3.2 only)** command line utility. Examples of both procedures are given below.

Using the **scconf** command:

```
# /usr/cluster/bin/scconf -a -q globaldev=d12
```

where:

d12 is the global device number.

A list of available global devices can be generated by using the /usr/cluster/bin/scdidadm -L command.

Or (for SC3.2 only):

```
# /usr/cluster/bin/clquorum add d12
```

Using the scsetup interactive utility:

1. Enter the scsetup utility:

   ```
   # /usr/cluster/bin/scsetup
   ```

   Or (for SC3.2 only):

   ```
   # /usr/cluster/bin/clsetup
   ```

   The main menu is displayed.

2. Select option 1 (Quorum) from the main menu.

   The Quorum Menu is displayed.

3. Select option 1 from the Quorum Menu **Add a quorum device**.

4. Follow the interactive instructions, and type in the global device number of the device to be used as the quorum device.

5. Verify that the quorum device has been added and is online with either of the following commands:

   ```
   # /usr/cluster/bin/scstat -q
   ```

Or (for SC3.2 only):

```
# /usr/cluster/bin/clquorum status
```

How to create a cluster file system

This procedure assumes that all components of the cluster are installed and configured (Solaris, Sun Cluster 3.x framework and VERITAS Volume Manager).

1. Create and register a VERITAS Volume Manager (VxVM) disk group.

   a. To create a VERITAS disk group:

   ```
   # vxdg init <dgname> c1t1d1,c1t1d2,c1t1d3… etc.
   ```

   where:

   <dgname> = name of the disk group to be created

   b. To register the disk group with the Sun Cluster 3.x framework:

   ```
   # /usr/cluster/bin/scsetup
   ```

2. Select **Device groups and volumes.**
3. Select **Register a VxVM disk group as a device group**.
4. Follow the prompts to register the newly created disk group.
5. Create your volumes, synchronize, and newfs them

   a. To create a VERITAS volume within the newly created disk group:

```
# vxassist -g <dgname> -U fsgen make <volname> <vol-size> <diskname> (or use vmsa)
```

   where:

   <dgname> = name of the disk group
   <volname> = name of the volume
   <volsize> = size of the volume
   <diskname> = name of the disk to use for the volume

   b. After volumes are created, the disk group needs to be synchronized. This is also the case when any volume changes are made to a disk group.

   ```
   # /usr/cluster/bin/scsetup
   ```

Or (for SC3.2 only):

```
# /usr/cluster/bin/clsetup
```

- Select **Device groups and volumes**.
- Select **Synchronize volume information for a VxVM device group**.

c. newfs your volumes:

for ufs:

```
newfs /dev/vx/rdsk/<dgname>/<volname>
```

for vxfs:

```
mkfs -F vxfs /dev/vx/rdsk/<dgname>/<volname>
```

6. Mount volumes:

```
mkdir /global/<mnt> ON ALL NODES5.
```

7. Add entries to vfstab ON ALL NODES THAT ARE DIRECTLY ATTACHED

for ufs:

```
/dev/vx/dsk/<dgname>/<volname>
/dev/vx/rdsk/<dgname>/<volname> /global/<mnt> ufs 2
yes global,logging
```

for vxfs:

```
/dev/vx/dsk/<dgname>/<volname>
/dev/vx/rdsk/<dgname>/<volname> /global/<mnt> vxfs
2 yes global,log6.
```

8. Mount the file system ON ONE NODE:

```
# mount /global/<mnt>
```

Configuring the Sun Cluster 3.x data service for Network File System (NFS)

Building upon the previous section the set up a Sun Cluster 3.x Cluster File System, the following steps can be used to setup HA-NFS for failover.

1. Install the Sun Cluster HA for NFS packages using the `/usr/cluster/bin/scinstall` utility.

   Run the scinstall utility with no options, and select **Add support for new data services to this cluster node**. Follow prompts to load the data services packages from the data services cd.

Perform the installations on all cluster nodes that will possibly run the data service.

Installation of the Sun Cluster HA for NFS can be verified by running the following command:

```
# pkginfo -l SUNWscnfs
```

2. Register and Configure Sun Cluster HA for NFS

   Verify that all of the cluster nodes are online:

```
# /usr/cluster/bin/scstat -n
```

   Or (for SC3.2 only):

```
# /usr/cluster/bin/clnode status
```

   Add the failover logical hostname/ip address to the `/etc/inet/hosts` file on ALL cluster nodes. The logical hostname is the name of the entity that will failover from one cluster node to another. An ip address needs to be associated with the logical host.

   Create a Pathprefix directory. This directory is used to maintain administrative and status information for Sun Cluster HA for NFS. For example make the directory on ONE node as follows:

```
# mkdir -p /global/nfs
```

3. Create a failover resource group that will contain the NFS resources:

```
# scrgadm -a -g <nfs-rg> -y Pathprefix=/global/nfs -h <node1,...>
```

   Or (for SC3.2 only):

```
#clresourcegroup create -n <node1...> -p PathPrefix=/global/nfs <nfs-rg>
```

   where:

   <nfs-rg> = name of the resource group
   <node1,...> = list of cluster nodes that can run the NFS data service

   For example:

```
# scrgadm -a -g nfs-res-group -y Pathprefix=/global/nfs node1,node2,node3
```

Or (for SC3.2 only):

```
# clresourcegroup create -n node1,node2,node3 -p PathPrefix=/global/nfs
   nfs-res-group
```

4. Configure name service mapping in the `/etc/nsswitch.conf` file on all cluster nodes to first check the local files before checking NIS or NIS+ for rpc lookups. Setting the hosts entry in `/etc/nsswitch` does not contact NIS/DNS before attempting to resolve names locally.

```
# hosts: cluster files [SUCCESS=return] nis# rpc: files nis
```

Note: Please also ensure that the ipnodes entry is of the following format:

ipnodes: files

5. Add the logical hostname resources to the failover resource group:

```
# scrgadm -a -L -g <nfs-rg> -l <log-host-name>
```

Or (for SC3.2 only):

```
# clreslogicalhostname create -g <nfs-rg> -h <log-host-name>
   <log-hostname-resource>
```

where:

<nfs-rg> = name of the resource group
<log-host-name> = name of the logical hostname
<log-hostname-resource> = name of the logical hostname resource

6. Create the administrative subdirectory below the Pathprefix directory created earlier. For example:

```
# mkdir /global/nfs/SUNW.nfs
```

In the directory created above, create a `dfstab.resource` file, and enter the share options for the NFS data service.

```
# cd /global/nfs/SUNW.nfs
# vi dfstab.nfs-res
```

The format of this file is the same as the `/etc/dfs/dfstab` file, and a typical entry would look like:

```
# share -F nfs -o ro -d <description> nsf/SUNW.nfs
```

7. Register the NFS resource type.

For Sun Cluster HA for NFS, the resource type is SUNW.nfs:

```
# scrgadm -a -t SUNW.nfs
```

Or (for SC3.2 only):

```
# clresourcetype register SUNW.nfs
```

8. Create the NFS resource in the failover resource group.

```
# scrgadm -a -j <r-nfs> -g <nfs-rg> -t SUNW.nfs
```

Or (for SC3.2 only):

```
# clresource create -g <nfs-rg> -t SUNW.nfs -p <r-nfs>
```

where:

<r-nfs> = any unique name for the resource
<nfs-rg> = name of the resource group
`SUNW.nfs` = name of the resource type

9. Enable the resources and switch the resource group into the online state:

```
# scswitch -Z -g <nfs-rg>
```

Or (for SC3.2 only):

```
# clresource group online -emM <nfs-rg>
```

Setting up Sun Cluster 3.x data service for RAC

The Sun Cluster 3.x data service for Oracle Real Application Cluster (RAC) enables these applications to run on Sun Cluster nodes and to be managed using Sun Cluster commands. It does not provide for automatic failover or monitoring. RAC has this functionality already built in. Unlike other Sun Cluster 3.x data services it is not registered to the Sun Cluster 3.x framework. This is also the case with the shared disk groups that are used by RAC. Solaris Volume Manager and VERITAS Volume Manager are both supported with the Cluster Feature unless otherwise stated in the *EMC Support Matrix*. The Cluster Features enables the ability to create shared disk groups. Shared disk groups are simultaneously imported on multiple cluster nodes. The Cluster Feature requires a separate license in addition to the base VERITAS Volume Manager license.

RAC also can be used without a volume manager with Sun Cluster 3.x. In this configuration, redundancy is provided by the RAID support on the storage array.

General setup guidelines for configuring Sun Cluster 3.x RAC data service

Detailed configuration instructions can be found in the *Sun Cluster 3.1 Data Service for Oracle Parallel Server/Real Application Clusters Guide***.** The following steps assumes that the Sun Cluster 3.x framework, RAC and Volume Manager are installed. Refer to the installation guides for those products for specific installation procedures. This examples assume that VERITAS Volume Manager is being used.

1. Install the Sun Cluster Support for RAC packages from the Sun Cluster 3.x Data Services distribution cd.

   For Solaris 10 x86:

   ```
   # cd /cdrom0/components/SunCluster Oracle RAC FRAMEWRK
      3.1/Solaris_10/Packages
   ```

   On all cluster nodes that will be running the RAC data service install the packages:

   ```
   # pkgadd -d . SUNWscucm SUNWscor
   ```

   Repeat these procedures on the other cluster nodes that will run the data service. Do not reboot the nodes until the shared memory settings have been set up in the `/etc/system` file on all nodes.

2. If Solaris Volume Manager is being used as a storage management scheme, then run the following commands on all the nodes of the cluster:

   ```
   # cd /cdrom/ cdrom0/components/SunCluster_Oracle_RAC
      CVM_3.1/Solaris_10/Packages
   # pkgadd -d . SUNWscmd
   ```

   If not already created, a database administrator group and Oracle user account need to be created.

   On each node, an example entry in to the `/etc/group` file for the dba group could look like the following:

   ```
   dba:*:600:root,oracle
   ```

   One each node, create an entry for the Oracle use ID in the `/etc/passwd` file. For example:

```
# useradd -u 600 -g dba -d /oracle-home oracle
```

The group / Oracle user ID should be the same on all nodes running the RAC data service.

3. Update the `/etc/system` file to provide appropriate shared memory resource. These values depend on available resources of the server nodes.

Note: This is an example of the `/etc/system` file parameter settings only:

```
*SHARED MEMEORY SETTINGS FOR ORACLE
set shmsys:shminfo_shmmax=4294967295
set semsys:seminfo_semmap=8024
set semsys:seminfo_semmni=8048
set semsys:seminfo_semmns=8048
set semsys:seminfo_semmsl=8048
set semsys:seminfo_semmnu=8048
set semsys:seminfo_semume=2048
set shmsys:shminfo_shmmin=2048
set shmsys:shmminfo_shmmni=2048
set shmsys:shminfo_shmseg=2048
set semsys:seminfo_semvmx=32767
```

Shut down and reboot all the cluster nodes that will run the RAC data service.

4. Create a shared disk group for used with the Sun Cluster 3.x RAC data service.

Create a shared disk group:

```
# metaset -s dg1 -a /dev/did/dsk/d0 /dev/did/dsk/d1
# metaset -s dg1 d0 1 1 /dev/did/dsk/d0s0
# metaset -s dg1 d1 1 1 /dev/did/dsk/d1s0
```

Use the following command to list disk groups:

```
# metaset -s dq1
```

At this point, a shared disk group is created and can be used to store the database associated with the RAC application.

11

Solaris x86 Symmetrix/CLARiiON over Fibre Channel

This chapter contains Symmetrix/CLARiiON support information specific to the Sun Solaris x86 operating system environment.

Solaris x86 Symmetrix/CLARiiON over Fibre Channel environment

This section contains Symmetrix/CLARiiON support information specific to the Sun Solaris x86 Operating System over Fibre Channel environment.

Hardware

Refer to the *EMC Support Matrix* or contact your EMC representative for the latest information on qualified Sun AMD64 based servers, Host Bus Adapters (HBA), and connectivity equipments.

Software

EMC supports Solaris 10 x86 Operating System or higher on Sun AMD64 based servers.

Boot device support

Booting from the Symmetrix/CLARiiON device is available to Solaris x86 hosts as described under *Boot Device Support* in the *EMC Support Matrix*.

Symmetrix/CLARiiON configuration

The Symmetrix/CLARiiON system is configured by an EMC Customer Engineer through the array service processor.

Refer to the *Fibre Bit Setting* section in the *EMC Support Matrix* for required and/or recommended director bit setting.

Useful Solaris utilities

The following are some Solaris utilities you can use to define and manage Symmetrix/CLARiiON devices. The use of these utilities is optional and for reference only.

- format – The Solaris disk format utility that allows you to format, partition, and label disk drives.
- newfs – Create a file system.

For information on managing disks and file systems, refer to the "Device and File Systems" section in the Sun document *System Administration Guide*, which is available on:

```
http://docs.sun.com/app/docs/doc/819-2723/819-2723
```

System and error messages

Solaris displays the system and error messages on the console and also logs them in a file called /var/adm/messages.

Sun ZFS (Zettabyte file system)

ZFS file system is a Sun product built into the Solaris 10 Operating System. It presents a pooled storge model that eliminates the concept of volumes as well as all of the related partition management, provisioning, and file system sizing matters. ZFS combines scalability anf flexibility while providing a simple command interface.

For more information on how to operate ZFS functionalities, refer to the Sun's *Solaris ZFS Administration Guide*, available at: http://docs.sun.com/app/docs/doc/819-5461.

CAUTION

EMC supports ZFS in Solaris 10_x86 11/06 or later. The Snapshot and Clone features of ZFS are supported only through Sun Microsystems.)

Solaris Volume Manager (SVM)

Solaris Volume Manager (SVM) is an application that built-in the Solaris 10 x86 Operating System. The SVM can be used to manage your system's storage needs.

For information about how to create *State Database Replicas,* create RAID-0/RAID-1/RAID-5 volumes, create Disk Set and other features, refer to the *Solaris Volume Manager Administration Guide,* which is available on:

`http://docs.sun.com/app/docs/doc/806-6111`

VERITAS Volume Manager (Solaris x64)

VERITAS Volume Manager (VxVM) and VERITAS File System (VxFS) are tools for disk and file management. VxVM can be used to create logical disks, mirrored and striped volumes. VxFS supports large file systems, file system expansion and a journaling file system.

Refer to the following documents for instructions on installing VxVM and VxFS, as well as creating disk groups, mirror volumes, striped volumes, and other related operations:

- *VERITAS Storage Foundation Installation Guide*
  Solaris x64 Platform Edition
- *VERITAS Volume Manager Administrator's Guide*
  Solaris x64 Platform Edition
- *VERITAS Volume Manager Hardware Notes*
  Solaris x64 Platform Edition
- *VERITAS File System Administrator's Guide*
  Solaris x64 Platform Edition

The above documents are available at:
http://seer.support.veritas.com/docs

CAUTION

VERITAS Dynamic Multipathing (DMP) functionality requires enabling the Symmetrix director C-bit flag.

Configuring MPxIO for Symmetrix/CLARiiON devices

MPxIO is a feature of the Sun StorEdge SAN Foundation Software that allows I/Os to failover from one path to another available path and automatically resumes on the original path once the original path is repaired.

To enable the MPxIO:

1. Set to file */kernel/drv/fp.conf*

   ```
   mpxio-disable="no";
   ```

2. **For Symmetrix devices only:**

   Set to file */kernel/drv/scsi_vhci.conf*

```
device-type-scsi-options-list="EMC     SYMMETRIX", "symmetrix-option";
                    symmetrix-option=0x1000000;
```

Note: After *device-type-scsi-options-list=*, there are **five** spaces between *EMC* and *SYMMETRIX*.

CAUTION

MPxIO functionality requires enabling the Symmetrix director C-bit flag.

Host configuration with Sun HBAs

CAUTION

EMC does not support FC-IP on Sun HBAs.

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved HBAs.

The Sun HBAs include the Sun-branded QLogic adapters and Sun-branded Emulex adapters.

The following are Sun-branded QLogic HBAs:

◆ SG-XPCI1FC-QF2	(2 GB single port PCI-X adapter)
◆ SG-XPCI2FC-QF2	(2 GB dual port PCI-X adapter)
◆ SG-XPCI1FC-QF4	(4 GB single port PCI-X adapter)
◆ SG-XPCI2FC-QF4	(4 GB dual port PCI-X adapter)
◆ SG-XPCIE1FC-QF4	(4 GB single port PCI Express adapter)
◆ SG-XPCIE2FC-QF4	(4 GB dual port PCI Express adapter)

The following are Sun-branded Emulex HBAs:

◆ SG-XPCI1FC-EM2	(2 GB single port PCI-X adapter)
◆ SG-XPCI2FC-EM2	(2 GB dual port PCI-X adapter)
◆ SG-XPCI1FC-EM4	(4 GB single port PCI-X adapter)
◆ SG-XPCI2FC-EM4	(4 GB dual port PCI-X adapter)
◆ SG-XPCIE1FC-EM4	(4 GB single port PCI Express adapter)
◆ SG-XPCIE2FC-EM4	(4 GB dual port PCI Express adapter)

EMC has qualified and supports Sun HBAs with Sun StorEdge SAN Foundation Software (also known as Leadville Stack driver). The Leadville Stack driver is embedded in the Solaris 10 x86 operating system.

- The Solaris 10 x86 HW2 is a minimum OS version that has been qualified for Sun branded QLogic 2 GB adapters.
- The Solaris 10 x86 Update 1 operating system with patch 119131-16 is a minimum version that has been qualified for Sun branded QLogic 4 GB adapters and Sun-branded Emulex 2 GB adapters.
- The Solaris 10 x86 Update 1 operating system with patch 120223-06 and 119131-16 is a minimum version that has been qualified for Sun-branded Emulex 4 GB adapters.

To install the EMC-qualified Sun HBAs into the Solaris x86 host and configure the host connection to the EMC storage array and for specific instructions on setting up that particular hardware, follow the installation guide that came with the HBA. If you need a copy of this guide, it can be obtained from the Sun website:

```
http://www.sun.com/products-n-solutions/hardware/docs/
   Network_Storage_Solutions/Adapters
```

Host configuration with Emulex HBAs

CAUTION

EMC does not support FC-IP on Emulex HBAs.

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved HBAs.

Sun AMD64 based servers support Emulex 2 GB/4 GB legacy adapters:

LP10000-E	(2 GB single port PCI-X adapter)
LP10000DC-E	(2 GB dual port PCI-X adapter)
LP11000-E	(4 GB single port PCI-X adapter)
LP11002-E	(2 GB dual port PCI-X adapter)
LPe11000-E	(4 GB single port PCI Express adapter)
LPe11002-E	(4 GB dual port PCI Express adapter)

Emulex legacy 2 GB/4 GB adapters are driven by the **emlxs** device driver. The **emlxs** driver is a part of the Sun StorEdge SAN Foundation Software (also known as Leadville stack driver). This SAN is embedded in the Solaris 10 Update 1 x86 operating system.

- The Solaris 10 x86 Update 1 operating system with patch 119131-16 is a minimum version that has been qualified for Emulex 2 GB adapters.
- The Solaris 10 x86 Update 1 operating system with patch 120223-06 and 119131-16 is a minimum version that has been qualified for Emulex 4 GB adapters.

If you intend to use Solaris 10 x86 prior to S10 Update 1 x86, there are two packages, SUNWemlxs and SUNWemlxu, that are required before installing required patch 120223-xx (refer to the *EMC Support Matrix* for the approval revision). These packages are available on the Sun website:

```
http://www.sun.com/download/products.xml?id=42c4317d
```

To install the EMC-qualified Emulex HBAs into the Solaris x86 host and configure the host connection to the EMC storage array and for specific instructions on setting up that particular hardware, follow the installation guide that came with the HBA. If you need a copy of this guide, it can be obtained from the Sun website:

```
http://www.sun.com/products-n-solutions/hardware/docs/
   Network_Storage_Solutions/Adapters
```

Host configuration with QLogic HBAs

CAUTION

EMC does not support FC-IP on QLogic HBAs.

Note: Refer to the *EMC Support Matrix* for the most up-to-date approved HBAs.

Sun AMD64 based servers support QLogic 2 GB/4 GB legacy adapters:

- QLA2340-E-SP (2 GB single port PCI-X adapter)
- QLA2342-E-SP (2 GB dual port PCI-X adapter)
- QLA2460-E-SP (4 GB single port PCI-X adapter)
- QLA2462-E-SP (4 GB dual port PCI-X adapter)
- QLE2460-E-SP (4 GB single port PCI Express adapter)
- QLE2462-E-SP (4 GB dual port PCI Express adapter)

QLogic legacy 2/4 GB adapters are driven by the **qlc** device driver. The **qlc** driver is a part of the *Sun StorEdge SAN Foundation Software* (also known as Sun SAN). This SAN is embedded in the Solaris 10 x86.

The Solaris 10 x86 Update 1 Operating System with patch 119131-16 is a minimum version that has been qualified for QLogic legacy 2/4 GB adapters.

To install the EMC-qualified QLogic HBAs into the Solaris x86 host and configure the host connection to the EMC storage array and for specific instructions on setting up that particular hardware, follow the installation guide that came with the HBA. If you need a copy of this guide, it can be obtained from the Sun website:

```
http://www.sun.com/products-n-solutions/hardware/docs/
   Network_Storage_Solutions/Adapters
```

12

Solaris x86 and Symmetrix over iSCSI

This chapter contains Symmetrix Multi-Protocol Channel Director (MPCD) iSCSI connectivity implementation details for the Sun Solaris x86 iSCSI software initiator kernel mode driver.

Hardware

Symmetrix iSCSI multiprotocol channel director (MPCD) is supported with Sun Gigabit Network Interface Cards (NIC) in the direct connect and the IP Switch environments.

Refer to the "iSCSI via Symmetrix Multi-Protocol Channel Director" section in Appendix A of the the *EMC Networked Storage Topology Guide* (available on http://Powerlink.EMC.com) for further information on the supported topologies.

Software

Sun iSCSI driver embedded in the Solaris 10_x86 01/06 or later. The iSCSI driver is included of two packages:

- SUNWiscsir - Sun iSCSI device driver
- SUNWiscsiu - Sun iSCSI management utilities

Addressing

Sun uses SCSI-2 device access protocol in addressing iSCSI devices, up to 256 (0 to 255) LUNs per network interface port.

Configuring Solaris iSCSI initiators

Refer to the Sun document *System Administration Guide* (available on http://docs.sun.com/app/docs/doc/819-2723?q=iscsi) to configure the Solaris iSCSI initiators.

Configuring Symmetrix iSCSI director

Refer to the section "Fibre Bit Settings" under "Symmetrix DMX Series" in the *EMC Support Matrix* for the recommended director bit setting for Sun servers.

Solaris x86 iSCSI/Symmetrix case studies

The following are two basic case studies that incorporate information of the Symmetrix iSCSI MPCD and Solaris x86 iSCSI host configurations.

Case study 1

Figure 20 show GigE Network adapters connecting directly to the iSCSI MPCD ports.

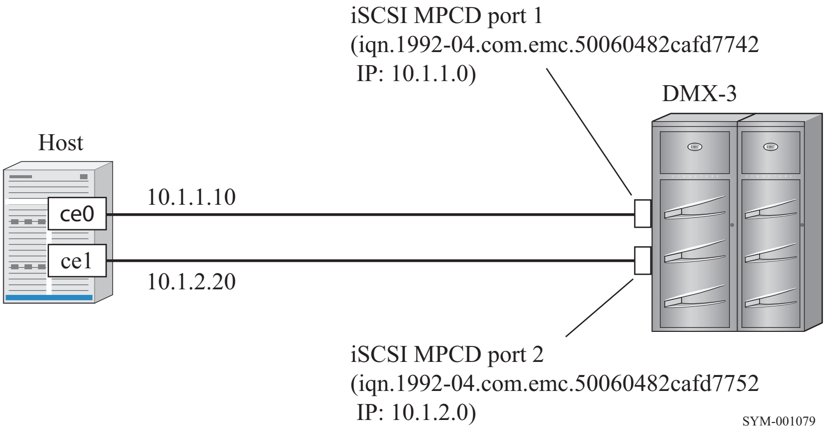


Figure 20 **Connection directly to iSCSI MPCD ports**

Case study 2

Figure 21 on page 226 shows GigE Network adapters connecting to the iSCSI MPCD ports via the IP Switch.

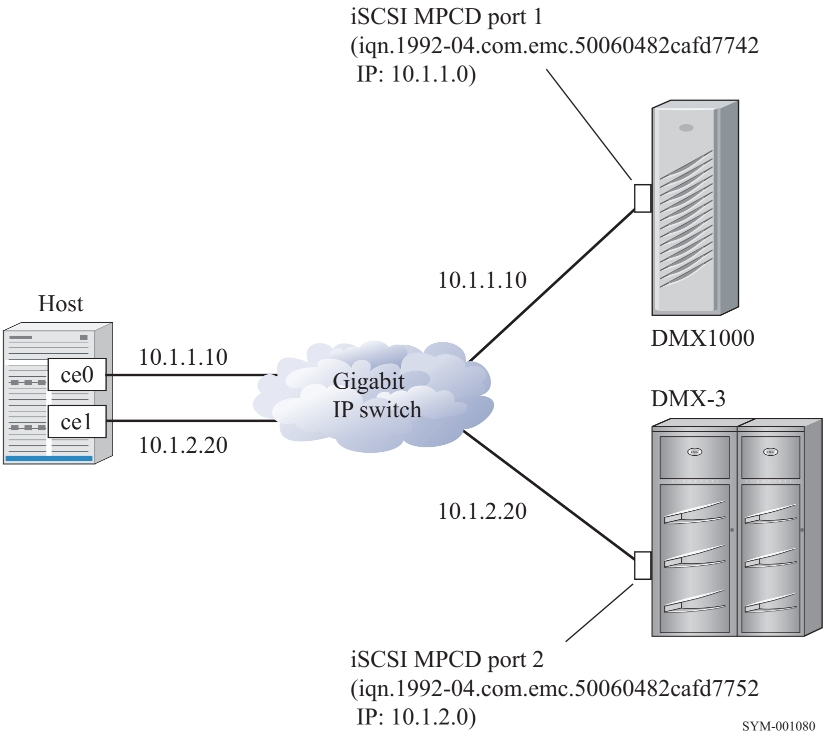


Figure 21 **Connection to iSCSI MPCD ports via IP switch**

Symmetrix configuration

"Case study 1" on page 225 and "Case study 2" on page 225 have the same iSCSI MPCD Channel Information settings.

1. Set "Primary IP Address" on the same subnet with the GigE Network adapters:

   Port 1: 10.1.1.0
   Port 2: 10.1.2.0

2. Set "Max Transmission":

   Port 1: 1500 (default)
   Port 2: 1500 (default)

3. Set "IP Mask" as same as the GigE Network adapters IP mask:

   Port 1: IP Mask = 255.255.255.0
   Port 2: IP Mask = 255.255.255.0

4. Set "IP DNS Group":

   Port 1: NONE (default)
   Port 2: NONE (default)

5. Set "SNMP":

   Port 1: YES (default)
   Port 2: YES (default)

6. Set "Default Gateway":

   Port 1: 0.0.0.0
   Port 2: 0.0.0.0

7. Set "ISNS IP Address":

   Port 1: 0.0.0.0
   Port 2: 0.0.0.0

Sun host configuration

"Case study 1" on page 225 and "Case study 2" on page 225 have the same host settings.

1. Enable network interface for each GigE Network adapter:

   ```
   # ifconfig ce0 plumb
   # ifconfig ce1 plumb
   ```

2. Set IP for each interface:

   ```
   # ifconfig ce0 10.1.1.10 netmask 255.255.255.0 up
   # ifconfig ce1 10.1.2.20 netmask 255.255.255.0 up
   ```

3. Add netmask value for the interfaces to the file /etc/inet/netmasks:

   ```
   10.1.1.0 255.255.255.0
   10.1.2.0 255.255.255.0
   ```

4. Add IP address of each interface to the file /etc/hosts:

   ```
   10.1.1.10 iSCSI0
   10.1.2.20 iSCSI1
   ```

5. Create host network file for each interface port:

   ```
   /etc/hostname.ce0  contains iSCSI0
   /etc/hostname.ce1  contains iSCSI1
   ```

6. You can use the static discovery method or SendTargets device discovery method:

   - Configure the static target discovery method:

     ```
     # iscsiadm add static-config
     iqn.1992-04.com.emc.50060482cafd7742,10.1.1.0:3260
     ```

     ```
     # iscsiadm add static-config
     iqn.1992-04.com.emc.50060482cafd7752,10.1.2.0:3260
     ```

   - Configure the SendTargets device discovery method:

     ```
     # iscsiadm add discovery-address 10.1.1.0:3260
     ```

     ```
     # iscsiadm add discovery-address 10.1.2.0:3260
     ```

7. Enable the iSCSI target discovery method

   - If you have configured the static discovery method, enable the static target discovery:

     ```
     # iscsiadm modify discovery –s enable
     ```

   - If you have configured the SendTargets discovery method, enable the SendTargets discovery:

     ```
     # iscsiadm modify discovery –t enable
     ```

CAUTION

You can only enable one discovery method at a time. If both SendTarget and Static discovery methods are enabled at the same time that may cause the host to PANIC.

8. Reboot the host with reconfigure for the changes to take effect:

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

9. If the host isn’t detected to any iSCSI devices, use the following command to create iSCSI device nodes:

   ```
   # devfsadm –i iscsi
   ```

Index