SPARCcluster™ High Availability™ Server Software Administration Guide

Sun

The Network Is the Computer™

Sun Microsystems Computer Company
2550 Garcia Avenue
Mountain View, CA 94043 USA
415 960-1300   fax 415 969-9131

Part No.: 802-3511-10
Revision A, October 1995

THIS PUBLICATION IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS PUBLICATION COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN, THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THE PUBLICATION. SUN MICROSYSTEMS, INC. MAY MAKE IMPROVEMENTS AND / OR CHANGES IN THE PRODUCT(S) AND / OR THE PROGRAMS(S) DESCRIBED IN THIS PUBLICATION AT ANY TIME.

Adobe PostScript

Contents

Figures

Preface

Solstice™ High Availability™ 1.0 (Solstice HA) is a software product that supports specific dual-server hardware configurations. It is compatible with the Solaris™ 2.4 software environment. When configured properly, the hardware and software together provide highly available data services. Solstice HA depends upon the mirroring and diskset capabilities and other functionality provided by Solstice DiskSuite™ 4.0, which is shipped with Solstice High Availability.

The *SPARCcluster High Availability Server Software Administration Guide* documents the software procedures to be followed when adding or replacing hardware, along with general software administration procedures.

Who Should Use This Book

The information in this manual is for Sun representatives and senior system administrators who have experience performing complex administrative procedures. The instructions and discussions in this manual are complex and intended for a technically advanced audience.

The instructions in this manual and the *SPARCcluster High Availability Software Planning and Installation Guide* assume the reader has expertise with the Solstice DiskSuite product.

In some instances, the manual will tell you to contact your support person rather than provide instructions for fixing a problem. This is done to help ensure the availability of the data services running on your Solstice High Availability servers.

System administrators with at least six years of UNIX® system experience will find this book useful when performing administrative tasks on Solstice High Availability configurations. This manual should be used with the *SPARCcluster High Availability Software Planning and Installation Guide* and the other manuals that are listed in "Related Documentation" on page xv.

Note – Junior and less experienced system administrators should not attempt to install, configure, or administer Solstice HA configurations.

How This Book Is Organized

This document has 10 chapters and two appendixes, as follows:

Chapter 1, "Introduction," introduces Solstice HA administration.

Chapter 2, "Preparing for Administration," offers a high-level overview of the functionality included with Solstice HA. The interactions between the various parts of Solstice HA are also discussed.

Chapter 3, "Monitoring the Solstice HA Servers," discusses the use of the `hastat(1M)` and `haload(1M)` commands to monitor the behavior of the systems.

Chapter 4, "Hardware Replacement and Repair," provides the software instructions to use when replacing or repairing hardware components such as local disks and cables.

Chapter 5, "Adding Hardware," tells the software procedure to follow when adding additional hardware such as disks, controllers, SBus cards, and network connections to Solstice HA configurations.

Chapter 6, "HA-NFS Maintenance," explains the maintenance procedures you will follow when migrating data to HA-NFS, administering HA-NFS, and working with UFS logs.

Chapter 7, "HA-ORACLE Maintenance," explains the maintenance procedures you will follow when administering an HA-DBMS.

Chapter 8, “Metadevice and Diskset Administration,” provides instructions for administering shared metadevices and explains metadevice actions that are not supported in Solstice HA configurations.

Chapter 9, “General Solstice HA Maintenance,” gives instructions for general maintenance procedures such as restarting failed systems in Solstice HA configurations.

Chapter 10, “Using the Terminal Concentrator,” tells how to use the terminal concentrator when performing administration of Solstice HA configurations.

Appendix A, “Error Messages,” explains the status, error, and log messages displayed by Solstice HA.

Appendix B, “Man Pages,” is a printed copy of the man pages for the Solstice HA product. Online versions of the man pages are included on the CD-ROM.

Related Documentation

Use the following manuals with the *SPARCcluster High Availability Server Software Administration Guide*:

- *SPARC: Installing Solaris Software*
- *SPARCcluster High Availability Software Planning and Installation Guide* (part number 802-3509*)*
- *SPARCcluster High Availability Server Hardware Planning and Installation* (part number 802-3510)
- *SPARCcluster High Availability Server Service Manual* (part number 802-3512)
- *Solstice DiskSuite 4.0 Administration Guide*
- *Solstice DiskSuite Tool 4.0 User’s Guide*
- *NFS Administration Guide*
- *SMCC NFS Server Performance and Tuning Guide* (part number 801-7289)
- *TCP/IP and Data Communications Administration Guide*
- *Name Services Administration Guide*
- *Name Services Configuration Guide*
- *SPARCserver 1000 System Installation Manual* (part number 801-2893)

- *SPARCserver 1000 System Service Manual* (part number 801-2895)
- *SPARCcenter 2000 Installation Manual* (part number 801-6975)
- *SPARCcenter 2000 Service Manual* (part number 801-2007)
- *SPARCstorage Array Model 100 Installation Manual* (part number 801-2205)
- *SPARCstorage Array Model 100 Service Manual* (part number 801-2206)
- *Multi-Disk Pack Installation and Service Manual (*part number 801-6119)
- *Memory Module Installation Guide* (part number 801-2030)
- *Solaris 2.x Handbook for SMCC Peripherals* (part number 801-5488)
- *FSBE/S SBus Card Manual* (part number 800-7508)
- *Installing SBus Cards in Deskside and Data Cabinet Systems* (part number 800-6366)
- *SunFast Ethernet Adapter User Guide* (part number 801-6109)
- *Fibre Channel SBus Card Installation Manual* (part number 801-6313)
- *Fibre Channel Optical Module Installation Manual* (part number 801-6326)
- *Oracle7® Server Administrator's Guide*
- *Oracle7 for Sun SPARC Solaris 2.x Installation and Configuration Guide*
- *Terminal Concentrator Binder Set*

What Typographic Changes Mean

The following table describes the typographic changes used in this book.

Table P-1 Typographic Conventions

Typeface or Symbol	**Meaning**	**Example**
`AaBbCc123`	The names of commands, files, and directories; on-screen computer output	Edit your `.login` file. Use `ls -a` to list all files. `machine_name% You have mail.`

`AaBbCc123`	What you type, contrasted with on-screen computer output	`machine_name%` **`su`**
		`Password:`

AaBbCc123	Command-line placeholder: replace with a real name or value	To delete a file, type `rm` *filename.*
AaBbCc123	Book titles, new words or terms, or words to be emphasized	Read Chapter 6 in *User's Guide.* These are called *class* options. You *must* be root to do this.

Shell Prompts in Command Examples

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

Table P-2 Shell Prompts

Shell	**Prompt**
C shell prompt	`machine_name%`
C shell superuser prompt	`machine_name#`
Bourne shell and Korn shell prompt	`$`
Bourne shell and Korn shell superuser prompt	`#`

Introduction 1

This chapter offers a high-level overview of the functionality included with Solstice HA. The interactions between the various parts of Solstice HA are also discussed.

Use the following table to locate specific information in this chapter:

1.1 Solstice HA Overview

Solstice HA is an unbundled software product that supports specific dual-server hardware configurations. When properly configured, the hardware and software together provide highly available data services.

Solstice HA provides an environment in which data services remain available after any single hardware or software point of failure has occurred in the configuration.

The hardware configuration is called SPARCcluster High Availability, while the software is referred to as a Solstice High Availability. The configuration includes two servers, multi-host disks, Solstice DiskSuite software, and the Solstice HA software. The data services available include Solstice HA-NFS and Solstice HA-ORACLE.

1.1.1 Hardware Overview

Each server in a Solstice HA configuration has two or more disks which are accessible only from that server. These are called *local disks*. They contain the Solaris software environment, the Solstice HA packages, and optionally other local data.

Disks in the configuration that are accessible from either of the servers are called *multi-host disks*. Multi-host disks are organized into one or two *disksets* during configuration. These disksets contain the data (information) for highly available services.

The servers in the configuration communicate via two private network connections. Solstice HA configuration and status information is communicated across these links.

The servers also have one or more public network connections that provide communication to clients of the highly available services.

The servers are referred to as being *siblings* of each other.

1.1.2 Software Overview

The Solstice HA and Solstice DiskSuite packages and the Solaris 2.4 distribution are installed on both servers in the configuration.

The Solstice HA software has the following components:

- Membership monitor
- Fault monitor
- Programs used by the membership monitor and fault monitor
- Various administration commands
- Solstice DiskSuite software package

The membership monitor, fault monitor, and associated programs allow one server to *take over* when hardware or software fails.

When a takeover occurs, the server assuming control becomes the I/O *master* for the failed server's disksets and redirects the clients of the failed server to itself. The takeover also includes actions that are specific to the HA-NFS and HA-ORACLE data services.

Administrators can use the same mechanism to manually direct one server to take over the data services for the sibling server. This is referred to as *switchover* and is performed using the `haswitch(1M)` command.

A switchover allows administrators to take a server offline for maintenance and to bring a previously offline server back online.

Solstice DiskSuite is required for Solstice HA operations. Solstice DiskSuite provides the following functionality:

- Diskset management
- Disk mirroring
- Disk concatenation
- Disk striping
- Hot spare pool device management
- UNIX file system logging
- Management of metadevice status and configuration of database replicas

After the Solaris software environment is installed, Solstice DiskSuite and Solstice HA software are installed on each server's local disks. The `hasetup(1M)` command provides an interface to set up the configuration. Either Solstice DiskSuite commands or the Solstice DiskSuite Tool (`metatool(1M)`) graphical user interface can then be used to create concatenations, stripes, mirrors, hot spares, and UFS logs. Solstice DiskSuite Tool provides an interface that makes large disk configurations easier to manage.

Administrators will use the `hastat(1M)`, `haload(1M)`, and `metastat(1M)` commands to monitor the status of the Solstice High Availability configuration and the Solstice DiskSuite metadevices.

1.2 Elements of Solstice HA

The Solstice HA product consists of a set of programs which provide the ability to:

- Monitor the Solstice HA configuration
- Configure the software in the Solstice HA configuration
- Monitor the services running on Solstice HA servers
- Monitor the status of the Solstice HA configurations
- Manipulate disksets

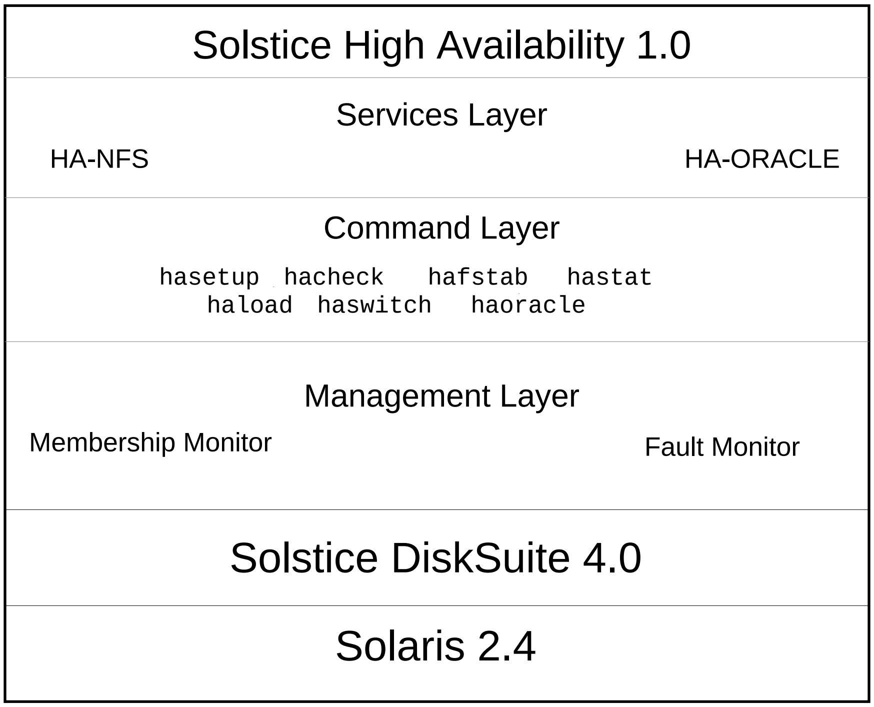

Figure 1-1 Diagram of the Solstice HA Elements

Figure 1-1 illustrates how Solstice HA fits on top of Solstice DiskSuite and Solaris 2.4. Discussions of each of these elements are given in the following subsections.

1.2.1 Solstice High Availability

Solstice High Availability is a software and hardware package that enables two servers to act as a highly available data facility. Solstice HA is built on Solstice DiskSuite, which provides the mirroring, concatenation, stripes, hot spares, UFS logging.

Each Solstice HA server acts as an I/O master for its respective diskset and runs data services that export data on that diskset.

In a symmetric configuration, each server is also a backup for the sibling server's data services. Solstice HA provides programs used by each server to monitor the status of data services running on itself as well as the data services running on the sibling machine in the configuration.

Solstice HA automates the decision to take over when the sibling server has a software or hardware failure. Takeover processing includes common actions, such as assuming I/O mastery of the failed server's diskset and redirecting the failed server's clients to itself. Takeover also includes actions that are specific to the data service.

Solstice HA additionally provides for administratively initiated switchover, which is the graceful switch of a diskset from one functional server to the sibling to reconfigure or bring a server back online.

Solstice HA is broken into three major layers: service, command, and management, as explained in the following three subsections.

1.2.1.1 Service Layer

Solstice HA supports two data services, HA-NFS and HA-ORACLE. The HA-ORACLE is version 7.0 relational database management system (DBMS).

1.2.1.2 Command Layer

Solstice HA provides utilities for configuring and administering the highly available data facility. These utilities allow:

- Configuring of the two Solstice HA servers (`hasetup(1M)`)
- Checking the configuration to ensure high availability (`hacheck(1M)`)
- Editing the `vfstab.`*logicalhost* and `dfstab.`*logicalhost* files (`hafstab(1M)`)

- Monitoring the status of the configuration (`hastat(1M)`)
- Transferring the data services from one server in the configuration to the sibling (`haswitch(1M)`)
- Monitoring the load on the servers (`haload(1M)`)
- Verify the HA-ORACLE installation (`haoracle(1M)`)

1.2.1.3 Management Layer

The Solstice HA management layer includes the *membership monitor* and the fault monitor.

The membership monitor detects which of the two servers in the Solstice HA configuration is running and which of the two servers has failed.

The principal functions of the membership monitor are to make sure the servers are in sync and to coordinate the configuration of the applications and services when the state of the configuration changes.

The membership monitor provides the following features:

- Reliability. No single point of failure in the servers can cause membership monitor failure.
- Fault detection. Detection of a server crash within the Solstice HA configuration.
- Server removal. Removal of the failed server from the Solstice HA configuration using a reliable fail-fast mechanism.

While the membership monitor detects total failure of a system in the Solstice HA configuration, the fault monitor detects failures of individual services.

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

- Probes of both the public and private networks
- Probes of both the local and remote exported NFS service
- Probes of both the local and remote ORACLE service

If the probe detects a service failure, the fault monitor may try to restart the service. If the service does not restart, the fault monitor probe initiates a takeover.

For HA-NFS service, the fault monitor checks the availability of each of the exported highly available NFS file systems.

Under certain circumstances the fault monitor will not initiate a takeover even though there has been an interruption of a service. These interruptions can include:

- The exported NFS file system is being checked with `fsck(1M)`.
- The NFS file system is locked via `lockfs(1M)`.
- The name service is not working. Because client HA-NFS depends on the name service database (NIS or NIS+), the HA-NFS services are only as reliable as the name service. The name service exists outside the Solstice HA configuration so you must take necessary measures to ensure its reliability. These measures may include use of uninterruptable power supply (UPS) on the name service servers. Refer to the *SPARCcluster High Availability Server Service Manual* for additional information. Because of the changes to `/etc/nsswitch.conf`, the server side of HA-NFS has a network name service dependency only on `netgroup`.

Note – Do not change any of the programs or files associated with the fault monitor daemon or probe. You can, however, change some of the parameters using Solstice HA commands.

1.2.2 Solstice DiskSuite

Solstice DiskSuite 4.0 is a software package that offers a metadisk driver and several UNIX file system enhancements that provide better performance, greater capacity, and improved availability.

The metadisk driver is the basic element of the Solstice DiskSuite product. This driver is implemented as a set of loadable, pseudo device drivers. The metadisk driver uses other physical device drivers to pass I/O requests to and from the underlying devices.

An overview of the metadisk driver elements is presented in the following subsections. For a complete discussion, refer to the *Solstice DiskSuite 4.0 Administration Guide*, which is included with the Solstice High Availability product.

1.2.2.1 Metadevices

Metadevices are the basic functional unit of the metadisk driver. After you create metadevices, you can use them like physical disk slices. These metadevices devices can be made up of one or more slices. The metadevices may be configured to use a single device, a concatenation of stripes, or stripe of devices.

1.2.2.2 Metadevice State Database Replicas

Metadevice state database replicas provide the nonvolatile memory necessary to keep track of configuration and status information for mirrors, submirrors, concatenations, stripes, UFS logs, and hot spares. The replicas also keep track of error conditions that have occurred. A majority of metadevice state database replicas must be preserved in the event a disk or SPARCstorage Array chassis fails.

The replicas are automatically placed on disks in the disksets by the `metaset` command. `metaset` places replicas on disks in a diskset. You must have at least three controllers attached to a server or you are at risk of not having enough replicas to have a majority of replicas in the event of a controller failure.

1.2.2.3 Disksets

A diskset is a pair of hosts and disk drives in which all the drives must be accessible by both hosts. There are one or two disksets in a Solstice HA configuration.

Only one server can master a diskset at any point in time. There is one metadevice state database per diskset and one per local diskset. You are instructed to create three metadevice state database replicas on the local disks during installation and configuration of Solstice HA. Numerous replicas are automatically placed on the disks in each diskset. The number and placement of the replicas on disks in the disksets is automatically determined by the `metaset(1M)` command.

1.2.2.4 Concatenations and Stripes

Each metadevice is either a concatenation or a stripe of slices. Concatenations and stripes work much the way the `cat(1)` command is used to concatenate two or more files together to create one larger file. When slices are concatenated, the addressing of the component blocks is done on the components sequentially. The file system can use the entire concatenation.

Striping is similar to concatenation except the addressing of the metadevice blocks is interlaced on the components, rather than addressed sequentially. When stripes are defined, an interlace size may be specified. The interlace size is a number followed by `k` for kilobytes, `m` for megabytes, or `b` for 512-byte blocks (for example, `8m`, `16k`, or `512b`).

1.2.2.5 Mirrors

All multi-host data must be placed on mirrored metadevices. This is necessary for the server to tolerate single-component failures.

To set up mirroring, you first create a metamirror. A metamirror is a special type of metadevice made up of one or more other metadevices. Each metadevice within a metamirror is called a submirror.

1.2.2.6 Hot Spares

The hot spare facility enables automatic replacement of failed submirror components, provided that a suitable spare component is available and reserved. Hot spares are temporary fixes, used until failed components are either repaired or replaced. Hot spares provide further security from downtime due to hardware failures.

1.2.2.7 UNIX File System Logging

UFS logging records UFS updates in a log before the updates are applied to the file system. UFS logging speeds up reboots, provides faster local directory operations, and decreases the time necessary for synchronous disk writes.

UFS logging eliminates file system checking at boot time because changes from unfinished system calls are discarded. A pseudo device, called the trans device, manages the contents of the log. Deciding on the placement of the log on multi-host disks in Solstice HA configurations is very important, because selecting the wrong location can decrease performance.

When using UFS logs in Solstice HA configurations, follow these guidelines:

- Set up one log per file system. Logs should not be shared between file systems.
- If you have heavy writing activity on a file system, use separate disks for the log and master.
- The recommended size for a UFS logs is one Mbyte per 100 Mbytes of file system size (one percent). The maximum useful log size is 64 Mbytes, for file systems larger than 6.4 Gbytes.

1.3 Solstice HA Commands

This subsection describes the commands associated with Solstice HA. A printed copy of each man page is provided in Appendix B, "Man Pages."

- `hacheck(1M)` – Validates Solstice HA configurations. This program ensures the configuration has been set up correctly and that the software and hardware provide highly available data services.
- `hafstab(1M)` – Provides a method of editing and distributing `dfstab(4)` and `vfstab(4)` files to the two servers in a Solstice HA configuration.
- `haload(1M)` – Monitors the load on the pair of Solstice HA servers. Monitoring is necessary because there must be some excess capacity between the two servers. If there is no excess capacity and a takeover occurs, the remaining server will be unable to take care of the combined workload.
- `haoracle(1M)` – Verifies the HA-ORACLE installation. The `haoracle` command also used to maintain the list of monitored databases in the HA-ORACLE configuration file, `haoracle_databases(4)`.
- `hasetup(1M)` – Provides an interface that allows initial configuration of the Solstice HA servers. The information entered on one of the Solstice HA servers is automatically updated on the other server. `hasetup` first attempts to discover most information about the configuration without user input.

You are asked about additional public network names, the type of configuration (symmetric or asymmetric), the data services being used, space for UFS logging, and placement of disks in disksets. The program then updates the Solstice HA configuration files with the information.

- `hastat(1M)` – Displays the current state of the Solstice HA configuration. In the default mode, the status of all the components of the configuration is displayed only once; the program then exits.
- `haswitch(1M)` – Transfers the specified diskset along with its associated data services and IP addresses to the specified sibling server.

You also use Solstice DiskSuite commands when performing administration procedures on Solstice HA configurations. These man pages are included with the Solstice DiskSuite distribution. Printed copies of the man pages can be found in the *Solstice DiskSuite 4.0 Administration Guide*.

1.4 System Files Associated With Solstice HA

There are several system files associated with Solstice HA. You can edit the `md.tab`, `vfstab.`*logicalhost*, and `dfstab.`*logicalhost* files. Do not edit the other files.

- `/etc/opt/SUNWmd/md.tab` – This file is used by the `metainit` and `metadb` commands as an optional input file. Each metadevice must have a unique entry in the file. Tabs, spaces, comments (using the pound sign (#) character), and line continuations (using the backslash (\) character) can be used in the file.

Note – The `md.tab` file is not automatically updated when the configuration is changed, so you must manually change the `md.tab` file. This manual does provide instructions and advice on using the `md.tab` file for initial configuration of Solstice DiskSuite metadevices.

- `/etc/opt/SUNWhadf/hadf/vfstab.`*logicalhost* – In a symmetric configuration, there are two instances of this file, one for each logical host. An asymmetric configuration has only one instance of this file. The `vfstab.`*logicalhost* files list the file systems mounted for the logical host.

- `/etc/opt/SUNWhadf/nfs/dfstab.`*logicalhost* – In a symmetric configuration there are two instances of this file, one for each logical host. An asymmetric configuration has only one instance of this file. This file is present only if you are running HA-NFS.
- `/etc/opt/SUNWhadf/hadf/cmm_confcdb` – This file contains configuration information for the membership monitor. Among other things, it identifies the two hosts of a Solstice HA configuration, private network connections, and membership monitor states and transitions.
- `/usr/opt/SUNWhadf/hadf/hadfconfig` – This file is read by the reconfiguration programs as part of the initial step of membership reconfiguration.
- `/`*logicalhost*`/statmon` – The `statmon` directory contains files that record lock manager and status monitor states when HA-NFS is running on a logical host.

Preparing for Administration 2

This chapter offers advice about preparing for administration of a Solstice High Availability configuration.

Use the following table to locate specific information in this chapter:

2.1 Saving Device Information

It is useful to have a record of the disk partitioning you have selected for the disks in your multi-host disksets. This multi-host partitioning information is needed when a disk replacement must be made.

The disk partitioning information for the local disks is not as critical because the local disks on both servers should have been partitioned identically. You can obtain the local disk partitioning from the sibling server in the event a local disk fails.

When a multi-host disk is replaced, the replacement disk must have the same partitioning as the disk it is replacing. Depending on how a disk has failed this information may or may not be available when replacement is performed. This is especially important if you have several different partitioning schemes in your disksets.

A simple way of recording this information is shown in the example script in Figure 2-1. This type of script should be run following Solstice HA configuration. In this example, the files containing the volume table of contents (VTOC) information are written to the `/etc/opt/SUNWhadf/vtoc` directory by the `prtvtoc(1M)` command.

```
#! /bin/sh
DIR=/etc/opt/SUNWhadf/vtoc
mkdir -p $DIR
cd /dev/rdsk
for i in *s7
do prtvtoc $i >$DIR/$i || rm $DIR/$i
done
```

Figure 2-1 Example Script for Saving VTOC Information

The script in Figure 2-1 will work because it is a requirement that each of the disks in a diskset has a slice 7, where the metadevice state database replicas reside.

If a local disk also has a valid slice 7, the VTOC information will also be saved by the example script in Figure 2-1. However, this should not occur for the boot disk, because typically a local disk does not a valid slice 7.

Note – Make certain that the script is run while none of the disks are owned by the sibling host. The script will work if the logical hosts are in maintenance mode (`MAINT`), the logical hosts owned by the local host, or if Solstice HA is not running.

You should duplicate this information on both servers in the Solstice HA configuration.

It is important this that this information be kept up to date as new disks are added to the disksets and when any of the disks are repartitioned.

2.2 Restoring Device Information

If you have saved the VTOC information for all multi-host disks, this information can be used when a disk is replaced. The example script shown in Figure 2-2 would use the VTOC information saved when you ran the script shown in Figure 2-1 to give the replacement disk the same partitioning as the failed disk.

```
#! /bin/sh
DIR=/etc/opt/SUNWhadf/vtoc
cd /dev/rdsk
for i in c1t0s0s7
do fmthard -s $DIR/$i $i
done
```

Figure 2-2 Example Script for Restoring VTOC Information

Note – The replacement drive must be of the same size and geometry (generally the same model from the same manufacturer) as the failed drive. Otherwise the original VTOC may not be appropriate for the replacement drive.

If you have failed to record this VTOC information, but you have mirrored slices on a disk by disk basis (for example, the VTOCs of both sides of the mirror are the same), it is possible to simply copy the VTOC information from the other mirror disk to the replacement disk. An example of how to perform this procedure is shown inFigure 2-3.

```
#! /bin/sh
cd /dev/rdsk
OTHER_MIRROR_DISK=c1t0d0s7
REPLACEMENT_DISK=c2t0d0s7
prtvtoc $OTHER_MIRROR_DISK | fmthard -s - $REPLACEMENT_DISK
```

Figure 2-3 Example Script to Copy VTOC Information From a Mirror

If you failed to save the VTOC information and did not mirror on a disk-by-disk basis, it is possible to examine the component sizes reported by the metastat(1M) command and reverse engineer the VTOC information. Because the computations used in this is a procedure are so complex, the procedure should only be performed by a trained service representative.

2.3 Recording the path_to_inst Information

It is important for you to record the /etc/path_to_inst information on removable media (that is, floppy disk or backup tape). The path_to_inst(4) file contains the minor unit numbers for disks in each SPARCstorage Array. This information will be necessary if the boot disk on either Solstice HA server fails and has to be replaced.

2.4 Instance Numbering

Instance names are occasionally reported in driver error messages. An instance name refers to the system devices such as ssd20 or le5.

You can determine the binding of an instance name to a physical name by looking at /var/adm/messages or dmesg(1M) output:

```
ssd20 at SUNW,pln0:
ssd20 is /io-unit@f,e0200000/sbi@0,0/SUNW,soc@3,0/SUNW,pln@a0000800,20183777/ssd@4,0

le5 at lebuffer5:  SBus3 slot 0 0x60000 SBus level 4 sparc ipl 7
le5 is /io-unit@f,e3200000/sbi@0,0/lebuffer@0,40000/le@0,60000
```

Once an instance name has been assigned to a device, it remains bound to that device.

Instance numbers are encoded in a device’s minor number. To keep instance numbers persistent across reboots, the system records them in the /etc/path_to_inst file. This file is read only at boot time, and is currently updated by the add_drv(1M) and drvconfig(1M) commands. For additional information refer to the path_to_inst man page.

When you perform a Solaris installation on a server, instance numbers can change if hardware was added or removed since the last Solaris installation. For this reason, use caution whenever you add or remove SBus cards on Solstice HA servers. Always install an SBus card in the next available empty SBus slot for that type of device.

The following example highlights the instance number problems that can arise in a configuration. In this example, the Solstice HA configuration consists of three SPARCstorage arrays with FC/S cards installed in SBus slots 1, 2, and 4 on each of the servers. The controller numbers are `c1`, `c2`, and `c3`. If the system administrator adds another SPARCstorage array to the configuration using a FC/S card in SBus slot 3, the corresponding controller number will be `c4`. If Solaris is reinstalled on one of the servers, the controller numbers `c3` and `c4` will refer to different SPARCstorage Arrays. The other Solstice HA server will still refer to the SPARCstorage Arrays with the original instance numbers. Solstice DiskSuite will not communicate with the disks connected to the `c3` and `c4` controllers.

Other problems can arise with instance numbering associated with the Ethernet connections. For example, each of the Solstice HA servers has three Ethernet SBus cards installed in slots 1, 2, and 3 and the instance numbers are `le1`, `le2`, and `le3`. If the middle card (`le2`) is removed and Solaris is reinstalled, the third SBus card will be renamed from `le3` to `le2`.

2.4.1 Reconfiguration Reboots

During some of the administrative procedures documented in this manuals, you are told to perform a reconfiguration reboot. A reconfiguration reboot is performed via the OpenBoot PROM `boot -r` command or by creating the file `/reconfigure` on the server and then rebooting.

Note – It is not necessary to perform a reconfiguration reboot to add disks to an existing SPARCstorage Array.

Be certain to avoid performing Solaris reconfiguration reboots when any hardware (especially a SPARCstorage Array or other disk) is not operational (powered off or otherwise defective). In such situations the reconfiguration reboot will remove the inodes in `/devices` and symbolic links in `/dev/dsk` and `/dev/rdsk` associated with the disk devices. These disks will become inaccessible to Solaris until a later reconfiguration reboot. A subsequent reconfiguration reboot may not restore the original controller minor unit numbering and cause Solstice DiskSuite to reject the disks. When the original numbering is restored, Solstice DiskSuite can access the associated metadevices.

If all hardware is operational a reconfiguration reboot may be safely performed to add a disk controller to a server. Such controllers must be added symmetrically to both servers (though a temporary unbalance is allowed while the servers are upgraded). Similarly, if all hardware is operational it is safe to perform a reconfiguration reboot to remove hardware.

2.5 Solstice HA Processes

The Solstice HA software has several processes running at any time on the two servers.

Caution – Never stop or `kill(1M)` a Solstice HA process unless you are told to do this as part of a maintenance procedure in this manual or in one of the other documents included in the *SPARCcluster Binder Set*.

The process that are running on the servers include the following:

```
/bin/sh /opt/SUNWhadf/bin/haload
/bin/sh /opt/SUNWhadf/clust_progs/runclocksync
/bin/sh /opt/SUNWhadf/fault_progs/faultdloop
/bin/sh /opt/SUNWhadf/fault_progs/net_probe_brother private
/bin/sh /opt/SUNWhadf/fault_progs/net_probe_brother public
/bin/sh /opt/SUNWhadf/fault_progs/nfs_probe_local_restart /var/opt/SUNWhadf/had
/bin/sh /opt/SUNWhadf/fault_progs/nfs_probe_one_common host1 1 0
/bin/sh /opt/SUNWhadf/fault_progs/nfs_probe_one_common host2 0 0
/bin/sh /opt/SUNWhadf/fault_progs/nfs_probe_remote
      /var/opt/SUNWhadf/hadf/ha_env FOREIGN
/bin/sh /opt/SUNWhadf/fault_progs/nfs_probe_local
      /var/opt/SUNWhadf/hadf/ha_env NATIVE
/opt/SUNWcluster/bin/clustd -f /etc/opt/SUNWhadf/hadf/cmm_confcdb
/usr/lib/autofs/automountd
/usr/lib/nfs/lockd -g 90
/usr/lib/nfs/statd -a host1 -p /host1
faultd
fdl_load -i 30 -p 90 host1 host2-priv1
haclksyn host1-priv1 host2-priv2
net_periodic_ping_other 8640000 30 5 120
      /var/opt/SUNWhadf/hadf/tmp/net_probe_brother.up.992
      /var/opt/SUNWhadf/hadf/tmp/net_probe_brother.down.992
net_periodic_ping_other 8640000 30 5 120
      /var/opt/SUNWhadf/hadf/tmp/net_probe_brother.up.1135
      /var/opt/SUNWhadf/hadf/tmp/net_probe_brother.down.1135
nfs_mon host1 host2 1 /var/opt/SUNWhadf/hadf/tmp/nfs_mon.cmd.host1.175
nfs_mon host2 host1 0 /var/opt/SUNWhadf/hadf/tmp/nfs_mon.cmd.host2.198
nfs_monitor_pids -i 10 116 847 862 840 838
scsirstd
```

Caution – Never stop or kill any executable found in the `/opt/SUNWhadf` tree.

2.6 Logging Into the Servers as Root

If you want to log in to Solstice HA servers as root through a terminal other than the console, you must edit the `/etc/default/login` file and comment out the following line:

```
CONSOLE=/dev/console
```

This will allow you to have root logins via `rlogin(1)`, `telnet(1)`, and other programs.

Monitoring the Solstice HA Servers 3

This chapter tells how to use the Solstice HA and Solstice DiskSuite commands to monitor the behavior of a Solstice HA configuration.

Use the following table to locate specific information in this chapter.

3.1 Overview of Solstice HA Monitoring

You will use five utilities in addition to the `/var/adm/messages` files when monitoring the behavior of a Solstice HA configuration. The utilities you use include `hastat(1M)`, `haload(1M)`, `metastat(1M)`, `metadb(1M)`, and `metatool(1M)`.

3.2 Monitoring the Solstice HA Configuration Status

hastat displays the current state of the Solstice HA configuration. The program displays status information about the hosts, logical hosts, private networks, public networks, data services, local disks, disksets, along with the most recent error messages.

An example of output from hastat is shown below:

```
# hastat
Configuration State: Stable
logicalhost1 - Owned by host1
logicalhost2 - Owned by host2

host1 -   1:56pm  up 2 day(s),  4:54,  2 users,  load average: 0.12, 0.09, 0.07
host2 -   1:56pm  up 2 day(s), 5 hr(s),  0 users,  load average: 0.11, 0.09, 0.12

Data Service HA-NFS: logicalhost1 - Unknown; logicalhost2 - Ok

Local metadevices: host1 - (none); host2 - (none)
Local metadb replicas: host1 - Ok; host2 - Ok
Diskset logicalhost1: Ok; MetaDB replicas in logicalhost1: Ok
Diskset logicalhost2: Ok; MetaDB replicas in logicalhost2: Ok

Private nets: Ok
Public nets: host1 - Ok; host2 - Ok

Extract of Message Log (examine /var/adm/messages for the full context):
Aug  9 09:11:14 host1 hadf: ERROR: nfs_mounttouchfile: Failed for host2:/palmer/root/whynot
with exit status 1
. . .
#
```

Figure 3-1 Example hastat Output

The status is reported as follows:

- Ok – The component's status is okay.
- Not Ok – The component is not functioning. For instance, no public networks are responding.
- Degraded – The component is working well enough to provide partial service to some clients, but needs some repair.

- Unknown – There is not enough information about the component to determine the status. For instance, when the sibling host is down, the remaining host will list the private nets as Unknown.

The following list explains the output displayed:

- Solstice HA configuration state – Either Down, Reconfiguring, or Stable. Down says the Solstice HA configuration is not functioning. The string Reconfiguring is displayed when the Solstice HA configuration is in the process of a transition from one state to another, because of a takeover or switchover. Stable says the server is functioning as expected.
- Logical hosts – The names of the logical hosts associated with the two disksets along with the name of the current owner, or the string Maintenance mode if the logical host is currently in the Maintenance state (taken down by an administrator).
- Physical servers – The names of both physical servers in the Solstice HA configuration are displayed with the current time, the length of time the server has been up (in days and hours), the number of users, and the load average over the past 1, 5, and 15 minutes.
- Status of data services – The data services running on which of the logical hosts. For HA-NFS the status is represented as OK, Not OK, or Degraded. For HA-ORACLE the status of each database is reported as running, maintenance, not configured correctly, or stopped. If a data service is not running on a logical host, that logical host is not listed for that data service. Not Ok indicates the data service has failed. If the status is Not OK or Degraded, check the Message Log or the message file (/var/adm/messages) to see if an error has been reported.
- Local metadevices – The status of local Solstice DiskSuite metadevices, reported as Ok, Not Ok, or Unknown. If the status is Not Ok, you should first check the Message Log or messages file (/var/adm/messages) to see if an error has been reported. If one has not, run the metastat(1M) command to discover the problem. If the local file systems are not on metadevices, this field displays a status of none.
- Local metadb replicas – The status of the metadevice state database replicas on the local disks, reported as Ok or Not Ok. If the status is Not Ok, one or more database replicas are inactive. Run the metadb(1M) command for additional information.

- Disksets – The status of the multi-host disksets reported as OK, Not OK, or Unknown. If the status is Not OK, you should first check the Message Log or message file to see if an error has been reported. If one has not, run the metastat -s *diskset* command to discover the problem. If hastat cannot determine the status it is reported as Unknown.
- Private networks – The status of private networks, displayed as either OK, Not OK, Degraded, or Unknown. A status of Not Ok or Degraded indicates a problem with one or both of the private network interfaces. You can check the Message Log or message file (/var/adm/messages) for additional information, or directly troubleshoot the interface for hardware or software faults using command such as ping(1M), swapping cables, or swapping controllers.
- Public networks – The status of public networks, displayed as either OK, Not OK, Degraded, or Unknown. You must check the Message Log or message file (/var/adm/messages) for additional information if the status is Not OK or Degraded.
- Recent error messages – The message log at the bottom of the display lists the last few messages from the /var/adm/messages file.

Note – Because the recent error messages list is a filtered extract of the log messages, the context of some messages may be lost. You should directly examine the /var/adm/messages file for a complete list of the messages.

3.3 Monitoring the Load of the Solstice HA Servers

haload is used to monitor the load on the pair of Solstice HA servers. Monitoring is necessary because there must be some excess capacity between the two servers. If there is no excess capacity and a takeover occurs, the remaining server may be unable to take care of the combined workload.

haload monitors both servers and logs occurrences of an overload. The administrator should take corrective actions to eliminate the possibility of an overload.

If an overload occurs, haload will exit with the special exit code 99.

haload may be invoked either automatically by Solstice HA or by the system administrator.

3.4 Monitoring Metadevice Actions

Metadevices can be monitored using the `metastat` command or the DiskSuite Tool (`metatool(1M)`). Complete details about the two commands can be found in the *Solstice DiskSuite 4.0 Administration Guide* and *Solstice DiskSuite Tool 4.0 User's Guide*.

By default, `metastat` prints information to the screen about all metadevices and hot spare pools that are in the local diskset on the local host. If you want to view diskset status, you must run the command on the server that owns the diskset. An example of the `metastat` command follows:

```
# metastat -s logicalhost1
logicalhost1/d0: Trans
    State: Okay
    Size: 14182560 blocks
    Master Device: logicalhost1/d125
    Logging Device: logicalhost1/d122

logicalhost1/d125: Mirror
    Submirror 0: logicalhost1/d127
      State: Okay
    Submirror 1: logicalhost1/d126
      State: Okay
    Pass: 1
    Read option: roundrobin (default)
    Write option: parallel (default)
    Size: 14182560 blocks

logicalhost1/d127: Submirror of logicalhost1/d125
    State: Okay
    Hot spare pool: logicalhost1/hsp000
    Size: 14182560 blocks
    Stripe 0:
       Device             Start Block  Dbase State       Hot Spare
        c1t0d0s0                    0      No     Okay
    Stripe 1:
       Device             Start Block  Dbase State       Hot Spare
        c1t1d0s0                    0      No     Okay
    Stripe 2:
       Device             Start Block  Dbase State       Hot Spare
        c1t1d1s0                    0      No     Okay
 ...
```

Individual metadevice status can be viewed by specifying the name of the metadevice on the `metastat` command line. For instance:

```
# metastat -s logicalhost1 d0
```

DiskSuite Tool displays status of metadevices and hot spares several ways. The problem list window of the DiskSuite Tool contains a scrolling list of current metadevice problems (but not a history of problems). The list is updated each time DiskSuite Tool learns of a change in status. Each item on the list is given a time stamp.

3.5 Monitoring Metadevice State Database Replicas

Use the `metadb` command to monitor the status of the metadevice state database replicas that reside on both local disks and in disksets.

To display the status of replicas that reside on local disks, execute `metadb` on the server where the disks are connected.

Complete details about the `metadb` command are in the *Solstice DiskSuite 4.0 Administration Guide.*

You can also use the metatool utility to check the status of metadevice state database replicas. Refer to Chapter 10 of the *Solstice DiskSuite Tool 4.0 User's Guide* for details.

To display the status of replicas that reside on disks in a diskset, execute the command shown below. The `-i` option prints the information message at the bottom of the output. The *setname* used as an argument to `metadb` is the name of the logical host.

```
# metadb -i -s setname
      flags           first blk      block count
    a m    luo      16             1034           /dev/dsk/c1t0d0s7
    a     luo      1050           1034           /dev/dsk/c1t0d0s7
    a     luo      16             1034           /dev/dsk/c1t1d0s7
    a     luo      1050           1034           /dev/dsk/c1t1d0s7
    a     luo      16             1034           /dev/dsk/c1t2d0s7
    a     luo      1050           1034           /dev/dsk/c1t2d0s7
    a     luo      16             1034           /dev/dsk/c1t3d0s7
    a     luo      1050           1034           /dev/dsk/c1t3d0s7
 o - replica active prior to last mddb configuration change
 u - replica is up to date
 l - locator for this replica was read successfully
 c - replica’s location was in /etc/opt/SUNWmd/mddb.cf
 p - replica’s location was patched in kernel
 m - replica is master, this is replica selected as input
 W - replica has device write errors
 a - replica is active, commits are occurring to this replica
 M - replica had problem with master blocks
 D - replica had problem with data blocks
 F - replica had format problems
 S - replica is too small to hold current data base
 R - replica had device read errors
#
```

3.6 Checking Message Files

The Solstice HA software writes messages to the /var/adm/messages files in addition to reporting these to the console. The following is an example of the messages reported when a disk error occurs.

```
...
Jun 1 16:15:26 host1 unix: WARNING: /io-
unit@f,e1200000/sbi@0.0/SUNW,pln@a0000000,741022/ssd@3,4(ssd49):
Jun 1 16:15:26 host1 unix: Error for command ‘write(I))’ Err
Jun 1 16:15:27 host1 unix: or Level: Fatal
Jun 1 16:15:27 host1 unix: Requested Block 144004, Error Block: 715559
Jun 1 16:15:27 host1 unix: Sense Key: Media Error
Jun 1 16:15:27 host1 unix: Vendor ‘CONNER’:
Jun 1 16:15:27 host1 unix: ASC=0x10(ID CRC or ECC error),ASCQ=0x0,FRU=0x15
...
```

Note – Because Solaris and Solstice HA error messages are written to the /var/adm/messages file, the /var directory may become full. Refer to “Maintenance of the /var File System” on page 9-10 for the procedure to correct this problem.

3.7 Using SunNet Manager to Monitor Solstice HA Servers

You can use SunNet Manager™ and its agents to monitor Solstice HA configurations. SunNet Manager enables you to set up procedures to get information such as:

- Ownership change
- Status of private links
- Host and network performance

This information can be presented in two ways:

- Graphically, using SunNet Manager
- Through custom scripts
- Event monitors that watch SunNet Manager data for significant changes

Note – Some of the SunNet Manager agents may have an adverse affect on the Solstice HA services.

The SunNet Manager agents that have been identified useful and safe to use in a Solstice HA configuration include `ping`, `hostif`, `hostmem`, `hostperf`, and `traffic`.

Refer to the SunNet Manager documentation set for instructions on setting up the agents.

3.7.1 SunNet Manager Requirements

The following requirements apply to the use of SunNet Manager in Solstice HA configurations:

- SunNet Manager should be installed on the workstation that you will be using to monitor the HA Servers.
- The SunNet Manager libraries and agents should be installed on the local disks of the Solstice HA servers so the activities can be monitored.

Note – You can choose to have a SunNet Manager Console window that is closed to an icon open automatically when an event is received. You specify this in the Console's Properties window, available by clicking SELECT in the Props button in the Console's control area.

You can receive notification by the blinking and coloring effect of the glyph. You can also be notified by either `mail(1)` or by sending the output to your customized script.

Hardware Replacement and Repair 4

This chapter provides the necessary software instructions to use when replacing or repairing hardware components such as disks and cables.

Use the *SPARCcluster High Availability Server Service Manual*, *SPARCstorage Array Model 100 Service Manual*, and the *Solstice DiskSuite 4.0 Administration Guide* with the information in this chapter.

Use the following table to locate specific information in this chapter.

4.1 Recovering From a Power Loss

Maintenance of SPARCcluster 1000 or SPARCcluster 2000 configurations includes handling such failures as power loss.

4.1.1 SPARCcluster 1000 Configuration Power Loss

In SPARCcluster 1000 configurations there are two types of power loss scenarios that can occur. They are:

- The SPARCcluster 1000 configuration has a single power cord and a failure takes down both Solstice HA servers.
- The two SPARCserver 1000s and the three SPARCstorage Arrays have separate power cords.

4.1.1.1 Single Power Cord

When power is lost to a SPARCcluster 1000 configuration in a single 56-inch data center expansion cabinet, the entire configuration will go down if power is supplied by a single cord. The two SPARCserver 1000 servers in the cabinet will reboot when power is restored.

You should immediately run `hastat` and `metastat` to look for any error conditions that may have happened due to the power outage.

When the reboot happens, one of the servers is likely to boot faster than the other and take ownership of both disksets if you have a symmetric configuration. You must run `haswitch(1M)` to reset the default diskset ownership.

The terminal concentrator will boot slower than the SPARCserver 1000s, which means you may miss any messages that appear when the server boots. The terminal concentrator can be run from a separate outlet to ensure it is available as early in the power process as possible.

If the server’s power is not cabled correctly, one of the servers may reboot before the SPARCstorage Arrays and some disks may be invisible to Solaris 2.4. In this event, the server may reboot again and see the SPARCstorage Array when it comes up.

You may need to use the instructions provided in Section 4.3.1, “Recovering From Power Loss,” on page 4-20 for returning the multi-host disks to service.

Note – If any SPARCstorage Array is not ready at Solaris boot time, the associated disks will not be accessible. If this occurs, one or both servers must be rebooted.

4.1.1.2 Separate Power Cords

If separate power cords are used on the two servers and the three SPARCstorage Arrays and you lost power to only one of the servers, the other server will detect the failure and initiate a takeover.

When power is restored to the server that failed, it will boot, wait for the membership state to become stable, and rejoin the configuration. Both disksets will be owned by the server that did not fail. Perform a manual switchover to restore the default diskset ownership.

If you lose power to one of the SPARCstorage Arrays, Solstice DiskSuite will detect errors on the affected disks and place the slices in error state. The SPARCstorage Array drivers will attempt to retry connections for up to one minute before reporting an error. Solstice DiskSuite mirroring will mask this failure from the Solstice HA fault monitoring. No switchover or takeover will occur.

When power is returned to the SPARCstorage Array, you must perform the procedure documented in Section 4.3.1, “Recovering From Power Loss.”

4.1.2 SPARCcluster 2000 Configuration Power Loss

In SPARCcluster 2000 configurations there are several types of power loss scenarios that can occur. These include:

- The power to both SPARCcenter 2000s fails, taking down the entire configuration.
- The power to one SPARCcenter 2000 fails, taking down the server and one SPARCstorage Array.
- The power to one SPARCcenter 2000 fails, taking down the server, two SPARCstorage Arrays, and the terminal concentrator.

4.1.2.1 Total Configuration Failure

If power to both servers in a SPARCcluster 2000 configuration fails, one of the servers may reboot faster than the other. If you have a symmetric configuration, the first server to reboot will take ownership of both disksets. In this event, you must return one of the disksets to the default master by using `haswitch`.

You should immediately run hastat and metastat to look for any error conditions that may have happened due to the power outage.

4.1.2.2 *Failure of a Server and One SPARCstorage Array*

If power is lost to one of the SPARCcenter 2000s and the SPARCstorage Array that is installed in the same cabinet, the other server will immediately initiate a takeover.

When the power is restored, the server will reboot, rejoin the configuration and begin monitoring activity. You must manually run haswitch to give ownership of the diskset back to the server that had lost power.

After the diskset ownership has been returned to the default master, any multi-host disks (submirrors, hot spares, and metadevice state database replicas) that reported errors must be returned to service. Use the instructions provided in Section 4.3.1, “Recovering From Power Loss,” on page 4-20 for returning the multi-host disks to service.

4.1.2.3 *Failure of a Server, Two SPARCstorage Arrays, and the Terminal Concentrators*

If power is lost to one of the SPARCcenter 2000s and the two SPARCstorage Arrays that are installed in the same cabinet, either a Solstice DiskSuite panic will occur because there is a minority of metadevice state database replicas or the Solstice HA software will cause a panic.

When any I/O is done to the disks in either of the two SPARCstorage Arrays, the problem will be noticed by Solstice DiskSuite. Briefly, DiskSuite will retry the I/O, then it will initiate a replica minority panic when it attempts to record the error status of the affected submirrors and discovers it has only a minority of replicas accessible.

Possibly, the HA-NFS fault probing may observe the problem as slow response before Solstice DiskSuite actually receives the disk I/O error. In this case a takeover may be initiated and a panic will occur during diskset takeover when a minority of the replicas are accessible.

The console message may not be visible if the terminal concentrator is also down.

When power is restored, the SPARCcenter 2000 may reboot before the terminal concentrator. Thus, any errors reported when the SPARCcenter 2000 is rebooting must be viewed using `dmesg(1M)` or by looking in `/var/adm/messages`. Depending on the specifics of your configuration, manual intervention may be required to return the Solstice HA configuration to service.

4.2 Replacement of Failed Disks

As part of standard Solstice HA administration, you should monitor the status of the configuration. See the instructions in Chapter 3, “Monitoring the Solstice HA Servers,” for instructions on the monitoring methods.

During the monitoring process you may discover problems with local and multi-host disks. The following subsections provide instructions for correcting these problems.

4.2.1 Overview of Multi-host Disk Replacement

The procedures in the following subsection describe a method for replacing a multi-host disk without interrupting Solstice HA services (online replacement). Consult the *Solstice DiskSuite 4.0 Administration Guide* for offline replacement procedures.

If a disk in a SPARCstorage Array must be replaced, you must first stop all I/O to the SPARCstorage Array tray containing the disk to be replaced. This is required so the tray can be spun down in preparation for drive replacement.

Caution – Before deleting the replicas and hot spares, you must make a record of the location (slice), number of replicas, and the hot spare information (names of the devices and all containing hot spare pools) so the actions can be reversed following the disk replacement.

You must delete any metadevices state database replicas from the affected tray to prevent replica IO operations during the replacement procedure. You must also offline or detach submirrors on the affected tray to stop their I/O activity. Finally, available hot spare devices must be deleted from hot spare pools to prevent them from begin brought into service during the disk replacement procedure.

When the I/O operations have been stopped the drives on the SPARCstorage Array tray can be spun down and the tray removed. The disk can be replaced and the tray returned to the array.

Before using the replacement disk drive it must be partitioned to match the partitioning of the replaced disk. This can be done using `format(1m)` or `fmthard(1m)`.

Metadevices state database replicas are added back to the tray in the same locations and with the same counts using the `metadb(1M)` command. Offline mirrors are brought back online and brought up to date resyncing only those dirty regions of the submirrors (optimized resync) using `metaonline(1M)`. Detached mirrors are attached and brought up to date with a submirror resync using `metattach(1M)` (this is expensive, but depending on specifics of submirror configuration is the only safe method). Finally the deleted hot spare devices are returned to their original hot spare pools using `metahs(1M)`.

▼ How to Replace a Failed Multi-host Disk

Replacement of a failed multi-host disk is a complex procedure. You should read the Section 4.2.1, “Overview of Multi-host Disk Replacement,” before you begin.

Note – This procedure can be used if a submirror component is in maintenance state, hot spare replaced, or is generating intermittent errors.

When `metastat(1M)` reports that a device is in maintenance state or some of the components have been replaced by hot spares, you must locate and replace the device. An example `metastat` output that shows device `c3t3d4s0` is in maintenance state follows:

```
host1# metastat -s logicalhost1
...
 d50:Submirror of logicalhost1/d40
      State: Needs Maintenance
      Stripe 0:
          Device       Start Block      Dbase      State          Hot Spare
          c3t3d4s0     0                No         Okay           c3t5d4s0
...
```

To locate and replace the disk, perform the following steps:

1. **Identify the disk to be replaced by examining `/var/adm/messages` and `metastat` output.**

```
host1# tail -f /var/adm/messages
...
Jun 1 16:15:26 host1 unix: WARNING: /io-
unit@f,e1200000/sbi@0.0/SUNW,pln@a0000000,741022/ssd@3,4(ssd49):
Jun 1 16:15:26 host1 unix: Error for command 'write(I))' Err
Jun 1 16:15:27 host1 unix: or Level: Fatal
Jun 1 16:15:27 host1 unix: Requested Block 144004, Error Block: 715559
Jun 1 16:15:27 host1 unix: Sense Key: Media Error
Jun 1 16:15:27 host1 unix: Vendor 'CONNER':
Jun 1 16:15:27 host1 unix: ASC=0x10(ID CRC or ECC error),ASCQ=0x0,FRU=0x15
...
```

   Based on the above information and `metastat` output, it is determined that drive `c3t3d4` must be replaced.

2. **Locate the diskset that contains the affected drive.**
   Locate drive `c3t3d4` by entering the following commands. Note that no output was displayed when the command was run with `logicalhost2`, but `logicalhost1` reported that the name was present. In the reported output, the `yes` field indicates that the disk contains a metadevice state database replica.

```
host1# metaset -s logicalhost2 | grep c3t3d4
host1# metaset -s logicalhost1 | grep c3t3d4
c3t3d4 yes
```

3. **Switch ownership of both logical hosts to one Solstice HA server using a command similar to the following:**

```
host1# haswitch host1 logicalhost1 logicalhost2
```

   The SPARCstorage Array tray that contains `c3t3d4` (the disk with the problem in this example) may also contain disks from both disksets. If this is the case, you must switch ownership of both disksets to the server where the `ssaadm(1M)` command will be used to spin down the disks.

4. **Determine the location of the problem disk.**
   To find the SPARCstorage Array tray where the problem disk resides, run the `ssaadm` command.

```
host1# ssaadm display c3
         SPARCstorage Array Configuration
Controller path: /devices/io-
unit@f,e1200000/sbi@0.0/SUNW,soc@0,0/SUNW,pln@a0000000,741022:ctlr
         DEVICE STATUS
         TRAY1          TRAY2          TRAY3
Slot
1        Drive:0,0      Drive:2,0      Drive:4,0
2        Drive:0,1      Drive:2,1      Drive:4,1
3        Drive:0,2      Drive:2,2      Drive:4,2
4        Drive:0,3      Drive:2,3      Drive:4,3
5        Drive:0,4      Drive:2,4      Drive:4,4
6        Drive:1,0      Drive:3,0      Drive:5,0
7        Drive:1,1      Drive:3,1      Drive:5,1
8        Drive:1,2      Drive:3,2      Drive:5,2
9        Drive:1,3      Drive:3,3      Drive:5,3
10       Drive:1,4      Drive:3,4      Drive:5,4

         CONTROLLER STATUS
Vendor:    SUNW
Product ID:  SSA100
Product Rev: 1.0
Firmware Rev: 2.3
Serial Num: 000000741022
Accumulate performance Statistics: Enabled
```

The `ssaadm` output for controller (`c3`) shows that `Drive 3,4` (`c3t3d4`) is the closest to you when you pull out the middle tray.

5. **Delete all hot spares that are have `Available` status and are in the same tray as the problem disk.**
   This includes all hot spares, regardless of their logical host assignment. In the following example, `metahs` reports the hot spares on `logicalhost1`, but none are present on `logicalhost2`.

   You should record all the information about the hot spares so they can be added back to the hot spare pools following the replacement procedure.

```
host1# metahs -s logicalhost1 -i
logicalhost1:hsp000 2 hot spares
        c1t4d0s0                Available       2026080 blocks
        c3t2d5s0                Available       2026080 blocks
host1# metahs -s logicalhost1 -d hsp000 c3t2d5s0
host1# metahs -s logicalhost2 -i
host1#
```

6. **Delete any metadevice state database replicas that are on disks in the tray that must be pulled. You must keep track of this information because you must replace these replicas in Step 18.**
   There may be multiple replicas on the same disk. Make sure you record the number of replicas deleted from each slice.

```
host1# metadb -s logicalhost1
 This command reports the replicas in diskset logicalhost1
host1# metadb -s logicalhost2
 This command reports the replicas in diskset logicalhost2
host1# metadb -s logicalhost1 -d replicas_in_tray
host1# metadb -s logicalhost2 -d replicas_in_tray
```

7. **Locate the submirrors that are using components that reside in tray 2.**
   One method to use would be to use the `metastat` command to create temporary files that contain the names of all metadevices. For instance:

```
host1# metastat -s logicalhost1 > /tmp/logicalhost1.stat
host1# metastat -s logicalhost2 > /tmp/logicalhost2.stat
```

Search the temporary files for the c3t3d*n* and c3t2d*n* components. If you used the hasetup(1M) defaults (two non-reserved user slices per disk), there will be a maximum of 20 components (10 disks * 2 slices).

The information in the temporary files will look like:

```
...
logicalhost1/d35: Submirror of logicalhost1/d15
   State: Okay
   Hot Spare pool: logicalhost1/hsp100
   Size: 2026080 blocks
   Stripe 0:
      Device      Start Block     Dbase     State      Hot Spare
      c3t3d3s0    0               No        Okay
logicalhost1/d54: Submirror of logicalhost1/d24
   State: Okay
   Hot Spare pool: logicalhost1/hsp106
   Size: 21168 blocks
   Stripe 0:
      Device      Start Block     Dbase     State      Hot Spare
      c3t3d3s6    0               No        Okay
...
```

8. **Detach all submirrors with components on the disk that is being replaced.**
   If you are detaching a submirror that has an errored component you must force the detach using the metadetach -f option.

```
host1# metadetach -s logicalhost1 d40 d50
```

9. **Take all other submirrors that have components in tray 2 offline.**
   Using the output from the temporary files in Step 7, run the metaoffline command on all submirrors in tray 2.

```
host1# metaoffline -s logicalhost1 d15 d35
host1# metaoffline -s logicalhost1 d24 d54
...
```

Run metaoffline as many times as necessary (maybe up to 20 times) to take all the submirrors offline. This forces Solstice DiskSuite to stop using the submirror components in tray 2 so that the spin down command can be issued.

10. **Spin down all disks in tray 2 of the SPARCstorage Array.**

```
host1# ssaadm stop -t 2 c3
```

Caution – The SPARCstorage Array tray should not be removed as long as the LED on the tray is illuminated. Also, you should not run any Solstice DiskSuite command while the tray is spun down as these may have the side effect of spinning up some or all of the drives in the tray.

11. **Pull tray 2 and replace the bad disk.**
    Instructions for the hardware procedure are found in the *SPARCstorage Array Model 100 Series Service Manual* (part number 801-2206) and the *SPARCcluster High Availability Server Service Manual.*

12. **Make sure all disks in tray 2 of the SPARCstorage Array spin up.**
    The disks in the SPARCstorage Array tray should automatically spin up following the hardware replacement procedure. If the tray fails to spin up automatically within two minutes, force the action by using the following command.

```
host1# ssaadm start -t 2 c3
```

13. **Use format(1M) or fmthard(1M) to repartition the new disk. Make sure you partition the new disk exactly as the disk that was replaced.**
    Saving the disk format information was recommended in Chapter 2, "Preparing for Administration."

14. **Bring all submirrors that were taken offline in Step 9 back online.**

```
host1# metaonline -s logicalhost1 d15 d35
host1# metaonline -s logicalhost1 d24 d54
...
```

Running `metastat` at this time would show that all the metadevices with components that reside in the second tray need maintenance.

When the submirrors are brought back online, Solstice DiskSuite will automatically perform resyncs on all the submirrors, bringing all the data back up to date.

Run `metaonline` as many times as necessary (maybe up to 20 times) to bring all the submirrors online.

15. **Attach submirrors that were detached in Step 8.**

```
host1# metattach -s logicalhost1 d40 d50
```

16. **Replace any hot spares in use in the submirrors attached in Step 15.**
If a submirror had a hot spare replacement in use before you detatched the submirror, this hot spare replacement will be in affect after the submirror is reattached. This step returns the hot spare to `Available` status.

```
host1# metareplace -s logicalhost1 -e d40 c3t3d4s0
```

17. **Add all hot spares that were deleted in Step 5.**

```
host1# metahs -s logicalhost1 -a hsp000 c3t2d5s0
```

18. **Add all metadevice state database replicas that were deleted from disks on tray 2.**
Use the information saved from Step 6 to replace the metadevice state database replicas.

```
host1# metadb -s logicalhost1 -a deleted_replicas
```

19. Switch each logical host back to its default master.

```
host1# haswitch host2 logicalhost2
```

4.2.2 Overview of Local Disk Replacement

In both the SPARCcluster 2000 and SPARCcluster 1000 configurations there are at least two local disks. One of the local disks is the boot disk which contains the Solaris operating environment. The other local disk contains your other local data.

Replacement of the boot disk is a difficult procedure and is detailed in the "How to Replace a Failed Local Boot Disk" on page 4-13. The procedure in that section covers the replacement of the failed local disk that contains the Solaris operating environment. In this procedure, the local disk on `host1` has failed.

Note – Depending on the severity of your boot disk failure, you may not be able to perform all the steps. If a server is already down, you may omit steps 1, 2, and 3.

The procedure for replacing a local disk that is not the boot disk is covered in "How to Replace a Failed Local Non-Boot Disk" on page 4-17.

▼ How to Replace a Failed Local Boot Disk

Note – This procedure expects that you installed and configured the two Solstice HA servers identically. This allows you to copy the various configuration files from the sibling rather than using backup tapes.

1. **Switch ownership of both logical hosts to one Solstice HA server using a command similar to the one shown.**
   If a takeover has been initiated by the sibling, it will not be necessary to run this command.

```
host2# haswitch host2 logicalhost1 logicalhost2
```

2. **Shut down the Solstice HA services on the host with the failed local disk.**

```
host1# /etc/init.d/SUNWhadf stop
```

3. **Halt the server that has the failed local disk.**

```
host1# halt
```

4. **Delete the server from the disksets on the sibling server.**
   Each of these commands may take several minutes to execute.

```
host2# metaset -s logicalhost1 -f -d -h host1
host2# metaset -s logicalhost2 -f -d -h host1
```

5. **Perform the disk replacement using the procedure in the** *SPARCcluster High Availability Server Service Manual***.**

6. **Install the Solaris operating environment using the instructions in the** *SPARCcluster High Availability Software Planning and Installation Guide***.**
   Make sure you install the same software clusters (packages) that are installed on the sibling server.

Note – Select the Do Not Reboot option during installation. This will allow you to restore some files to the new root slice in `/a` before doing the reconfiguration reboot.

During installation, numerous Disk Reserved messages will be displayed because the sibling host owns all the SPARCstorage Array disks. These messages can safely be ignored.

7. **Edit the** `/a/etc/nsswitch.conf` **file, changing the host line to specify** `files` **first.**
   Later you can copy the `/etc/nsswitch.conf` file from the sibling, but the host line must specify files first for the procedure you are performing here to work.

8. **Restore a copy of the** /etc/path_to_inst **file and place it in** /a/etc/path_to_inst**.**
   Because the two servers were installed identically, the file can be copied from the sibling.

9. **Restore a copy of the** /etc/system **file and place it in** /a/etc/system**.**
   This file contains modification for Solstice HA operation.

10. **(Optional) Edit the** /a/etc/default/login **file to allow root logins on terminals other than the console.**

11. **Add all host names, all private host names, and all logical host names to the** /a/etc/hosts **file.**
    Refer to Chapter 7 of the *SPARCcluster High Availability Software Planning and Installation Guide* for instructions.

12. **Configure the private networks. Run the** ifconfig **command as shown below.**

```
host1# ifconfig be0 plumb
host1# ifconfig be0 host1-priv1 netmask + broadcast + -trailers up
```

Refer to Chapter 7 of the *SPARCcluster High Availability Software Planning and Installation Guide* for additional information.

13. **Copy the configuration files from the sibling.**
    Use the following commands to copy the network configuration files from the sibling.

```
host1# rcp -p host2-priv1:/etc/nsswitch.conf /a/etc
host1# rcp -p host2-priv1:/etc/syslog.conf /a/etc
host1# rcp -p host2-priv1:/etc/netmasks /a/etc
host1# rcp -p host2-priv1:/kernel/drv/md.conf /a/kernel/drv
host1# rcp -p host2-priv1:/.rhosts /a/.rhosts
host1# rcp -p host2-priv1:/.profile /a (optional command)
```

14. **(Optional) Copy the appropriate entries from the old** vfstab **file from tape and make mount points for formerly mounted file systems.**

15. **(Optional) Restore the** crontab **file from backup tape.**

16. **(Optional) Enable core dumps in** /a/etc/init.d/sysetup**.**

17. **(Optional) Restore** `/a/etc/resolv.conf` **(if you are using DNS).**

18. **Reboot the server.**

```
host1# reboot
```

19. **Install the Solstice DiskSuite and Solstice HA packages and recommended patches.**

20. **Restore the** `/etc/hostname.*` **files for both private and secondary public networks.**
    If these files do not exist, you must re-create them by using the instructions in Chapter 7 of the *SPARCcluster High Availability Software Planning and Installation Guide*.

21. **Execute a reconfiguration reboot.**
    The reconfiguration reboot builds the appropriate device special inodes for Solstice DiskSuite. To execute a reconfiguration reboot, enter the following:

```
host1# reboot -r
```

22. **Add the three replicas back on slice 4 of the new boot disk.**
    In this example `/dev/dsk/c0t0d0s4` is used.

```
host1# metadb -afc 3 /dev/dsk/c0t0d0s4
```

23. **Add the server to the disksets.**
    Note that the following command are executed on `host2`.

```
host2# metaset -s logicalhost1 -a -h host1
host2# metaset -s logicalhost2 -a -h host1
```

24. **Copy the Solstice HA configuration files from the sibling.**
    Enter the following commands from the sibling host (`host2`) to copy the appropriate files.

```
host2# cd /etc/opt/SUNWhadf/hadf
host2# rcp -p cmm_confcdb hadfconfig host1-priv1:/etc/opt/SUNWhadf/hadf
host2# rcp -p hafmconfig vfstab.* host1-priv1:/etc/opt/SUNWhadf/hadf
host2# cd /etc/opt/SUNWhadf/nfs
host2# rcp -p dfstab.* host1-priv1:/etc/opt/SUNWhadf/nfs
```

25. **Run the `hacheck(1M)` command.**
    Instructions for using this command can be found in Chapter 11 of the *SPARCcluster High Availability Software Planning and Installation Guide.*

26. **Create a hard link from `/etc/rc3.d/S20SUNWhadf` to `/etc/init.d/SUNWhadf`.**
    This link automatically starts Solstice HA when the server is brought up in multi-user mode. It will not automatically start Solstice HA when the server is brought up in single user mode.

```
host1# ln /etc/init.d/SUNWhadf /etc/rc3.d/S20SUNWhadf
```

27. **Start the Solstice HA services.**

```
host1# /etc/init.d/SUNWhadf start
```

28. **Switch the logical host back to its default master.**

```
host1# haswitch host1 logicalhost1
```

▼ How to Replace a Failed Local Non-Boot Disk

This procedure covers the replacement of the failed local disk that does not contain the Solaris operating environment. In this example, `host2` has the disk that failed.

1. **Switch ownership of both logical hosts to the server that is not experiencing problems. Use a command similar to the following:**

```
host1# haswitch host1 logicalhost1 logicalhost2
```

2. **Shut down the Solstice HA services on the server that is having problems.**

```
host2# /etc/init.d/SUNWhadf stop
```

3. **Locate any local metadevice state database replicas that may have been placed on the problem disk. Use the `metadb` command to find the replicas.**
   Errors may be reported for the replicas located on the failed disk. In this example, `c0t1d0` is the problem device.

```
host2# metadb
   flags         first blk        block count
  a m     u        16             1034           /dev/dsk/c0t0d0s4
  a       u        1050           1034           /dev/dsk/c0t0d0s4
  a       u        2084           1034           /dev/dsk/c0t0d0s4
  W   pc luo       16             1034           /dev/dsk/c0t1d0s4
  W   pc luo       1050           1034           /dev/dsk/c0t1d0s4
  W   pc luo       2084           1034           /dev/dsk/c0t1d0s4
host2#
```

The output shown above shows there are three metadevice state databases on slice 4 of each of the local disks, `c0t0d0s4` and `c0t1d0s4`. The `W` in the flags field of the `c0t1d0s4` slice indicates the device has write errors.

4. **Make a record of the slice name where the replicas reside and the number of replicas, then delete the metadevice state databases.**
   The number of replicas is obtained by counting the number of appearances of a slice in `metadb` output in Step 3. In this example, we are deleting the three replicas that exist on `c0t1d0s4`.

```
host2# metadb -d /dev/dsk/c0t1d0s4
```

5. **Shut down Solaris and turn off the server.**

```
host2# halt
```

6. **Perform the disk replacement using the procedure in the** *SPARCcluster High Availability Server Service Manual***.**

7. **Turn the server on and reboot it in single user mode.**

```
ok boot -s
```

8. **Repartition the new disk with the same slice information as the failed disk.**

9. **Run** `newfs(1M)` **on the new slices to create file systems.**

```
host2# newfs raw_device
```

10. **Mount the appropriate file systems.**

11. **Restore data from backup tapes.**

12. **If you deleted replicas in Step 4, add the same number back to the appropriate slice.**
    In this example, `/dev/dsk/c0t1d0s4` is used.

```
host2# metadb -ac 3 /dev/dsk/c0t1d0s4
```

13. **Reboot the server.**

```
host2# reboot
```

14. **When the host has rejoined the Solstice HA configuration (this usually takes about one minute), switch the logical host back to its default master.**

```
host2# haswitch host2 logicalhost2
```

4.3 SPARCstorage Array Maintenance

Maintenance of the SPARCstorage Arrays in a SPARCcluster 1000 or SPARCcluster 2000 configuration involves the following:

- Recovering from power loss
- Repairing a lost connection
- Replacing a failed SPARCstorage Array (changing the World Wide Name)
- Removing a SPARCstorage Array tray
- Replacing a SPARCstorage Array tray
- Replacing failed SPARCstorage Array components (disk, battery, backplane, controller, optical module, fan tray, or fibre channel cable)

4.3.1 Recovering From Power Loss

When power is lost to one SPARCstorage Array, I/O operations to the submirrors, hot spares, and metadevice state database replicas will generate Solstice DiskSuite errors. The errors are reported at the slice level rather than the drive level. Errors are not reported until I/O operations are made to the disk. Hot spare activity may be initiated if affected devices have assigned hot spares.

You must monitor the configuration for these events using `hastat(1M)` and `metastat(1M)` as explained in Chapter 3, “Monitoring the Solstice HA Servers.”

When power is restored, you will use the `metastat` command to identify the errored devices. Errored devices are returned to service using the command:

```
# metareplace -s logicalhost -e metamirror component
```

The `-e` option transitions the state of component to the available state and resyncs the failed component.

Note – Components that have been replaced by a hot spare should be the last devices replaced using the `metareplace` command. If the hot spare is replaced first, it could replace another errored submirror as soon as it becomes available.

A resync can be performed on only one component of a submirror (metadevice) at a time. If all components of a submirror were affected by the power outage, each component must be replaced separately. It takes approximately 10 minutes for a resync to be performed on a 1.05-Gbyte disk.

If both disksets in a symmetric configuration were affected by the power outage, a resync can be run on the affected submirrors concurrently by logging into each host separately and running `metareplace`.

Depending on the number of submirrors and the number of components in these submirrors, the resync actions can require a considerable amount of time. A single submirror that is made up of 30 1.05-Gbyte drives might take about five hours to complete. A more realistic configuration made up of five component submirrors might take only 50 minutes to complete.

After the loss of power, all metadevice state database replicas on the affected SPARCstorage Array chassis will enter an errored state. While these will be reclaimed at the next takeover (`haswitch` or `reboot(1M)`) you may want to manually return them to service by first deleting and then adding them back as metadevices. Because metadevice state database replica recovery is not automatic, it is safest to manually perform the recovery immediately after the SPARCstorage Array returns to service. Otherwise, a new failure may cause a majority of replicas to be out of service and cause a kernel panic. This is the expected behavior of Solstice DiskSuite when too few replicas are available.

Note – Make sure you add back the same number of replicas that were deleted on each slice. Multiple replicas can be deleted with a single `metadb` command. It may require multiple invocations of `metadb -a` to add back the replicas deleted by a single `metadb -d`. This is because if you need multiple copies of replicas on one slice these must be added in one invocation of `metadb` using the `-c` flag.

4.3.2 *Repairing a Lost Connection*

When a connection from a SPARCstorage Array to one of the hosts fails, the failure is probably due to either a fiber optic cable or a SBus FC/S or FC/OM cards.

In either event, the host on which the failure occurred will begin generating errors when the failure is discovered. This takes about one minute. Later accesses to the SPARCstorage Array will generate additional errors. The host will exhibit the same behavior as though power had been lost to the SPARCstorage Array.

In symmetric configurations, I/O operations from the other host to the SPARCstorage Array are unaffected by this type of failure.

To diagnosis the failure, inspect the SPARCstorage Array's display. The display will show whether the A or B connection has been lost.

To replace the cable, use the following procedure. In this example, the connection to `host2` from one SPARCstorage Array must be replaced.

▼ How to Repair a Lost Connection

1. **Replace the failed cable.**
   Refer to the *SPARCstorage Array Model 100 Series Service Manual* for detailed instructions.

2. **Recover from Solstice DiskSuite errors as described in Section 4.3.1, "Recovering From Power Loss."**
   Solstice DiskSuite cannot detect whether the loss of power is due to a failed SPARCstorage Array or a power loss.

4.3.3 *Changing a SPARCstorage Array World Wide Name*

Some SPARCstorage Array failures may make it necessary to replace the entire chassis.

These failures can be caused by a faulty controller or other reasons. To guard against this type of failure, all metadevices have been set up with only one submirror of a mirror on a SPARCstorage Array chassis. Thus, loss of data will not occur with this type of failure.

The SPARCstorage Array controller has a unique identifier known as the World Wide Name (WWN). The WWN is like the host ID stored in the host IDPROM of a desktop SPARCstation. The last four digits of the SPARCstorage Array WWN are displayed on the LCD panel of the chassis. The WWN is part of the `/devices` path associated with the SPARCstorage Array and its component drives.

When you replace the SPARCstorage Array chassis in a Solstice HA configuration, you can change the WWN of the replacement chassis to be that of the chassis you are replacing. This may be easier than reconfiguring Solstice DiskSuite.

If the SPARCstorage Array controller or the entire chassis must be replaced, the Solstice HA servers will discover the new WWN when they are rebooted. This confuses the identity of disks within a diskset. To avoid this potential confusion, the WWN of the new controller can be changed to the WWN of the old controller. (This is similar to swapping the IDPROM when replacing a System Board in a desktop SPARCstation.)

▼ How to Change a SPARCstorage Array World Wide Name

1. **Determine the symbolic link value of the SPARCstorage Array.**
   Assuming that controller `c1` failed, enter the following `ls(1)` command. The command will report the symbolic link value of the SPARCstorage Array.

```
# ls -l /dev/rdsk/c1t0d0s0
lrwxrwxrwx   1 root     root         92 Jun 25 12:11 /dev/rdsk/c1t0d0s0 -> ../../devices/io-
unit@f,e0200000/sbi@0,0/SUNW,soc@3,0/SUNW,pln@a0000000,7412bf/ssd@0,0:a,raw
```

   Another way to discover the WWN is with the `ssaadm` command. When you run `ssaadm(1M)` with the `display` option and specify a controller, all the information about the SPARCstorage Array is displayed. The serial number reported by `ssaadm` is the WWN.

2. **Obtain the WWN from the `pln` path.**
   The WWN is the last 12 hexadecimal digits of the path component containing the characters `pln` from the symbolic link value of the SPARCstorage Array (not including commas).

```
SUNW,pln@a0000000,7412bf
```

3. **Change the WWN.**
   Use the `ssaadm` command to change the WWN. For example, the following command would change the WWN to `0000007411f3`.

```
# ssacli -s -w 0000007411f3 download c1
```

Note – The leading zeros must be entered as part of the WWN to make a total of 12 digits.

4.3.4 Removing a SPARCstorage Array Tray

Before removing a SPARCstorage Array, tray you must halt all I/O and spin down all drives in the tray. The drives automatically spin up if I/O requests are made. Thus, it is necessary to stop all I/O before the drives are spun down.

Stop Solstice DiskSuite I/O activity with the `metaoffline(1M)` command, which takes the submirror offline. (The `metadetach(1M)` command could be used to stop the I/O, but the resync cost is greater.) When the submirrors on a tray are taken offline, the corresponding mirrors will only provide one-way mirroring (that is, there will be no data redundancy). When the mirror is brought back online, an automatic resync occurs.

Use the `metastat(1M)` command to identify all submirrors containing slices on the tray to be removed. Also, use the `metadb(1M)` command to identify any replicas on the tray. Any available hot spare devices must also be identified and the associated submirror identified using the `metahs(1M)` command.

With all affected submirrors offline, I/O to the tray will be stopped.

The `ssaadm` command is used to spin down the tray. When the tray lock light is out the tray may be removed and the required task performed.

4.3.5 Replacing a SPARCstorage Array Tray

When you have completed work on a SPARCstorage Array tray, replace the tray in the chassis. The disks will automatically spin up. However if the disks fail to spin up, you can use the `ssaadm` command to manually spin up the entire tray. There is a short delay (several seconds) between starting drives in the SPARCstorage Array.

After the disks have spun up, you must place online all the submirrors that were taken offline. When the `metaonline(1M)` command is run, an optimized resync operation automatically brings the submirrors up to date. The optimized resync copies only the regions of the disk that were modified while the submirror was offline. This is typically a very small fraction of the submirror capacity. You must also replace all metadevice state database replicas (`metadb(1M)`) and add back hot spares (`metahs(1M)`).

Note – If you used `metadetach(1M)` to detach the submirror rather than `metaoffline`, the entire submirror must be resynced. This typically takes about 10 minutes per Gbyte of data.

4.3.6 Replacing SPARCstorage Array Components

The SPARCstorage Array components that can be replaced include the disks, fan tray, battery, tray, power supply, backplane, controller, optical module, and fibre channel cable.

Some of the SPARCstorage Array components can be replaced without powering down the SPARCstorage Array. Other components require the SPARCstorage Array to be powered off. Consult the SPARCstorage Array documentation for details.

To replace SPARCstorage Array components which require power off without interrupting Solstice HA services you perform the steps necessary for tray removal for all three trays in the SPARCstorage Array before turning off the power. This will include taking submirrors offline, deleting hot spare devices from hot spare pools, deleting metadevice state database replicas from drives, and spinning down the three trays.

After these preparations, the SPARCstorage Array can be powered down and the components replaced.

Note – Because the SPARCstorage Array controller contains a unique World Wide Name, which identifies it to Solaris, special procedures apply for SPARCstorage Array controller replacement. Contact your service provider for assistance.

After component replacement and power on follow the tray replacement procedures for all three trays.

4.4 Replacing Network Cables and Interfaces

There are three types of failures that require the replacement of network cables and interfaces.These include:

- Public or client Ethernet cable failure
- Private network cable failure
- Public or private Ethernet interface failure

The following procedures provide instructions for these replacements.

▼ How to Replace a Public or Client Ethernet Cable

1. **Switch ownership of both logical hosts to the Solstice HA server that does not need an Ethernet cable replaced.**
   For instance, if the cable is being replaced on `host1` enter the following:

```
host1# haswitch host2 logicalhost1 logicalhost2
```

2. **Replace the cable using the appropriate hardware instructions in the** *SPARCcluster High Availability Server Service Manual*.

3. **Switch ownership of the logical hosts back to the appropriate default master.**
   For instance:

```
host1# haswitch host1 logicalhost1
```

How to Replace a Private Network Cable

When a private network cable fails, both servers will be aware that a private network connection is not working. The Solstice HA services should not be affected because of the second private network cable.

1. **Unplug the faulty Ethernet cable and replace it with a new one.**
   You can use either Sun Microsystems' replacement parts number 530-2149 or 530-2150. If you are not using standard Sun parts, be sure the replacement Ethernet cable has the pairs crossed. Refer to Appendix B of the *SPARCcluster High Availability Server Service Manual* for cable information.

How to Replace a Public or Private Ethernet Interface

1. **Switch ownership of both logical hosts to the Solstice HA server that does not need a the Ethernet interface replaced.**
   For instance, if the interface is being replaced on `host1` enter the following:

```
host1# haswitch host2 logicalhost1 logicalhost2
```

2. **Shut down the Solstice HA software on `host1`.**

```
host1# /etc/init.d/SUNWhadf stop
```

3. **Halt the server.**

```
host1# halt
```

4. **Power off the server.**

5. **Replace the appropriate public or private Ethernet interface using the instructions in the** ***SPARCcluster High Availability Server Service Manual*****.**

6. **Power on the server.**
   The server will automatically rejoin the Solstice HA configuration.

7. **Switch ownership of the logical hosts back to the appropriate default master.**
   For example:

```
host1# haswitch host1 logicalhost1
```

4.5 System Board Replacement

The Solstice DiskSuite component of Solstice HA is sensitive to the device numbering and can become confused if System Boards are moved around.

When the server is booted initially, the SPARCstorage Array entries in the `/dev` directory are tied to the connection slot.

For example, when the server is booted, System Board 0 and SBus slot 1 will be part of the identify of the SPARCstorage Array. If the board or SBus card is shuffled to a new location Solstice DiskSuite will be confused because Solaris will assign new controller numbers to the SBus controllers when they are in a new location.

Note – The SBus cards can be moved as long as the type of SBus card in a slot remains the same.

Shuffling the fiber cables that lead to the SPARCstorage Arrays can also create problems. The System Boards on each of the Solstice HA servers must be configured identically (that is, the same type of SBus cards in each slot). When SBus cards are switched you must also reconnect the SPARCstorage Arrays back to the same SBus slot they were connected to before the changes.

4.6 Replacing SBus Cards

Replacement of SBus cards in Solstice HA servers can be done by switching over the data services to the server that is functioning and performing the hardware procedure to replace the board. The logical hosts should be switched back to the default masters following the procedure.

▼ How to Replace an SBus Card

1. **Switch ownership of both logical hosts to the Solstice HA server that does not need an SBus card replaced.**
   For instance, if the board is being replaced on `host2` enter the following:

```
host1# haswitch host1 logicalhost1 logicalhost2
```

2. **Stop Solstice HA on the affected server.**
   The the `SUNWhadf stop` command must be run on the host that has the failed SBus card.

```
host2# /etc/init.d/SUNWhadf stop
```

3. **Halt the affected server.**

```
host2# /etc/halt
```

4. **Power off the server.**

5. **Perform the hardware replacement procedure.**
   Refer to the instructions in the appropriate hardware service manual that contains instructions for replacement of the specific SBus card.

6. **Power on the server.**
   The server will automatically rejoin the Solstice HA configuration.

7. **Switch the logical hosts back to the default masters.**

```
host1# haswitch host2 logicalhost2
```

Adding Hardware 5

This chapter tells the software procedure to follow when adding hardware such as disks, SPARCstorage Arrays, and public network connections to Solstice HA configurations.

Use the following table to locate specific information in this chapter.

5.1 Adding a SPARCstorage Array

Additional SPARCstorage Arrays can be added to a Solstice HA configuration at any time.

You must review the metadevice distribution in your Solstice HA configuration before adding a SPARCstorage Array. The discussions in the *SPARCcluster High Availability Software Planning and Installation Guide*, and Chapter 8, “Metadevice and Diskset Administration,” in this manual will help determine the impact of the SPARCstorage Array on the distribution of metadevices.

How to Add a SPARCstorage Array

1. **Shut down one of the Solstice HA servers (`host1`).**
   Use the procedure in Section 9.3, “Shutting Down Solstice HA Servers,” on page 9-5 to shut down the server.

2. **Install the Fibre Channel SBus card (FC/S) in the server.**
   Use the instructions in the *SPARCcluster High Availability Server Service Manual* to install the FC/S card.

Note – Install the FC/S card in the first available empty SBus slot, following all other cards in the server. This will ensure the controller numbering will be preserved if the Solaris operating environment is reinstalled. Refer to Section 2.4, “Instance Numbering,” on page 2-4 for more information.

3. **Connect the cables to the SPARCstorage Array and FC/S card.**
   Use the instructions in the *SPARCcluster High Availability Server Service Manual.*

4. **Perform a reconfiguration reboot of the server.**

```
ok boot -r
```

5. **When the server reboots, switch ownership of the Solstice HA services to the other host.**
   Use the `haswitch(1M)` command.

```
host1# haswitch host1 logicalhost1 logicalhost2
```

6. **Repeat Step 1 through Step 4 on the sibling Solstice HA server.**

Note – The hardware must be installed identically on each of the servers. This means the new SBus card must be installed on the same System Board and SBus slot on each server.

7. **Switch ownership of the logical hosts back to the appropriate default master.**
   For example:

```
host1# haswitch host2 logicalhost2
```

8. **Add the disks in the SPARCstorage Arrays to the selected diskset.**
   Use the instructions in Section 8.3.1, "Adding a Disk to a Diskset," on page 8-4 to add the disks to disksets.

5.2 Adding a Disk to a SPARCstorage Array

Adding a disk to a SPARCstorage Array involves taking all the metadevices in the tray offline. It is likely that the tray will contain disks from each of the disksets in a symmetric configuration.

How to Add a Disk to a SPARCstorage Array

1. **Switch ownership of both logical hosts to one of the Solstice HA servers.**
   In order to run the `ssaadm(1M)` command you must switch over both logical hosts to a single server. If you have an asymmetric configuration no switchover is required.

```
host1# haswitch host1 logicalhost1 logicalhost2
```

2. **Determine the controller number of the SPARCstorage Array where the disk will be added.**
   The World Wide Name displayed on the front of the SPARCstorage Array also appears as part of the `/devices` entry to which the `/dev` entry containing the controller number points. For example:

```
host1# ls -l /dev/rdsk | grep -i world_wide_number | tail -1
```

If the World Wide Name displayed on the front of the SPARCstorage Array is 36cc, the following output would be displayed and you would find the controller number to be c2:

```
host1# ls -l /dev/rdsk | grep -i 36cc | tail -1
lrwxrwxrwx  1 root   root       94 Jun 25 22:39 c2t5d2s7 -> ../../devices/io-
unit@f,e1200000/sbi@0,0/SUNW,soc@3,0/SUNW,pln@a0000800,201836cc/ssd@5,2:h,raw
host1#
```

3. **Locate an appropriate empty disk tray slot in the SPARCstorage Array for the disk that is being added.**
   Use the ssaadm(1M) command with the display option to view the empty slots in the SPARCstorage Array. The empty slots are shown with a NO SELECT status.

```
host1# ssaadm display c2
                          DEVICE STATUS
      TRAY 1                 TRAY 2                 TRAY 3
slot
1     Drive: 0,0             Drive: 2,0             Drive: 4,0
2     Drive: 0,1             Drive: 2,1             Drive: 4,1
3     NO SELECT              NO SELECT              NO SELECT
4     NO SELECT              NO SELECT              NO SELECT
5     NO SELECT              NO SELECT              NO SELECT
6     Drive: 1,0             Drive: 3,0             Drive: 5,0
7     Drive: 1,1             NO SELECT              NO SELECT
8     NO SELECT              NO SELECT              NO SELECT
9     NO SELECT              NO SELECT              NO SELECT
10    NO SELECT              NO SELECT              NO SELECT
host1#
```

4. **Determine the tray to which you will add the new disk.**
   In the remainder of the procedure, tray 2 will be used as an example.

   The slot selected for the new disk is tray 2 slot 7. The new disk will be known as c2t3d1.

5. **Locate all hot spares in the tray of the SPARCstorage Array.**
   To discover the status and location of all hot spares, run the `metahs(1M)` command with the `-i` option on each of the logical hosts.

```
host1# metahs -s logicalhost1 -i
...
host1# metahs -s logicalhost2 -i
...
```

Note – Save a list of the hot spares. The list will be used later in this maintenance procedure. Be sure to note the hot spare devices and the hot spare pools they are in.

---

6. **Delete all hot spares in the tray.**
   Use the `metahs` command with the `-d` option to delete the hot spares.

```
host1# metahs -s logicalhost1 -d hot_spare_pool
host1# metahs -s logicalhost2 -d hot_spare_pool
```

7. **Locate all metadevice state database replicas that are on disks in the tray of the SPARCstorage Array.**
   Run the `metadb(1M)` command on each of the logical hosts to locate all metadevice state databases. In an asymmetric configuration there will be only one logical host. Direct the output into temporary files.

```
host1# metadb -s logicalhost1 > /tmp/mddb1
host1# metadb -s logicalhost2 > /tmp/mddb2
```

Note – Save the list of the metadevice state database replicas. The list will be used later in this maintenance procedure.

Use the `metadb` command to determine on which disks in this tray metadevice state database replicas reside. Save this information for the step in which you restore the replicas.

8. **Delete the metadevice state database replicas that are on disks in the tray where the disk will be added.**
   Keep a record of the number and local of the replicas you delete. These must be restored in a later step.

```
host1# metadb -s logicalhost1 -d replicas
host1# metadb -s logicalhost2 -d replicas
```

9. **Run the `metastat` command to determine all the metadevice components in the tray.**
   The output from `metastat` should be directed to a temporary file so the information can be used later when deleting and re-adding the metadevices.

```
host1# metastat -s logicalhost1 > /tmp/log1
host1# metastat -s logicalhost2 > /tmp/log2
```

10. **Take all submirrors on the tray offline.**
    Use the temporary files to create a script to take all submirrors on the tray offline. If there are only a few submirrors on the tray you can run the `metaoffline(1M)` command to take each offline. The following is an example script.

```
#!/bin/sh
# metaoffline -s <setname> <metamirror> <submirror>

metaoffline -s logicalhost1 d15 d35
metaoffline -s logicalhost2 d15 d35
...
```

11. **Spin down the disks in the SPARCstorage Array tray.**
    Use the `ssaadm(1M)` command to perform this step.

```
host1# ssaadm stop -t 2 c2
```

12. **Remove the tray, insert the disk, and replace the tray.**
    Use the instructions in the *SPARCstorage Array Model 100 Series Service Manual* to perform the hardware procedure of adding the disk.

13. **Make sure all disks in the SPARCstorage Array tray spin up.**
    The disks in the SPARCstorage Array tray should automatically spin up following the hardware procedure. If the tray fails to spin up automatically within two minutes, force the action by using the following command:

```
host1# ssaadm start -t 2 c2
```

14. **Bring the submirrors in the tray back online.**
    Modify the script you created in Step 10 to bring the submirrors back online.

```
#!/bin/sh
# metaonline -s <setname> <metamirror> <submirror>

metaonline -s logicalhost1 d15 d35
metaonline -s logicalhost2 d15 d35
...
```

15. **Add back the hot spares that were deleted in Step 6.**

```
host1# metahs -s logicalhost1 -a hot_spare_devices
host1# metahs -s logicalhost2 -a hot_spare_devices
```

16. **Be sure to restore the original count of metadevice state database replicas to the devices. The replicas were removed in Step 8.**

```
host1# metadb -s logicalhost1 -a replicas
host1# metadb -s logicalhost2 -a replicas
```

17. **Run the** `drvconfig(1M)` **and** `disks(1M)` **commands to create the new entries in** `/devices`**,** `/dev/dsk`**, and** `/dev/rdsk` **for the drive that was added.**

```
host1# drvconfig
host1# disks
```

18. **Switch ownership of both logical hosts to** `host2`**.**

```
host1# haswitch host2 logicalhost1 logicalhost2
```

19. **Run the** `drvconfig` **and** `disks` **commands on** `host2`**.**

```
host2# drvconfig
host2# disks
```

20. **Add the disk to a diskset.**
    In this example, the disk is being added to the diskset that is mastered by `host2`.

```
host2# metaset -s logicalhost2 -a drivename
```

21. **Perform usual administration actions.**
    You can now perform the usual administration steps that are done when a new drive is brought into service. These include partitioning the disk, adding it to the configuration as a hot spare, or configuring it as a metadevice.

22. **Switch each logical host back to its default master.**

```
host2# haswitch host1 logicalhost1
```

5.3 *Adding a Public Network*

Adding a public network connection in a Solstice HA configuration involves both software and hardware procedures.

▼ How to Add a Public Network Connection

1. **Run `haswitch` to move the data services that are running to a single host.**
   In this example, `host2` will be the first to receive the new public network connection.

```
host1# haswitch host1 logicalhost1 logicalhost2
```

2. **Stop the membership monitor.**

```
host2# /etc/init.d/SUNWhadf stop
```

3. **Create the entries in `/etc/inet/hosts`, `/etc/inet/netmasks`, and network name services.**
   Add the new hostnames to the local `/etc/inet/hosts` file on both Solstice HA servers and the network name service. If this is a new network number make the appropriate entries in the `/etc/inet/netmasks` file.Assign additional logical hostnames on this new network for each logical host. Make the entries to the `/etc/inet/hosts` files and network name service at any time, before the new network is added.

4. **Create the `/etc/hostname.`*xxn* file for the new interface.**
   The `/etc/hostname.`*xxn* file must contain the hostname associated with the new interface. (Replace the *xxn* suffix with the type and number of the interface (for example `qe3` or `le4`). If this is performed before the server is halted, `ifconfig(1M)` will automatically assign an address to the network interface when the server is rebooted.

   If the name of the new interface (for instance, `qe3`) cannot be determined before the reboot, run the `prtconf(1M)` command after the next reboot and use that output to learn the new address.

   In that case either another reboot or several manual invocations of `ifconfig` are needed to bring the interface into service.

5. **Halt the server.**

```
host2# halt
```

6. **Install the network hardware using the instructions in the appropriate hardware manual.**

7. **Perform a reconfiguration reboot on the server.**
   The reboot will automatically start the membership monitor, however the server will not take back ownership of the diskset or data services.

```
ok boot -r
```

8. **Repeat the entire procedure on the sibling server.**

9. **Edit the `hadfconfig(4)` file on both servers.**
   After the new interfaces are appropriately configured and entries have been added to the name service, `/etc/hosts` file, and the `/etc/hostname.`*xxn* file, the interface will be added to the `/etc/opt/SUNWhadf/hadf/hadfconfig` file. Entries to the `hadfconfig` file are made in the following format:.

```
HOSTNAME host1-nnn logicalhost1-nnn host2-nnn logicalhost2-nnn
```

Following the naming convention, in the above example the *nnn* represents the third octet of the associated network number.

Note – If this is an asymmetric configuration with a single logical host (diskset), the string `logicalhost2-`*nnn* is replaced with a hyphen (`-`). These changes must be made manually on both machines for correct operation.

Services will be offered via the logical hosts on the next membership reconfiguration. Refer to “Forcing a Membership Reconfiguration” on page 9-2.

10. **Run the** hacheck(1M) **command on both hosts.**

```
host2# hacheck
```

5.4 *Adding Board-Level Modules*

Adding or replacing board-level modules such as SIMM and CPUs involves both software and hardware procedures.

▼ How to Add Board-Level Modules

1. **Run** haswitch **to move the data services that are running to a single host.**
   In this example, host2 will be the first to receive the board-level module.

```
host1# haswitch host1 logicalhost1 logicalhost2
```

2. **Stop the membership monitor.**

```
host2# /etc/init.d/SUNWhadf stop
```

3. **Halt the server.**

```
host2# halt
```

4. **Power off the server.**

5. **Perform a reconfiguration reboot.**

```
ok boot -r
```

6. **Install the board-level module using the instructions in the appropriate hardware manual.**

7. **Power on the server.**
   The server will rejoin the configuration.

8. **Run** haswitch **to move the data services to the host that has just received the additional board-level modules.**

9. **Repeat Step 2 through Step 6 on the sibling server.**
   In order to maintain a symmetric hardware configuration, both servers must have exactly the same hardware installed.

10. **Switch each logical host back to its default master.**

HA-NFS Maintenance 6

This chapter explains the maintenance procedures for administering HA-NFS and working with UFS logs.

Use the following table to locate specific information in this chapter.

6.1 Adding a Logging UFS File System to a Logical Host

▼ How to Add a Logging UFS File System to a Logical Host

1. **Add an entry for the logging UFS file system to the `vfstab.`*logicalhost* file using `hafstab(1M)`.**

2. **Run `mount(1M)` to mount the new file system.**
   Alternatively, you can wait until the next membership reconfiguration for the file system to be automatically mounted.

3. **Add the HA-NFS file system to the logical host.**
   You would perform this step only if the logging UFS file system is going to be an HA-NFS file system. If this is the case, follow the procedure in Section 6.3, "Adding an HA-NFS File System to a Logical Host," on page 6-2.

6.2 Removing a Logging UFS File System From a Logical Host

▼ How to Remove a Logging UFS File System From a Logical Host

1. **Remove the logging UFS file system entry from the `vfstab.`*logicalhost* file by using the `hafstab` command.**
   Refer to Chapter 10 of the *SPARCcluster High Availability Software Planning and Installation Guide* for instruction on using the `hafstab` command.

2. **Run the `umount(1M)` command to unmount the file system.**

3. **(Optional) Clear the associated trans device and its mirrors using either the `metaclear(1M) -r` or the `metatool(1M)` command.**
   Refer to the *Solstice DiskSuite 4.0 Administration Guide* or the *Solstice DiskSuite Tool 4.0 User's Guide* for instructions on clearing trans devices.

6.3 Adding an HA-NFS File System to a Logical Host

▼ How to Add an HA-NFS File System to a Logical Host

1. **Make the appropriate entry for each logging UFS file system that will be shared by HA-NFS in the `vfstab.`*logicalhost* file by using the `hafstab` command.**

2. **Make the corresponding entry in the `dfstab.`*logicalhost* file by using the `hafstab` command.**
   Refer to Chapter 10 of the *SPARCcluster High Availability Software Planning and Installation Guide* for instruction on using the `hafstab` command.

3. **Execute a reconfiguration reboot of the server.**
   Alternatively, the file system may shared manually. If the procedure is performed manually, the fault monitoring processes will not be started either locally or remotely until the next membership reconfiguration is performed.

```
host1# touch /reconfigure
```

6.4 Removing an HA-NFS File System From a Logical Host

▼ How to Remove an HA-NFS File System From a Logical Host

1. **Remove the entry for the HA-NFS file system from the `dfstab.`*logicalhost* file by using the `hafstab` command.**
   Refer to Chapter 10 of the *SPARCcluster High Availability Software Planning and Installation Guide* for instructions on using the `hafstab` command.

2. **Run the `unshare(1M)`command.**
   The fault monitoring system will try to access the file system until the next membership reconfiguration. Errors will be logged but a takeover of services will not be initiated by the Solstice HA software.

3. **(Optional) Remove the logging UFS file system from the logical host. If you want to retain the UFS file system for a non-HA-NFS purpose, such as an HA-ORACLE file system, skip to Step 4.**
   To perform this task, use the procedure in Section 6.2, “Removing a Logging UFS File System From a Logical Host,” on page 6-2.

4. **Execute a reconfiguration reboot of the server.**

```
host1# touch /reconfigure
```

6.5 Changing Share Options on an HA-NFS File System

If you use the `rw`, `rw=`, `ro`, or `ro=` options to the `share` command, HA-NFS fault monitoring will work best if you grant access to all the physical hostnames associated with both Solstice HA servers. Refer to the list of hostnames you entered during the configuration of your Solstice HA servers.

If you use netgroups in the share command, rather than the names of individual hosts, add all of the Solstice HA hostnames to the appropriate netgroup. Ideally, you should grant both read and write access to all the Solstice HA hosts' hostnames to enable the NFS fault probes to do a complete job.

▼ How to Change Share Options on an HA-NFS File System

1. **Make the appropriate share changes to the `dfstab.`*logicalhost* file using the `hafstab(1M)` command.**
   Refer to Chapter 10 of the *SPARCcluster High Availability Software Planning and Installation Guide* for instruction on using the `hafstab` command.

2. **Perform a membership reconfiguration using the instructions in Section 9.1, "Forcing a Membership Reconfiguration," on page 9-2.**
   If a reconfiguration is not possible at this time, you can run the `share(1M)` command with the new options. Some changes may cause the fault monitoring subsystem to issue messages. For instance, a change from read-write to read-only will generate messages.

HA-ORACLE Maintenance 7

This chapter gives instructions for the maintenance procedures that may need to be performed when you are running HA-ORACLE.

Use the following table to locate specific information in this chapter.

7.1 Overview of HA-ORACLE Operations

Once configured, the administration of HA-ORACLE should be the same as administration of any ORACLE database.

During administration, you should never edit or delete the `haoracle_support(4)` or the `haoracle_config_v1(4)` files.

Note – You should never use the ORACLE mirroring utilities to create mirrors in a Solstice HA configuration. You must use Solstice DiskSuite mirroring for ORACLE database on both UFS file systems and on raw devices.

7.2 *HA-ORACLE Fault Monitoring*

The `status` field in the `haoracle_databases(4)` file should always be set to `on`. When the field is set to `on`, the instance is considered to be under Solstice HA fault monitoring. HA-ORACLE will start up and shut down the instance and will provide fault monitoring.

Caution – You should never set the `status` field to `off`, because it will impact availability. When the field is set to `off`, the database is considered to be in maintenance mode and is not maintained by HA-ORACLE. The database will not be started or shut down during membership transactions and there will be no fault monitoring. If the database has a problem, it will not be detected by HA-ORACLE.

7.3 *Shutting Down and Starting Databases*

Before performing any maintenance on an HA-ORACLE database, the database should be taken out of service. Use the following procedure to perform this task:

1. **Stop the database instance by using the** `haoracle(1M)` **command.**

```
# haoracle stop instance_name
```

2. **Shut down the database instance by using the ORACLE** `sqldba(1M)` **command.**

3. **Restart the database instance by using the haoracle command.**

```
# haoracle start instance_name
```

7.4 *Adding an HA-ORACLE Database Instance*

Use the instructions in Chapter 10 of the *SPARCcluster High Availability Software Planning and Installation Guide* to install, set up, or create a ORACLE database instance in a Solstice HA configuration.

Metadevice and Diskset Administration

8

This chapter provides instructions for administering shared metadevices and explains metadevice actions that are not supported in Solstice High Availability configurations.

Use the following table to locate specific information in this chapter.

8.1 Overview of Metadevice and Diskset Administration

Metadevices and disksets are created and administered using either Solstice DiskSuite 4.0 command-line utilities or the DiskSuite Tool (`metatool(1M)`) graphical user interface.

Your primary source of information about administration of Solstice DiskSuite devices will be the *Solstice DiskSuite 4.0 Administration Guide* and *Solstice DiskSuite Tool 4.0 User's Guide*.

Read the information in this chapter before using the DiskSuite documentation to administer disksets and metadevices in a Solstice HA configuration.

Disksets are groups of disks. The primary administration task you will perform on disksets involves adding and removing disks.

Before using a physical device that you have placed in a diskset, you must set up a metadevice using the component. A metadevice consist of concatenations, stripes, mirrors, UFS logs, or hot spares.

Note – Metadevice names begin with "`d`" and are followed by a number. By default there are 128 unique metadevices in the range 0 to 127. Each UFS logging device you create will use at least seven metadevice names. Thus, in a large Solstice HA configuration, you may need more than the 128 default metadevice names. Refer to Appendix A of the *Solstice DiskSuite 4.0 Administration Guide* for instructions on changing the number.

The *Solstice DiskSuite 4.0 Administration Guide* tells you the only modifiable field in the `/kernel/drv/md.conf` file is the `nmd` field. You can also modify the `md_nsets` field from 4 to 3.

8.2 Mirroring Guidelines

Corresponding submirror components must be on different SPARCstorage Array chassis. This ensures all mirrored data will survive a SPARCstorage chassis failure.

The need for online service of Solstice HA servers dictates a somewhat stronger requirement, that no two components in different submirrors of a single mirror may be in the same SPARCstorage Array chassis or tray. This allows submirrors to be taken offline and either the chassis to be powered off or the tray to be spun down and removed.

Additional information about mirroring guidelines can be found in the *SPARCcluster High Availability Software Planning and Installation Guide.*

8.3 Diskset Administration

Diskset administration consists of adding and removing disks from the diskset. The steps for these procedures are included in the following subsections.

Note – If the logical hosts are up and running you should never perform diskset administration using either the `-t` (take ownership) or `-r` (release ownership) options of the `metaset(1M)` command. These options are used internally by the Solstice HA software and must be coordinated between the two servers.

If the logical host is in maintenance mode, as reported by `hastat(1M)`, you can safely use the `metaset -t` command to take ownership of the diskset. However, before returning the logical host to service you must release the diskset ownership using the `metaset -r` command.

To place a logical host in maintenance mode, run the `haswitch -m` command.

If a SPARCstorage Array tray must be spun down to perform the administration tasks, you must first have ownership of any diskset on the tray. Ownership may be changed using the `haswitch(1M)` command.

You cannot use the `ssaadm(1M)` command to perform maintenance on a SPARCstorage Array if the sibling host has ownership of the logical host. If the logical host is either in maintenance mode or is locally owned you can use the `ssaadm` command.

Before using the `ssaadm` command, you must make sure that no drives in the tray are owned by the sibling server. If you see the following message when running the ssaadm command with the `display` argument, there are drives in the tray owned by the sibling server.

```
host1# ssaadm display c1
ssaadm: Close: I/O error
host1#
```

The procedures in Chapter 4, “Hardware Replacement and Repair,” or Chapter 5, “Adding Hardware,” provide additional details about performing maintenance procedures.

8.3.1 Adding a Disk to a Diskset

If the disk that is being added to a diskset will be used as a submirror, you must have two disks available on two different SPARCstorage Arrays to allow for mirroring. However, if the disk will be used as a hot spare, you can add only one.

▼ How to Add a Disk to a Diskset

1. **Ensure there is no data on the disk because the partition table will be rewritten and space for a metadevice state database replica will be allocated on the disk.**

2. **Insert the disk device into the SPARCstorage Array using the instructions in the** *SPARCstorage Array Model 100 Series Service Manual.*

3. **Invoke the following command from the Solstice HA server that owns the diskset:**

```
# metaset -s logicalhost -a diskname
```

8.3.2 Removing a Disk From a Diskset

You can remove a disk from a diskset at any time, provided that none of the slices on the disk are currently in use in metadevices or hot spare pools. To find out if any of the slices are in use, you would use the `metastat(1M)` command.

The steps to follow when removing a disk from a diskset are in the following section.

▼ How to Remove a Disk From a Diskset

1. **Ensure that none of the slices are in use as metadevices or as hot spares.**

2. **Invoke the following command from the Solstice HA server that owns the diskset:**

```
# metaset -s logicalhost -d diskname
```

The metaset command automatically discovers if a metadevice state database replica existed on the disk and if so it finds a suitable location for a replacement replica on another disk.

8.4 Multi-host Metadevice Administration

The following subsections contain information about the differences in administering metadevices in the multi-host Solstice HA environment versus a single host environment.

Unless noted in the following subsections, the instructions in the *Solstice DiskSuite 4.0 Administration Guide* and the *Solstice DiskSuite Tool 4.0 User's Guide* can be used.

Note – Before using the instructions in either of the Solstice DiskSuite manuals, check in the SPARCcluster documentation set. The instructions in the Solstice DiskSuite manuals deal only with single host configurations.

The following subsections tell you the Solstice DiskSuite command-line programs you should use when performing a task. Optionally, you can use the metatool(1M) graphical user interface for all the tasks unless directed otherwise. You must remember to use the -s option when running metatool. The -s option allows you to specify the diskset name.

8.4.1 Managing Metadevices

For ongoing management of metadevices, you must constantly monitor the metadevices for errors in operation, as discussed in Chapter 3, “Monitoring the Solstice HA Servers.”

Use the hastat(1M) command to monitor the status of the disksets. When hastat reports a problem with a diskset, you can use the metastat(1M) command to locate the errored metadevice.

You must use the -s option when running either metastat or metatool. The -s option allows you to specify the diskset name.

8.4.2 Adding a Mirror to a Diskset

Mirrored metadevices may be used directly by HA-DBMS applications. They may also be used as part of a logging UFS file system for either HA-NFS or HA-ORACLE applications.

Idle slices on disks within a diskset can be configured into metadevices by using the `metainit(1M)` command.

8.4.3 Removing a Mirror from a Diskset

HA-ORACLE applications may use raw mirrored metadevices for database storage. While these are not mentioned in the `dfstab.`*logicalhost* or `vfstab.`*logicalhost* files, they appear in the related HA-ORACLE configuration files. The mirror must be removed from these files and the HA-ORACLE system made to stop using the mirror. At that point the mirror may be cleared by using the `metaclear(1M)` command.

8.4.4 Taking Submirrors Offline

Before replacing or adding a disk drive in a SPARCstorage Array tray, all the metadevices on that tray must be taken offline.

In symmetric configurations, taking the submirrors offline for maintenance is complex because there may be disks from each of the two disksets on the same tray in the SPARCstorage Array. That means you must take the metadevices from each diskset offline before removing the tray.

You will use the `metaoffline(1M)` command to take all submirrors on every disk in the tray off line.

8.4.5 Creating New Metadevices

After a disk is added to a diskset, you can create new metadevices using `metainit(1M)` or `metatool`. If the new devices are going to be hot spares, you will use the `metahs(1M)` command to place the hot spares in a hot spare pool.

8.4.6 Replacing Errored Components

When replacing an errored metadevice component, you will use the `metareplace(1M)` command.

You can return drives to service that have sustained transient errors (for example, a chassis power failure) by using the `metareplace -e` command.

A replacement slice (or disk) must be available as a replacement. This device could be an existing device that is not in use or a new device you have added to the diskset.

8.4.7 Deleting a Metadevice

Before deleting a metadevice, you must ensure that none of the components in the metadevice is in use, either by a database or by HA-NFS. You will use the `metaclear(1M)` command to delete the metadevice.

8.4.8 Growing a Metadevice

To grow a metadevice you must have another two slices (disks) in different SPARCstorage Arrays available. Each of the two new slices will be added to a different submirror with the `metainit` command. You then use the `growfs(1M)` command to grow the file system.

Caution – When running the `growfs` command, clients can experience an interruption of service.

If a takeover occurs while the file system is growing, the file system will not be grown. You must reissue the `growfs` command (from the command line) after the takeover completes.

Note – The file system that contains /*logicalhost*/`statmon` cannot be grown. This is because the `statd(1M)` program modifies this directory it would be blocked for extended periods while the file system is growing. This would have unpredictable effects on the network lock protocol. This is only a problem for configurations using HA-NFS.

The following example shows the `growfs` command used to enlarge the `d30` metadevice on `host1` by 512 Mbytes. The warning printed by `growfs` reports the number of inode blocks that will not be allocated. The errors indicate that while the file system was being grown it was locked against modifications. These errors are part of the expected behavior.

```
# growfs -s 1000000 -M /host1/ufsd30 /dev/md/host1/rdsk/d30
Warning: inode blocks/cyl group (133) >= data blocks (4) in last cylinder group.
    This implies 64 sector(s) cannot be allocated.
/dev/md/host1/rdsk/d30:     999936 sectors in 992 cylinders of 14 tracks, 72 sectors
    488.2MB in 62 cyl groups (16 c/g, 7.88MB/g, 3776 i/g)
super-block backups (for fsck -F ufs -o b=#) at:
 32, 16240, 32448, 48656, 64864, 81072, 97280, 113488, 129696, 145904, 162112,
 178320, 194528, 210736, 226944, 243152, 258080, 274288, 290496, 306704,
 322912, 339120, 355328, 371536, 387744, 403952, 420160, 436368, 452576,
 468784, 484992, 501200, 516128, 532336, 548544, 564752, 580960, 597168,
 613376, 629584, 645792, 662000, 678208, 694416, 710624, 726832, 743040,
 759248, 774176, 790384, 806592, 822800, 839008, 855216, 871424, 887632,
 903840, 920048, 936256, 952464, 968672, 984880,
NFS2 setattr failed for server ha-drag: RPC: Timed out
Jun 30 14:40:55 ha-drag hadf: ERROR: Error: nfs_mon: ftruncate of
'/var/opt/SUNWhadf/hadf/nfs_probe_mountpoints/_host1_ufsd30.locking/.probe_nfs/probefile'
failed, errno=145 'Connection timed out'
Jun 30 14:40:55 host1 hadf: ERROR: Error: nfs_mon: This problem is with this host.
#
```

8.4.9 *Managing Hot Spare Pools*

Hot spare devices can be added and deleted from hot spare pools at any time, providing they are not in use. In addition, you can create new hot spare pools and associate them with submirrors by using the `metahs(1M)` command.

8.4.10 *Managing UFS Logs*

All UFS logs on multi-host disks are mirrored. When a submirror fails, it is reported as an errored component. You repair the failure by using either `metareplace` or `metatool`.

If the entire mirror that contains the UFS log fails, you must unmount the file system, back up any accessible data, repair the error, repair the file system (using `fsck(1M)`), and remount the file system.

8.4.10.1 *Adding UFS Logging to a Logical Host*

All UFS file systems within a logical host must be logging UFS file systems to ensure the failover or `haswitch` timeout criteria can be met. A logging UFS file system may be used by HA-NFS or by an HA-DBMS data service.

The logging UFS file system is set up by first creating a trans device with a mirrored log and a mirrored UFS master file system. Both the log and UFS master device must be mirrored following the mirroring guidelines explained in Section 8.2, "Mirroring Guidelines," on page 8-2.

During Solstice HA configuration, the `hasetup(1M)` command optionally reserves space on slice 6 of each drive in a diskset for use as a UFS log. The size of these slices was determined during configuration. The slices can be used for UFS log submirrors. If the slices are smaller than the desired log size, several can be concatenated. Typically one Mbyte per 100 Mbytes is adequate for UFS logs. Ideally, log slices would be drive-disjoint from the UFS master device.

Note – If you must repartition the disk to gain space for UFS logs, you must preserve the existing slice 7 which starts on cylinder 0 and contains at least two Mbytes. This space is required and reserved for metadevice state database replicas. The `Tag` and `Flag` fields (as reported by the `format(1M)` command) must be preserved for slice 7. The `metaset(1M)` command sets the `Tag` and `Flag` fields correctly when the initial configuration is built.

After the trans device has been configured, the UFS file system must be created by using `newfs(1M)` on the trans device. This typically takes about two minutes per Gbyte of UFS master device size.

After the `newfs` process is completed, the UFS file system can be added to the `vfstab.`*logicalhost* file with the `hafstab(1M)` command. If this file system is for use by an HA-DBMS application, the HA-DBMS configuration files should be updated.

If the file system will be shared by HA-NFS, follow the procedure in Section 6.3, "Adding an HA-NFS File System to a Logical Host," on page 6-2.

The new file system will be automatically mounted at the next membership monitor reconfiguration. It may be manually mounted if membership monitor reconfiguration is not possible or desirable.

8.5 Local Metadevice Administration

Local disks are not mirrored, however some local file systems may have UFS logs. If an error occurs, you will perform the same actions as when the entire mirror fails, as specified in “Managing UFS Logs” on page 8-8.

If the entire root disk fails, use the instructions in “How to Replace a Failed Local Boot Disk” on page 4-13.

8.6 Destructive Metadevice Actions

The metadevice actions that are not supported in Solstice HA configurations include:

- Creation of a diskset with fewer than three disks
- Creation of a diskset that is attached to fewer than three controllers
- Creation of a one-way mirror in a diskset
- Creation of a configuration with too few metadevice state database replicas on the local disks
- Creation of metadevice state database replicas on multi-host disks without the use of the `metaset(1M)` command, unless directed to do so in explicit instructions in this or another SPARCcluster manual.

General Solstice HA Maintenance 9

This chapter gives instructions for general maintenance procedures such as restarting failed servers in Solstice HA configurations.

Use the following table to locate specific information in this chapter.

9.1 Forcing a Membership Reconfiguration

A membership reconfiguration can be forced by changing ownership of a logical host.

A switchover (using `haswitch(1M)`) accomplishes this task, however you will be required to perform a second switchover in order to restore the original configuration, that is, have the logical hosts associated with the default masters.

Another method to perform a membership reconfiguration is to use the internal utility, `clustm`. To perform a membership reconfiguration, enter the following:

```
# /etc/SUNWcluster/bin/clustm reconfigure hadf
```

9.2 Public Network Administration

Adding and removing public network connections in Solstice HA configurations involves software procedures in addition to making the hardware connections. The instructions for adding a public network can be found in Chapter 5, “Adding Hardware.” Use the following procedure to remove a public network.

Caution – If you perform an initial install of Solaris in the future, the removal of the network interface SBus cards may cause the numbering of the remaining network to change.

▼ How to Remove a Public Network

1. **Notify users the subnet is going to be removed.**
   Make sure the users are off the subnet.

2. **Remove, or comment out, the appropriate** `HOSTNAME` **line in the** `/etc/opt/SUNWhadf/hadf/hadfconfig` **file on both Solstice HA servers.**

3. **Perform a membership monitor reconfiguration.**
   The logical hosts on the associated network will cease offering services following the membership reconfiguration. To perform a membership reconfiguration, enter the following:

```
# /etc/SUNWcluster/bin/clustm reconfigure hadf
```

4. **On each server, determine which interface and logical interfaces will be removed.**
   After the membership reconfiguration, Solstice HA will forget about the network, but does not completely clean up. The `ifconfig -a` command will report that the associated logical interfaces are still up, as follows:

```
host1# ifconfig -a
le5: flags=863<UP,BROADCAST,NOTRAILERS,RUNNING,MULTICAST> mtu
    1500 inet 192.9.76.12 netmask ffffff00 broadcast 192.9.76.255
     ether 8:0:20:1c:b2:92
le5:1: flags=843<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
        inet 192.9.76.18 netmask ffffff00 broadcast 192.9.76.255
le5:2: flags=842<BROADCAST,RUNNING,MULTICAST> mtu 1500
        inet 192.9.77.13 netmask ffffff00 broadcast 192.9.77.255
```

5. **Execute the `ifconfig down` commands on both servers. The following commands must be entered on both servers.**

```
host1# ifconfig le5:1 down
host1# ifconfig le5:2 down
host1# ifconfig le5 down
host1# ifconfig le5 unplumb
```

6. **On each server, remove the `/etc/hostname.`*nnn* file that is associated with the interface.**
   This step is only necessary if the hardware is removed from both servers.

```
host1# rm /etc/hostname.le5*
```

7. **Switch ownership of both logical hosts to one Solstice HA server using a command similar to the following:**

```
host1# haswitch host1 logicalhost1 logicalhost2
```

8. **Stop the membership monitor and halt the server that will have the hardware removed first.**
   After entering the following commands, turn off the server.

```
host2# /etc/init.d/SUNWhadf stop
host2# halt
```

9. **Remove the hardware from the server that has been halted.**
   Use the procedure from the *SPARCcluster High Availability Server Service Manual* to remove the hardware.

10. **Perform a reconfiguration reboot on the server.**

```
ok boot -r
```

11. **Switch ownership of both logical hosts to the server that has already had the hardware removed.**

```
host2# haswitch host2 logicalhost1 logicalhost2
```

12. **Stop the membership monitor and halt the other server.**
    After entering the following commands, turn off the server.

```
host1# /etc/init.d/SUNWhadf stop
host1# halt
```

13. **Remove the hardware from the server that has been halted.**
    Use the procedure from the *SPARCcluster High Availability Server Service Manual* to remove the hardware.

14. **Perform a reconfiguration reboot on the server.**

```
ok boot -r
```

15. **Switch ownership of the logical hosts to the default master.**

```
host1# haswitch host1 logicalhost1
```

9.3 Shutting Down Solstice HA Servers

You may have to shut down one or both Solstice HA servers to perform hardware maintenance such as adding or removing SBus cards. The following sections describe the procedure for shutting down a single server or the entire configuration.

▼ How to Shut Down One Server

If you want the data in a logical host (diskset) to remain available when a server is shut down, you must first switch ownership of the diskset to the other server using the `haswitch` command.

If it is not necessary to have the data available, the logical host (diskset) can be placed in maintenance mode. Refer to Section 9.4.1, “Putting Logical Hosts in Maintenance Mode,” on page 9-7 for additional information.

Note – It is possible to halt (`halt(1M)`) a Solstice HA server and allow a takeover to restore the logical host services on the other server. The `halt` may cause the server to panic. However, the `haswitch` command offers a more reliable method of switching ownership of the logical hosts.

To stop running Solstice HA on a server while leaving services running on the sibling, enter the following commands:

```
host1# haswitch host2 logicalhost1 logicalhost2
host1# /etc/init.d/SUNWhadf stop
```

At this point you should halt the server.

```
host1# halt
```

How to Shut Down a Solstice HA Configuration

You may want to shut down both servers in a Solstice HA configuration should a bad environmental condition exist, such as a cooling failure or a severe lightning storm.

1. **Stop the membership monitor and `halt` one of the servers.**

```
host1# /etc/init.d/SUNWhadf stop
host1# halt
```

2. **Stop the membership monitor and `halt` the sibling server.**

```
host2# /etc/init.d/SUNWhadf stop
host2# halt
```

How to Halt a Solstice HA Server

Either server can be shut down using either `halt` or `uadmin(1M)`. If the membership monitor is running when a host is shut down, the server will most likely take a “Fastfail timeout” and display the following message:

```
Panic: Failfast timeout unit “abort_thread”
```

This can be avoided by stopping the membership monitor before shutting down the server. Refer to Section 9.5, “Stopping the Membership Monitor,” on page 9-8 for additional information.

9.4 Switching Over Data Services

You will use the haswitch command to move data services from one Solstice HA server to the other. The command also allows you to put logical hosts in maintenance mode.

For example, to execute a switchover of data services from host1 to host2 (with the logical hosts being named logicalhost1 and logicalhost2), enter the following command:

```
host2# haswitch host2 logicalhost1 logicalhost2
```

9.4.1 Putting Logical Hosts in Maintenance Mode

To put the disksets of a logical host in maintenance mode, use the -m option of the haswitch command. Maintenance mode is useful for some administration on file systems and disksets.

Note – Unlike other ownership of a logical host, maintenance mode persists across server reboots. A logical host can be removed from maintenance mode only by a manual switchover to a specific Solstice HA server.

An example use of the maintenance option would be:

```
# haswitch -m logicalhost1
```

This command stops the data services associated with logicalhost1 on the Solstice HA server that currently owns the diskset and also halts the fault monitoring services associated with both Solstice HA servers. The command will also execute an unshare(1M) and umount(1M) of any file systems on the logical host. The associated diskset ownership will be released.

The command may be run on either host, regardless of current ownership of the logical host and diskset.

9.5 Stopping the Membership Monitor

To put the server in any mode other than multi-user, or to halt or reboot the server, you must first stop the Solstice HA membership monitor. You can then use your site's preferred method for further server maintenance.

The membership monitor can be stopped only when no logical hosts are owned by the local Solstice HA server. To stop the membership monitor on one host, run the following commands:

```
host1# haswitch host2 logicalhost2
host1# /etc/init.d/SUNWhadf stop
```

If a logical host is owned by the server when the `stop` command is run, ownership will be transferred to the other Solstice HA host before the membership monitor is stopped.

If the other Solstice HA server is down, the command will take down the data services in addition to stopping the membership monitor.

To stop the membership monitor on both Solstice HA servers, run the `haswitch` command and stop the membership monitor on one of the servers. Then run the following command on the second server:

```
# /etc/init.d/SUNWhadf stop
```

9.6 Changing the Time in Solstice HA Configurations

A simple time synchronization protocol is run on both Solstice HA servers that ensures the clocks stay close to each other. The "window of error" is roughly three seconds. If a failover or switchover occurs during that period, the time stamp difference for HA-NFS clients will go unnoticed.

This synchronization is only within the Solstice HA configuration. No reference is made by the servers to external time standards that may be used at your site. For this reason, the time on the Solstice HA servers may drift out of sync with other hosts on the network.

⚠ **Caution –** There is no way for an administrator to adjust the time of the servers in a Solstice HA configuration. Never attempt to perform a time change using either the `date(1)` or the `rdate(1M)` commands.

9.7 *Setting the OpenBoot PROM*

For correct Solstice HA operation, the OpenBoot PROM options on both servers should be set to the factory defaults with the exception of the `watchdog-reboot?` variable which should be set to `true`. The default setting for `auto-boot?`, which is `true`, should not be changed. These settings ensure that Solstice HA servers will boot upon power up and after a kernel watchdog reset.

Note – Under some circumstances, the Solstice HA software may execute a `halt(1M)` command on a server rather than a `reboot(1M)` command. This is ordinarily confined to initial configuration problems. In this case, the server must be manually booted to return to service.

▼ How to Set the OpenBoot PROM

1. **Boot the Solstice HA server in single user mode or run the `eeprom(1M)` command.**

2. **Run the following commands to set the variables from the OpenBoot PROM. The OpenBoot `printenv` command was used to check the values.**

```
ok set-defaults
Setting NVRAM parameters to default values.
ok setenv watchdog-reboot? true
watchdog-reboot?= true
ok printenv
Parameter Name      Value            Default Value
...
sbus-probe-list1    0123             0123
sbus-probe-list0    0123             0123
fcode-debug?        false            false
auto-reboot?        true             true
watchdog-reboot?    true             false
...
```

Alternatively, the `eeprom` command can be used to set the OpenBoot PROM variables. For example:

```
# eeprom 'auto-boot?'
auto-boot?=true
# eeprom 'watchdog-reboot?'
watchdog-reboot?=false
# eeprom 'watchdog-reboot?=true'
# eeprom 'watchdog-reboot?'
watchdog-reboot?=true
#
```

9.8 *Maintenance of the `/var` File System*

Because Solaris and Solstice HA software error messages are written to the `/var/adm/messages` file, the `/var` file system may become full. If this happens when the server is running, the server will continue to run. In most instances, you will not be able to log into the server that has the full `/var` file system. Should the server go down, Solstice HA will not start and a login will not be possible.

If the server goes down you must reboot in single user mode (`boot -s`).

If the server reports a full /var file system and continues to run Solstice HA services, follow the steps in the following section. In the following procedure, host1 has a full /var file systems.

▼ How to Repair a Full /var File System

1. **Perform a switchover.**

```
host2# haswitch host2 logicalhost1 logicalhost2
```

2. **Stop the Solstice HA services.**
   If you have an existing shell open to host1, enter the follow:

```
host1# /etc/init.d/SUNWhadf stop
```

   If you do not have a shell open to host1, use the procedure in "How to Enter the OpenBoot PROM on a Solstice HA Server" on page 10-5 to connect to the console and halt the server.

3. **Reboot the server in single user mode.**

```
ok boot -s
INIT: SINGLE USER MODE

Type Ctrl-d to proceed with normal startup,
(or give root password for system maintenance): root_password
Entering System Maintenance Mode
#
```

4. **When the server boots, locate and remove or copy the offending file.**
   Removing or copying the /var/adm/messages file to another location will not alter the Solstice HA performance.

```
# find /var -size +20480 -print
/var/adm/messages
# rm /var/adm/messages
```

   Alternatively you can search for files that have been recently modified. The following command shows all the files modified in the past 24 hours.

```
# find /var -mtime -1 -print
/var/adm/messages
/var/adm/utmp
/var/adm/utmpx
...
# rm /var/adm/messages
```

5. **Enter multi-user mode.**
   When you enter the following, the server will come up multi-user mode and will automatically rejoin the configuration.

```
# exit
```

9.9 Solstice HA Packages Maintenance

The only Solstice HA or Solstice DiskSuite packages that can safely be removed are the AnswerBook documents. By default, the AnswerBooks for Solstice DiskSuite reside in /opt/SUNWabmd and the AnswerBooks for Solstice HA reside in /opt/SUNWabha. To remove these packages use the procedure in “How to Remove Solstice HA Packages.”

If new distributions of the Solstice HA packages arrive, you can upgrade the servers using the procedure in “How to Upgrade Solstice HA Packages.”

How to Remove Solstice HA Packages

1. **Remove the packages from each of the Solstice HA servers by using the** `pkgrm(1M)` **command on each server.**

```
host1# pkgrm SUNWabha SUNWabmd
```

```
host2# pkgrm SUNWabha SUNWabmd
```

How to Upgrade Solstice HA Packages

1. **Switch ownership of both logical hosts to the Solstice HA server that will not be upgraded first.**
   In this example, `host2` will be the first one upgraded with new packages, so `host1` is taking ownership of both logical hosts.

```
host2# haswitch host1 logicalhost1 logicalhost2
```

2. **Stop the membership monitor on** `host2`**.**

```
host2# /etc/init.d SUNWhadf stop
...
```

3. **Remove the existing packages by using the** `pkgrm` **command.**
   It does not matter in which order the packages are removed.

```
host2# pkgrm SUNWhaor SUNWhanfs SUNWhagen SUNWcmm SUNWff ...
```

4. **Insert the CD that contains the new software into the CD-ROM drive and change directories to root.**
   There is a brief delay while Solaris scans the CD and mounts the proper file systems.

```
host2# cd /
```

5. **Enter the following command to install the three Solstice DiskSuite packages. Enter** `y` **at any prompts about changing modes on directories.**

```
host2# pkgadd -d /cdrom/cdrom0 SUNWhagen SUNWhanfs ...
```

6. **Check the contents of the packages.**

```
host2# pkgchk -n SUNWhagen SUNWhanfs ...
```

7. **Start the membership monitor.**

```
host2# /etc/init.d/SUNWhadf start
```

8. **Switch ownership of both logical hosts to the host that has just been upgraded.**

```
host2# haswitch host2 logicalhost1 logicalhost2
```

9. **Repeat the upgrade procedure shown in Step 2 through Step 7.**

10. **Switch ownership of the logical hosts back to the appropriate default master.**
    For example:

```
host1# haswitch host1 logicalhost1
```

9.10 Bringing Up Servers Without Starting Solstice HA

You may need to bring up a server without starting the Solstice HA software. One reason you would need to start a server without running Solstice HA is if the `vfstab.`*logicalhost* file is lost or becomes corrupt. This is possible because the Solstice HA software is started at run level 3.

▼ How to Bring Up Servers Without Starting Solstice HA

1. **Boot the server to single-user mode.**

```
# boot -s
...
INIT: SINGLE USER MODE

Type Ctrl-d to proceed with normal startup,
(or give root password for system maintenance): root_password
Entering System Maintenance Mode

#
```

2. **Bring the server up to run level 2 by using the** `init(1M)` **command.**
   Run level 2 has all normal file systems mounted and non-Solstice HA network services started.

```
# init 2
INIT: New run level: 2
The system is coming up.  Please wait.
checking ufs filesystems
/dev/rdsk/c0t0d0s7: is stable.
/dev/rdsk/c0t0d0s5: 665 files, 73552 used, 111295 free
/dev/rdsk/c0t0d0s5: (199 frags, 13887 blocks, 0.1% fragmentation)
NIS domainname is host1.West.COM
starting router discovery.
starting rpc services: rpcbind keyserv ypbind kerbd done.
Setting netmask of lo0 to 255.0.0.0
Setting netmask of be0 to 255.255.255.0
Setting netmask of be1 to 255.255.255.0
Setting netmask of le0 to 255.255.255.0
Setting default interface for multicast: add net 555.0.0.0:
gateway host1-drag
syslog service starting.
volume management starting.
The system is ready.
soc0: port 0: Fibre Channel is ONLINE
soc1: port 0: Fibre Channel is ONLINE
soc2: port 0: Fibre Channel is ONLINE

console login:
```

3. **After you perform the desired maintenance procedure, start Solstice HA by rebooting the server.**
   If the server is brought back up using `init 3`, some daemons are restarted and many system error messages appear. These are avoided with the following command:

```
# reboot
```

9.11 Changing the Host Name of a Server or a Logical Host

Changing the host name of a server in a Solstice HA configuration is a complex procedure. This procedure should only be performed by a trained service representative.

Renaming a logical host is not possible.

Using the Terminal Concentrator 10

This chapter gives instructions for using the terminal concentrator when performing administration of Solstice HA configurations.

Use the following table to locate specific information in this chapter.

10.1 Connecting to the Solstice HA Server Console

You can perform administrative tasks from a window connected to either Solstice HA server. The procedures for initial set up of a terminal concentrator and how to set up security can be found in the *SPARCcluster High Availability Software Planning and Installation Guide* and in the terminal concentrator documentation.

The following procedure tells you how to create connections from the administrative workstation in a Solstice HA configuration.

▼ How to Connect to the Solstice HA Server Console

1. **Open a** `shelltool(1)` **window on the desktop of the administration workstation.**

2. **Note the size of the** `shelltool` **window.**
   Run the `tput(1)` command and note the size of the `shelltool` window. This number will be used in Step 6.

```
# tput lines
35
#
```

3. **Open a** `telnet(1)` **connection to one of the Solstice HA servers through the terminal concentrator.**
   Run the following command to open the `telnet` connection to a server in a window.

```
# telnet terminal_concentrator_name 5002
Trying 192.9.200.1 ...
Connected to 192.9.200.1.
Escape character is '^]'.
```

4. **Optionally, open another** `shelltool` **window and run the following command to open a** `telnet` **connection to the other server.**

```
# telnet terminal_concentrator_name 5003
Trying 192.9.200.1 ...
Connected to 192.9.200.1.
Escape character is '^]'.
```

Note – If you set up security as described in the *SPARCcluster High Availability Software Planning and Installation Guide*, you will be prompted for the port password. After establishing the connection, you will be prompted for the login name and password.

5. **Login to the server.**

```
Console login: root
Password: root_password
```

6. **Reset the terminal rows attribute to the number found in Step 2. You will use the `stty(1)` command for this procedure.**
   A problem exists with the default `shelltool` size and the size reported by the console terminal type. By entering the following, the information these two values will agree.

```
# stty rows 35
```

10.2 Resetting Terminal Concentrator Connections

This section contains instructions for resetting the terminal concentrator connection.

If another user has a connection to the Solstice HA server console port on the terminal concentrator, you can reset the port to disconnect that user. This procedure will be useful if you need to immediately perform an administrative procedure.

If you cannot connect to the terminal concentrator, the following message will be displayed:

```
# telnet xx-tc 5002
Trying 192.9.200.1 ...
telnet: Unable to connect to remote host: Connection refused
#
```

If you use the port selector, you may see a port busy message.

▼ How to Reset a Terminal Concentrator Connection

1. **Connect to the terminal concentrator and select the command line interface (`cli`). Remember to press an extra return after making the connection.**

```
# telnet xx-tc

Enter Annex port name or number: cli
```

2. **Enter the** `su` **command and passwd.**
   By default, the password is the IP address of the terminal concentrator.

```
annex: su
Password:
```

3. **Discover which port you want to reset.**
   The port in this example is port 2. Use the terminal concentrator's built-in `who` command to show connections.

```
annex# who
```

4. **Reset the port.**
   Use the terminal concentrator's built-in `reset` command to reset the port. This example breaks the connection on port 2.

```
annex# admin reset 2
```

5. **Disconnect from the terminal concentrator.**

```
annex# hangup
```

6. **Reconnect to the desired port.**

```
# telnet xx-tc 5002
```

10.3 Entering the OpenBoot PROM

This section contains information for entering the OpenBoot PROM from the terminal concentrator.

▼ How to Enter the OpenBoot PROM on a Solstice HA Server

1. Connect to the desired port.

```
# telnet xx-tc 5002
Trying 192.9.200.1 ...
Connected to 129.9.200.1 .
Escape character is '^]'.
```

2. Enter the `telnet` command mode by typing the `telnet` escape character.

```
^]
telnet>
```

3. Send a break to the server.

```
telnet> send brk
```

4. You may now execute the OpenBoot commands.

10.4 Troubleshooting the Terminal Concentrator

Terminal concentrator connections made through a router can exhibit an intermittent problem.

Connections from a host that resides on the same network as the terminal continue to work normally. However, connections to the terminal concentrator via a router display a random interruption. These connections may come alive for random periods, then go dead again. When the connection is dead, new terminal concentrator connection attempts will timeout. The terminal concentrator shows no signs of rebooting.

The diagnosis of this problem is that with routed traffic for many routes on your network you may encounter terminal concentrator routing table overflow and loss of the route to one or more networks. This problem will not occur for connecting hosts on the same network as the terminal concentrator, but rather it will occur for hosts reaching the terminal concentrator through a router. Later routed traffic may re-establish a needed route only to disappear again later.

The solution to this problem is to establish a default route within the terminal concentrator and disable the routed feature. The routed feature must be disabled to prevent the default route from being lost.

The following procedure shows how this can be done. The file config.annex is created in the terminal concentrator's EEPROM file system and defines the default route to be used. The config.annex file can also be used to define rotaries which allow a symbolic name to be used instead of a port number. The routed feature is disabled using the terminal concentrator's set command.

1. **Open an xterm(1) connection to the terminal concentrator.**

```
# telnet xx-tc
Trying 192.9.200.2 ...
Connected to xx-tc.
Escape character is '^]'.


Rotaries Defined:
    cli                              -

Enter Annex port name or number: cli


Annex Command Line Interpreter   *   Copyright 1991 Xylogics, Inc.
```

2. **Enter the su command and administrative password.**
   By default, the password is the IP address of the terminal concentrator.

```
annex: su
Password: administrative_password
```

3. **Edit the** `config.annex` **file.**
   When the terminal concentrator's editor starts, you will see the following message:

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

   Enter the information that is highlighted in the following example, substituting the appropriate IP address for your default router.

```
Ctrl-W: save and exit Ctrl-X: exit Ctrl-F: page down  Ctrl-B: page up
%gateway
net default gateway 192.9.200.2 metric 1 active
```

4. **Disable the local** `routed`**.**

```
annex# admin set annex routed n
    You may need to reset the appropriate port, Annex subsystem or
     reboot the Annex for changes to take effect.
annex#
```

5. **Reboot the terminal concentrator.**

```
annex# boot
```

Error Messages A

This appendix contains the error messages returned when problems arise with Solstice High Availability.

Use the following table to locate specific information in this appendix.

A.1 Overview of Error Messages

Errors that deal with command usage and other simple error messages are not documented in this appendix. Examples and explanations of these messages are:

```
command name: flag unknown flag
```

The above message indicates that an unsupported option (*flag*) was used with the Solstice HA command, *command name*.

```
command name: command line error
```

The above message indicates that a command-line error was detected when the Solstice HA command, *command name*, was invoked.

```
command name: must be root to run this command
```

The above message indicates that the user must be superuser (root) to invoke the Solstice HA command, *command name*.

A.2 Membership Monitor Messages

The following messages are returned by the membership monitor daemon. All of the errors result in the daemon calling the abort programs and a takeover being performed by the sibling.

```
add_hostname: nodeid id is out of range, nodeid
```

The *nodeid* specified in the `cmm_confcdb` file has a value greater than 32. If this message is displayed, contact your service representative.

```
comm_addnode: duplicate nodeid id, nodeid
```

You should contact your service provider if this message is displayed.

```
newipaddr: unknown host hostname
```

The cluster monitor is unable to obtain information about the private network names specified in the `cmm_confcdb` file. A possible problem is that the private network names specified are not in the `/etc/hosts`. Add the private network names to the `/etc/hosts` file. This problem should have been discovered by either `hasetup(1M)` or `hacheck(1M)` during initial configuration.

```
node cannot bind to any host address
```

The cluster monitor is unable to bind to any of the addresses specified in the cmm_confcdb file. These addresses would be the names mapping to the private Ethernets. The problem could be that the interfaces are not configured or that the cable connections are wrong. You should check the cable connections and test the connections using ping(1M). This problem should have been discovered by either hasetup or hacheck during initial configuration.

```
t_bind cannot bind to requested address
```

The port number specified in cmm_confcdb is already in use. To correct the problem, specify a new port number. Both servers should be using the same number. Specify the new port number using the hasetup command.

```
nodetimeout cannot be lower than msgtimeout
```

The default values in the cmm_confcdb file are corrupted. If this message is displayed, contact your service representative.

```
cdbmatch value, fullkey, cdb_sperrno
```

The values in the cmm_confcdb file are corrupted. If this message is displayed, contact your service representative.

```
invalid value for parameter failfast
```

The *fastfailmode* parameter in the `cmm_confcdb` file is wrong. If this message is displayed, contact your service representative.

```
parameter parameter must be an integer
```

A parameter in the `cmm_confcdb` file is corrupted. If this message is displayed, contact your service representative.

```
parameter parameter must be either true or false
```

A parameter in the `cmm_confcdb` file is corrupted. If this message is displayed, contact your service representative.

```
parameter parameter must be an numeric
```

A parameter in the `cmm_confcdb` file is corrupted. If this message is displayed, contact your service representative.

```
parameter parameter not found in config file
```

A parameter is missing from the `cmm_confcdb` file. If this message is displayed, contact your service representative.

```
unsupported version number number (supported number)
```

The version number of the `cmm_confcdb` does not match the version number expected by the cluster monitor. The default file has a value of 2. Install the same Solstice HA packages on both systems.

```
must be superuse to start
```

A user without superuser privileges attempted to invoke the cluster monitor. This only occurs if the user attempted to start the cluster monitor from the command line. Either allow the monitor to start automatically when the system is rebooted or run the following command:

```
# /etc/init.d/SUNWhadf start
```

```
Aborting node with stale seqnum number
```

A server is in an unexpected state.

```
received signal signal
```

The cluster monitor, `clustd`, received either a `SIGTERM`, `SIGPOLL`, `SIGALRM`, or `SIGINT` signal.

```
transition timeout
```

The cluster reconfiguration program was not completed within the specified transition timeout. This error will result in a takeover being performed by the other system in the configuration.

```
unknown scheduling class class
```

An invalid scheduling class was specified in the `cmm_confcdb` file. The only valid choice is RT. To correct the error, restore the template `cmm_confcdb` file and enter a valid class using `hasetup`.

Man Pages B

This appendix contains the manual pages associated with Solstice High Availability. The manual pages included in this appendix are:

- `hacheck(1M)` – Checks and validates Solstice HA configurations
- `hafstab(1M)` – Edits and distributes `dfstab(4)` and `vfstab(4)` files in a Solstice HA configuration
- `haload(1M)` – Monitors the load on the Solstice HA servers
- `haoracle(1M)` – Performs HA-ORACLE administration
- `hasetup(1M)` – Sets up Solstice HA configurations
- `hastat(1M)` – Monitors status of Solstice HA configurations
- `haswitch(1M)` – Performs a switchover of services in a Solstice HA configuration
- `hadfconfig(4)` – Contains the Solstice HA configuration information
- `haoracle_config(4)` – Contains HA-ORACLE fault monitoring information
- `haoracle_databases(4)` – Contains the table of HA-ORACLE databases
- `haoracle_support(4)` – Contains the table of HA-ORACLE releases supported by Solstice HA

Index

CETTE PUBLICATION PEUT CONTENIR DES MENTIONS TECHNIQUES ERRONEES OU DES ERREURS TYPOGRAPHIQUES. DES CHANGEMENTS SONT PERIODIQUEMENT APPORTES AUX INFORMATIONS CONTENUES AUX PRESENTES, CES CHANGEMENTS SERONT INCORPORES AUX NOUVELLES EDITIONS DE LA PUBLICATION. SUN MICROSYSTEMS INC. PEUT REALISER DES AMELIORATIONS ET / OU DES CHANGEMENTS DANS LE(S) PRODUIT(S) ET / OU LE(S) PROGRAMME(S) DECRITS DANS DETTE PUBLICATION A TOUS MOMENTS.

Reader Comments

We welcome your comments and suggestions to help improve this manual. Please let us know what you think about the *SPARCcluster High Availability Server Software Administration Guide,* part number *802-3511-10*.

- The procedures were well documented.

Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
❑	❑	❑	❑	❑

Comments ____

- The tasks were easy to follow.

Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
❑	❑	❑	❑	❑

Comments ____

- The illustrations were clear.

Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
❑	❑	❑	❑	❑

Comments ____

- The information was complete and easy to find.

Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
❑	❑	❑	❑	❑

Comments ____

- Do you have additional comments about the *SPARCcluster High Availability Server Software Administration Guide*?

____

____

____

____

Name: ____

Title: ____

Company: ____

Address: ____

____

Telephone: ____

Email address: ____

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL
FIRST CLASS MAIL PERMIT NO. 1 MOUNTAIN VIEW, CA

POSTAGE WILL BE PAID BY ADDRESSEE

SUN MICROSYSTEMS, INC.
Attn: Manager, Hardware Publications
MS MPK 14-101
2550 Garcia Avenue
Mt. View, CA 94043-9850