Sun StorEdge™ Availability Suite 3.2 Remote Mirror Software Administration and Operation Guide

Sun Microsystems, Inc.
www.sun.com

Please
Recycle

Adobe PostScript™

Contents

Figures

Tables

Preface

The *Sun StorEdge Availability Suite 3.2 Remote Mirror Software Administration and Operations Guide* describes the Sun StorEdge™ remote mirror software and its use. This guide is intended for system administrators who have experience with the Solaris™ operating environment and related disk storage systems.

Note – This software was previously known as the Sun StorEdge Network Data Replicator (SNDR) software.

Before You Read This Book

To fully use the information in this document, you must have thorough knowledge of the topics discussed in these books:

■ *Sun StorEdge Availability Suite 3.2 Software Installation Guide*

■ *Sun StorEdge Availability Suite 3.1 Software Release Notes*

To operate correctly, the remote mirror software must be installed as described in the installation and release documentation.

How This Book Is Organized

Chapter 1 provides general information about the remote mirror software.

Chapter 2 describes the remote mirror synchronization and replication modes.

Chapter 3 describes how to configure and prepare to use the software.

Chapter 4 explains remote mirror software operations including startup, synchronization, and recovery rehearsals.

Chapter 5 describes use of the `sndradm` command.

Appendix A describes the `dsstat` reporting and monitoring command.

Using UNIX Commands

This document might not contain information on basic UNIX® commands and procedures such as shutting down the system, booting the system, and configuring devices. See the following for this information:

- Software documentation that you received with your system
- Solaris™ operating environment documentation, which is at

  `http://docs.sun.com`

Typographic Conventions

Typeface or Symbol[1]	Meaning	Examples
AaBbCc123	The names of commands, files, and directories; on-screen computer output.	Edit your `.login` file. Use `ls -a` to list all files. `% You have mail.`
AaBbCc123	What you type, when contrasted with on-screen computer output.	`%` **su** `Password:`
AaBbCc123	Book titles, new words or terms, words to be emphasized. Command-line variable; replace with a real name or value.	Read Chapter 6 in the *User's Guide*. These are called *class* options. You *must* be superuser to do this. To delete a file, type `rm` *filename*.
[ ]	In syntax, brackets indicate that an argument is optional.	`scmadm` [–d *sec*] [–r *n*[:*n*][,*n*]...] [–z]
{*arg* | *arg*}	In syntax, braces and pipes indicate that one of the arguments must be specified.	`sndradm -R b {p|s}`

1 The settings on your browser might differ from these settings.

Shell Prompts

Shell	Prompt
C shell	*machine-name*%
C shell superuser	*machine-name*#
Bourne shell and Korn shell	$
Bourne shell and Korn shell superuser	#

Related Documentation

Application	Title	Part Number

Man pages	`sndradm`
	`iiadm`
	`dsstat`
	`kstat`
	`svadm`
		N/A

Latest release information	*Sun StorEdge Availability Suite 3.2 Software Release Notes*	817-2782
	Sun Cluster 3.0 and Sun StorEdge Software Release Note Supplement	816-5128
Installation and User	*Sun StorEdge Availability Suite 3.2 Software Installation Guide*	817-2783
	SunATM 3.0 Installation and User's Guide	805-0331
	SunATM 4.0 Installation and User's Guide	805-6552
	Sun Gigabit Ethernet FC-AL/P Combination Adapter Installation Guide	806-2385
	Sun Gigabit Ethernet/S 2.0 Adapter Installation and User's Guide	805-2784
	Sun Gigabit Ethernet/P 2.0 Adapter Installation and User's Guide	805-2785
	Sun Enterprise 10000 InterDomain Networks User Guide	806-4131
System administration	*Sun StorEdge Availability Suite 3.2 Point-in-Time Copy Software Administration and Operation Guide*	817-2781
Cluster usage	*Sun Cluster 3.0 and Sun StorEdge Software Integration Guide*	816-5127
Configuration	*Sun Enterprise 10000 InterDomain Network Configuration Guide*	806-5230

Accessing Sun Documentation

You can view, print, or purchase a broad selection of Sun documentation, including localized versions, at:

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

Contacting Sun Technical Support

If you have technical questions about this product that are not answered in this document, go to:

`http://www.sun.com/service/contacting`

Sun Welcomes Your Comments

Sun is interested in improving its documentation and welcomes your comments and suggestions. You can submit your comments by going to:

`http://www.sun.com/hwdocs/feedback`

Please include the title and part number of your document with your feedback:

Book Title, part number 8xx-xxxx-xx.

Introduction

This chapter describes the following:

Overview

The Sun StorEdge Availability Suite 3.2 remote mirror software is a volume-level replication facility for the Solaris operating environment. You can use this software to replicate disk volumes between physically separate primary and secondary hosts in real time.

As part of a disaster recovery and business continuance plan, the remote mirror software enables you to keep up-to-date copies of critical data at remote sites. You can also rehearse your data recovery strategy to fail over data to remote sites. Later, you can write any data changes that occurred to the original disk. To transport data, the remote mirror software uses any Sun network adapter that supports TCP/IP.

The remote mirror software is active while your applications are accessing the data volumes and continually replicates the data to remote sites. The software operates at the volume level on storage devices attached to one or more hosts.

You can also update the data on the secondary site volume by issuing a command to *synchronize* the primary and secondary site volumes. You can restore data from the secondary volume to the primary volume by issuing a command to *reverse synchronize* the volumes. Reverse synchronization is also known as performing a *reverse update*. Updates from the primary site to the secondary site are also known as *forward resynchronizations*.

Features of the Remote Mirror Software

TABLE 1-1 describes the remote mirror software features.

TABLE 1-1   Remote Mirror Software Features

Feature	Enables You To. . .
Disk-based asynchronous queues	Create asynchronous queues on disk as well as in memory, which enables significantly larger queues.
Active logging	Continue logging operations whenever the remote mirror software is disabled or interrupted.
Disaster rehearsal rollback	Simulate a disaster without committing data changes to volumes.
Groups of remote mirror volume sets	Control the remote mirror software volume sets simultaneously by grouping them. This feature is essential in installations requiring you to maintain consistent contents of a group of volumes.
	An I/O group is a collection of remote mirror software sets that have the same group name, primary and secondary interfaces, and mirroring mode. Mixed groups (those where mirroring modes are asynchronous for one set and synchronous for another set) are not allowed.
	The remote mirror software maintains write ordering for volumes in a group to ensure that the data on the secondary volumes is a consistent copy of the corresponding primary volumes.
Multihop sets	Replicate data from one primary volume to a secondary volume; the secondary volume then replicates the data again to another secondary volume, and so on, in daisy-chain fashion. See "One-to-Many, Many-to-One, and Multihop Volume Sets" on page 14.
Mutual backup	Concurrently transmit data copies to and receive data copies from a remote volume. Also known as a *bilateral relationship*.

TABLE 1-1 Remote Mirror Software Features *(Continued)*

Feature	Enables You To. . .
One-to-many sets	Replicate data from one primary volume to many secondary volumes residing on one or more sites. When you perform a forward resynchronization, you can synchronize one volume set or all volume sets. Issue a separate command for each set. You can also update the primary volume using a specific secondary volume. See "One-to-Many, Many-to-One, and Multihop Volume Sets" on page 14.
Optimized resynchronization	Resynchronize volumes following disk, link, system, and storage platform outages; you only replicate those blocks that were modified since the last synchronization
RAID support	Use RAID volumes as part of your remote mirror software strategy. Volumes can be any RAID level.
Security and Internet protocols	Use version 3.2 securely and efficiently in the Sun Solaris 8 and 9 operating environments, which support Internet Security Protocol (IPsec). The Solaris 8 and 9 operating environments also support Internet Protocol Version 6 (IPv6). (The Solaris 7 operating environment does not support IPv6). Version 3.2 does not require the use of an `.rhosts` file. Place the host names of each host running the remote mirror version 3.2 software in the `/etc/hosts` file on each machine.
Stop and start control of remote replication	Have fine control over replication at the logical volume level. You can include volumes containing critical data by specifying these volumes as part of a remote mirror software volume set. You can also exclude volumes with noncritical data from remote mirror software options.
Sun StorEdge point-in-time copy software integration	Make point-in-time volume copies of your data for additional point-in-time coordination and recovery.
Synchronous and asynchronous volume replication between physically separate sites	Plan for disaster recovery and business continuance using physically separate primary and secondary sites. The remote mirror software design is link-neutral, meaning that it can use any Sun network adapter that supports TCP/IP. Continue remote data access despite local disk failure (depending on how the remote mirror software is configured).

Features of 3.2 Version

The following features are new to the remote mirror software with the Sun StorEdge Availability Suite 3.2 software release.

Disk-Based Asynchronous Queues

In the 3.2 version of the remote mirror software, data can be queued on disk as well as in memory. Memory-based queues are the default.

Disk-based queues enable:

- Larger local queues in asynchronous replication mode
- Larger bursts of I/O activity without affecting application response time

If a disk-based queue fills up, the remote mirror software goes to nonblocking, or scoreboarding, mode.

Blocking Mode

Blocking mode is the default when operating the remote mirror software in asynchronous mode. Blocking mode ensures write ordering of the packets to the secondary site.

If the asynchronous queue fills up when the software is running in blocking mode, the response time to the application can be affected adversely. Write operations must be acknowledged prior to being removed from the queue, so they can prevent, or block, further write operations to the queue until space is available.

Nonblocking Mode

Nonblocking mode is optional in asynchronous operation of the remote mirror software. In nonblocking mode, if the asynchronous queue fills up, the queue is discarded and the remote mirror software goes into logging mode.

In logging mode, the bitmap is used to scoreboard writes. The application's writes are not blocked, but write ordering is lost during scoreboarding. However, the application sees no significant degradation in response time.

Do an update synchronization to synchronize data on the primary and secondary sites after filling of the queue and subsequent entry into logging mode.

Multiple Asynchronous Flusher Threads

The remote mirror software now has the ability to use multiple flusher threads to increase the drain rate from the asynchronous queues. This allows multiple I/Os per consistency group or set on the network at one time. The default number of queue-flushing threads is two.

If you want an operation that is similar to that of the Sun StorEdge Availability Suite 3.1 remote mirror software, set the flusher threads to one. You can increase the number of threads above two for enhanced performance in a low-latency network environment.

When using multiple threads, writes can often arrive at the secondary site out of sequence. To prevent any problems at the secondary site, sequence numbers are added to all data writes at the primary site. The secondary site manages the incoming data based on the sequence numbers. Write ordering is essentially restored at the secondary site. Writes that arrive out of order are stored in memory until previous writes arrive.

The use of multiple asynchronous flusher threads on the primary site requires more memory at the secondary site. Each group or ungrouped set that the secondary site is tracking can have a maximum of 64 requests pending and in memory at the secondary site. Memory requirements depend on the number of groups or sets tracked, the maximum request count of 64, and the size of the writes.

When the number of requests for a group or set reaches 64, the secondary site prevents the primary site from issuing any more requests for that group or set.

If memory is not available when a packet arrives, the packet is rejected and all groups and sets go into logging mode at the secondary site.

See "Memory Requirements" on page 11 for further information.

Protocol Changes

The protocol of the remote mirror software has been improved to take advantage of the software's improved asynchronous flushing rate and its improved usage of network bandwidth.

The software protocol has been enhanced to work efficiently with the new disk-based asynchronous queues and the associated multiple flusher threads.

Write Coalescing

If possible, the remote mirror software combines, or coalesces, multiple sequential writes to the primary volume into a single network write. The size of the writes from the application and the network packet size affect the remote mirror software's ability to coalesce writes. Write coalescing provides two important advantages:

- The asynchronous queue's drain rate improves.
- Network bandwidth utilization improves.

Remote Mirror Software Architecture

The core remote mirror software code is a kernel module that interfaces to the network storage control module (nsctl) framework. The software is configurable on any device that is accessible through the nsctl framework. The sndradm CLI is the external user interface used to manage the remote mirror software.

FIGURE 1-1 shows the relationship between the storage volume (SV)driver, the remote mirror software, and the rest of the nsctl framework. I/O commands and data enter and exit the remote mirror software through the Sun StorEdge storage volume (SV) driver software. Mediated by nsctl, the data flows through the remote mirror software (and, optionally the point-in-time copy software) and the storage device block cache (SDBC) drivers to its destination on the storage array or in user space.

The SV driver intercepts the I/O commands to the remote mirror volumes and routes them through the Sun StorEdge I/O stack to the storage device driver or the volume manager. The SV driver is a very thin layer in the I/O stack and operates by interposing commands onto the entry points for the underlying device driver.

The I/O commands originating in user space are intercepted at the top of the Sun StorEdge I/O stack. The SV driver routes them through the stack and feeds them back to the storage device driver or volume manager at the bottom of the stack. Data also flows in the opposite direction, from the storage back to user space. Because the remote mirror software is located at the top of the stack before the point-in-time copy software, the remote mirror volume sets must be placed in logging mode (sndradm -l) before performing point-in-time enable, copy, update, or reset operations on any volume that is also part of a remote mirror volume set.

If the volume set is not in logging mode, the point-in-time copy operation fails and the remote mirror software reports that the operation is denied. See also "Using the Remote Mirror Software With the Point-in-Time Copy Software" on page 29.
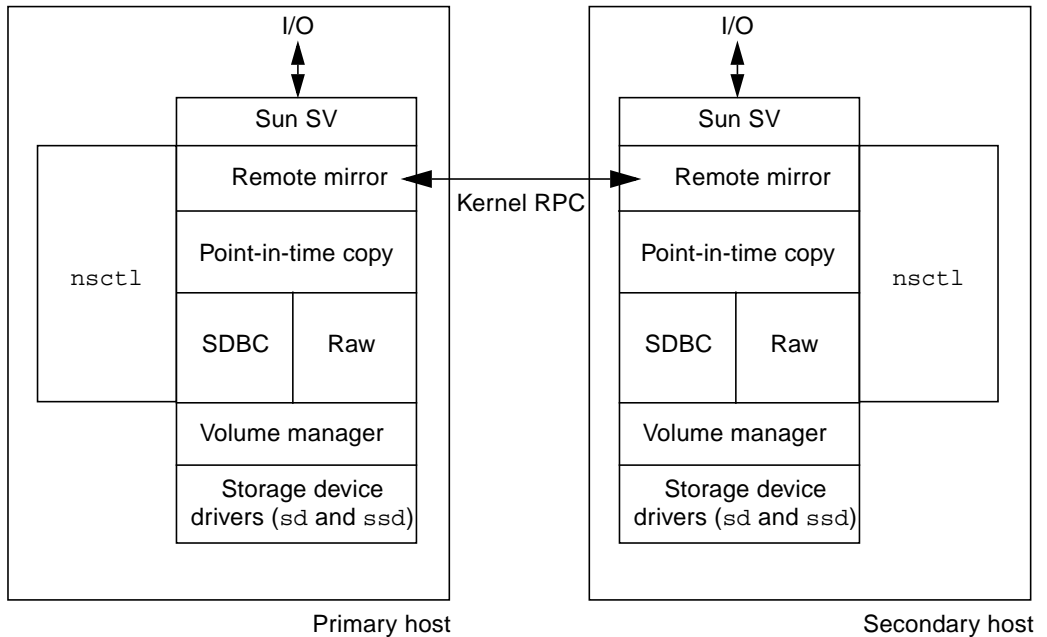
FIGURE 1-1   Remote Mirror Software Architecture

File System Replication

The remote mirror software is not a file or file system replicator. It is a volume replicator. When you replicate a volume on the primary site (Site-A) to a volume on the secondary site (Site-B), Site-B receives an exact bit-for-bit copy. Make sure that any Site-A file systems you want to replicate are mounted.

If a file system has data cached but not committed to disk, and has returned successfully, to the calling application just before a system failure occurs, data can be lost. To prevent this from happening, mount your file system with the `forcedirectio` option. Using this option has a significant impact on overall performance, so test your system to be sure that its use is necessary.

When replicating, the primary host file system is mounted. Do not mount the file system on Site-B until you are ready to fail over or write to that site's volume. Changes appear on a replicated file system volume only after mounting the volume.

A file system on Site-B can be mounted only in read-only mode while the volume set continues to replicate. Once Site-B volumes are placed into logging mode, the file system can be mounted for read and write operations.

Volume Sets

Note – In this document, "volume" refers to a raw disk partition or a volume created by a volume manager.

The remote mirror software replicates data to volume sets that you define. A volume set consists of a primary volume residing on a local (primary) host and another volume residing on a remote (secondary) host. The volume set also includes a bitmap volume on each host that tracks write operations and differences between the volumes. See "Customizing Volume Sets" on page 42.

The secondary volumes can be updated synchronously in real time or asynchronously using a store-and-forward technique. Typically, a primary volume is first explicitly copied to a designated secondary volume to establish matching contents. As applications write to the primary volume, the remote mirror software replicates changes to the secondary volume, keeping the two images consistent.

In synchronous mode, a write operation is not confirmed as complete until the remote volume has been updated. In asynchronous mode, a write operation is confirmed as complete before the remote volume has been updated.

The size of the secondary volume must be equal to or greater than the corresponding primary volume. If you initiate a resynchronization on a volume set where the secondary volume is smaller than the primary, the software fails with an error. Refer to "Setting Up Bitmap Volumes" on page 41 for more information about volume sizes.

What Kind of Volumes Can I Use?

Note – See "Volume Manager Software" on page 12.

Volumes are defined here as logical volumes that can be linear, striped, or RAID volumes. You can create logical volumes by using the Solstice DiskSuite™, Solaris Volume Manager (available in the Solaris 9 OE), or VERITAS Volume Manager software.

You can use Redundant Array of Independant Disks (RAID) volumes as part of your remote mirror software strategy. Volumes can be any RAID level. The RAID levels of volumes in a volume set do not have to match.

Caution – You cannot use the remote mirror software to replicate the boot device or any system-level partitions like `/var` and `/usr`.

Raw Partitions and Volumes

When selecting a volume to be used in a volume set (including the configuration location), ensure that the volume does not contain disk label private regions (for example, slice 2 on a Solaris OE-formatted volume). The disk label region is contained in the first sectors of a disk. To be safe, ensure that cylinder 0 is not part of any logical volume that is replicated.

Caution – When creating volume sets, do not create secondary or bitmap volumes using partitions that include cylinder 0, data loss might occur. See "VTOC Information" on page 17.

How Many Volumes Can I Configure?

By default, the remote mirror and point-in-time copy software support a configuration of 4096 volumes and 64 Mbyte for caching. You can increase both amounts if system resources allow. The number of volumes allowed is divided between both software products. For example, if you use the point-in-time software only, you can have 341 volume sets, each consisting of master, shadow, and bitmap volumes.

For more information, see "Increasing the Default Number of Volumes Allowed" on page 38.

> **Caution –** In a clustered environment, only one system administrator or root user at a time is allowed to create and configure Sun StorEdge volume sets. This restriction helps avoid creating an inconsistent Sun StorEdge Availability Suite volume set configuration.

Two administrators should not be writing to the Sun StorEdge Availability Suite software configuration at the same time. The operations that access the configuration include but are not limited to:

- Creating and deleting volume sets
- Adding and removing volume sets from I/O groups
- Assigning new bitmap volumes to a volume set
- Updating the disk device group or resource name
- Any operation that changes the Sun StorEdge Availability Suite software and related volume set configuration

I/O Groups

The software enables you to group volume sets in *I/O groups*. You can assign specific volume sets to an I/O group to perform replication on those volume sets and not on others you have configured. Grouping volume sets also guarantees write ordering: Write operations to the secondary volume occur in the same order as the write operations to the primary volume.

An I/O group is a collection of software sets that have the same group name, primary and secondary interfaces, and mirroring mode. Mixed groups (those where mirroring modes are asynchronous for one set and synchronous for another set) are not allowed.

Using an I/O group, you can issue a remote mirror software command that is executed on every member of the group. Volume sets can be controlled as a single unit.

I/O group operations are atomic. The change from replicating mode to logging mode is guaranteed to occur on every set in an I/O group and to fail on every set if it fails on a single set in the group.

The software maintains write ordering for volumes in a group to ensure that the data on each secondary volume is a consistent copy of the corresponding primary volume. See "Order-Dependent Writes and Volume Set Grouping" on page 15.

> **Note –** The I/O group concept does not matter for synchronous replication. Write-ordering is preserved among those volume sets configured as `sync`.

The autoresynchronization feature supports I/O grouping. It allows the feature to be enabled or disabled on a per-group basis and controls the resynchronization operation atomically on the group.

I/O grouping has an adverse affect on the remote mirror software asynchronous operation, as I/O flushing is reduced to a single thread. In this case, consider the size of the data to be transferred because all I/O is routed through a single queue.

Groups in a Clustered Environment

You can also group volume sets according to their cluster or resource tag to perform replication in a clustered environment. The remote mirror software is cluster aware in the Sun Cluster 3.0 Update 3 and Sun Cluster 3.1 environments, providing high availability (HA) for the Sun StorEdge software.

See the *Sun Cluster 3.0 and Sun StorEdge Software Integration Guide* for more information about configuring the Sun StorEdge Availability Suite software in a Sun Cluster environment.

Memory Requirements

This section discusses the remote mirror software and the possible memory requirement it places on the secondary host when using multiple asynchronous flusher threads.

With releases prior to the Sun StorEdge Availability Suite 3.2 software, a single asynchronous thread was created for each group of sets on the primary host. Asynchronous I/O requests were placed on an in-memory queue and serviced by this single thread. Because there was only one thread, only one RPC request could be in-flight for each group of sets and each request had to be completed before a new request could be issued.

The version 3.2 software enables the number of asynchronous service threads to be set on a per-group basis, which allows multiple in-flight RPC requests and speeds servicing of the asynchronous queue. Enabling more than one RPC request creates the possibility that these requests can arrive out of order with respect to the order in which the writes were issued on the primary host. In other words, a request might arrive before a previous request has completed its I/O.

The write order must be maintained within a group. Therefore, out of order requests must be stored in memory on the secondary host until the missing request comes in and completes execution.

A maximum of 64 outstanding requests per group are stored on the secondary host, after which it stalls the primary host from issuing any more requests. This hard limit is applied only to the number of possible outstanding requests, not to the size of their payload. For example, if the I/O consists of 4Kbyte writes with 6 groups, the total memory requirement could be 4Kbyte x 6 x 64 = 1536 Kbyte. However, with an I/O size of 1 Mbyte, this could rise to 1Mbyte x 6 x 64 = 384 Mbyte.

Volume Manager Software

The remote mirror and point-in-time copy software does not support metatrans devices (also known as trans metadevices) created by the Solstice DiskSuite or Solaris Volume Manager software.

Metatrans devices are intended for use with UNIX file systems (`ufs`) without using any other layered services. Use the `ufs` logging mount option as an alternative to the use of metatrans devices. The Sun StorEdge Availability Suite software supports `ufs` logging, which should be used when available instead of metratrans devices.

Network Protocols and TCP/IP Connection

The remote mirror software requires a TCP/IP connection between the primary and secondary hosts. A dedicated TCP/IP link is not required.

Although the remote mirror software is most likely to be used with SunATM™ link-level interfaces, the remote mirror software design is link-neutral, meaning that it can use any Sun network adapter that supports the TCP/IP protocol.

Each host must have the proper Asynchronous Transfer Method (ATM) or Ethernet hardware installed to support the TCP/IP link. The remote mirror software operates over any TCP/IP networking technology but has been qualified only on 10, 100, and 1000 Mbit Ethernet and SunATM155 and SunATM622 technologies.

When using ATM, ensure that the configuration supports TCP/IP by using either Classical IP or LAN Emulation mode. For more information on configuring the SunATM interface for these protocols, refer to the *SunATM Installation and User's Guide.*

Use of TCP/IP Ports

The remote mirror software on both the primary and secondary nodes listens on a well-known port advertised in `/etc/services`. The default is port 121. Remote mirror write traffic flows from primary to secondary host over a socket with an arbitrarily assigned address on the primary host and the well-known address on the secondary host. The health-monitoring heartbeat travels over a different connection, with an arbitrarily assigned address on the secondary host and the well-known address on the primary host. The remote mirror protocol utilizes SUN RPCs over these connections.
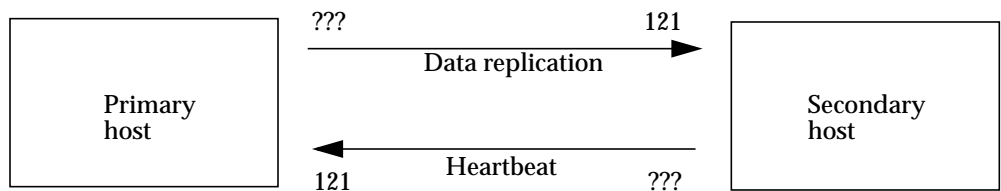
FIGURE 1-2   Use of TCP/IP Port Addresses by the Remote Mirror Software

Default Listening Port

Port 121 is the default TCP/IP port for use by the remote mirror `sndrd` daemon. To change the port number, edit the `/etc/services` file using a text editor. See the *Sun StorEdge Availability Suite 3.2 Software Installation Guide* for more information.

Note – If you change the port number, you must change it on all remote mirror hosts within a configuration set (that is, primary and secondary hosts, or all hosts in one-to-many, many-to-one, and multihop configurations). In addition, you must restart all affected hosts, so that the port number change can take effect.

Firewalls and Remote Mirror Software

Because RPCs require an acknowledgment, the firewall must be opened to allow the well-known port address to be in either the source or destination fields of the packet.

In the case of write replication traffic, packets destined for the secondary host have the well-known port number in the destination field. Acknowledgments of these RPCs contain the well-known address in the source field.

For health monitoring, the heartbeat originates from the secondary host with the well-known address in the destination field. The acknowledgment contains this address in the source field.

If the option is available, be sure to configure the firewall to allow RPC traffic as well.

One-to-Many, Many-to-One, and Multihop Volume Sets

The remote mirror software enables you to create one-to-many, many-to-one, and multihop volume sets.

One-to-many replication enables you to replicate data from one primary volume to many secondary volumes residing on one or more hosts. The primaryvolume plus each secondary volume forms a single volume set. For example, with one primary and three secondary host volumes, you need to configure three volume sets: primary A and secondary B1, primary A and secondary B2, and primary A and secondary B3.

Note – Consistency groups across interfaces are not supported in a one-to-many configuration because write ordering cannot be guaranteed.

Many-to-one replication enables you to replicate volumes across more than two hosts through more than one network connection. The software supports the replication of volumes located on many different hosts to volumes on a single host. The terminology differs from the one-to-many configuration terminology, where the *one* and the many referred to are *volumes*.

Multihop replication indicates that the secondary host volume of one volume set acts as the primary host volume of another volume set (it is still the secondary volume of the first volume set). In the case of one primary host volume A and one secondary host volume B, the secondary host volume B appears as primary host volume A1 to the secondary host volume B1.

See , and for more information about these scenarios.

Order-Dependent Writes and Volume Set Grouping

Write ordering is maintained for groups of asynchronously replicating volume sets. (The general definition of write ordering is that write operations directed to the target volume occur in the same order as write operations to the source.) The group of target volumes is a copy of the group of source volumes.

This feature can be valuable in cases where you can avoid application requirements that limit operations. For example, a database application might limit partition sizes to no greater than 2 Gbyte. In this case, you might group volume sets to create a large virtual "volume" that preserves write operations. Otherwise, you risk having inconsistent data by trying to update volume sets individually instead of as a group.

When an application has multiple logical volumes assigned, application data integrity can be maintained by one of the following techniques:

- Specifying that all remote mirror software volumes associated with that application are in `sync` mode
- Using the Sun StorEdge Availability Suite 3.2 point-in-time copy software to make periodic, recoverable point-in-time copies

If you use the point-in-time copy software, the remote point-in-time snapshot is taken while the application is in the recoverable state. For example, most database applications allow for a hot backup. If a remote point-in-time copy was made of the entire replicated database while the primary hosr was in hot backup mode, a consistent remote database becomes available by using the point-in-time copy and the log files taken while the database was in hot backup mode.

Note – You must place the remote mirror volume set in logging mode for the point-in-time copy software to successfully perform an enable, copy, update, or reset operation on a remote mirror volume. If the volume set is not in logging mode, the point-in-time copy operation fails and the remote mirror software reports that the operation is denied.

Mutual Backup

Typically, volumes are replicated from local Site-A to remote Site-B. However, as applications are geographically distributed, a storage system at remote Site-B might be both a remote volume backup to local Site-A and also a direct storage resource for applications on Host-B. Under these circumstances, you can replicate Host-B volumes to Site-A.

This reciprocal backup arrangement supported by the remote mirror software is known as *mutual backup* or *dual backup*.

With mutual backup, the remote mirror software volumes considered primary by Site-B are administered from the Site-B session. Site-B replicated-volume devices are considered secondary volumes at Site-A.

In this case, you need to configure two *unique* volume sets for each site. For example:

- Volume set 1 contains primary Site-A volumes replicated to secondary Site-B
- Volume set 2 contains primary Site-B volumes replicated to secondary Site-A

See FIGURE 1-3.

Site-A primary host

| Site-A
Application1
Primary volume A | → Replicated to Site-B → | **Site-B secondary host**

Site-B
Secondary volume B |

Site-A secondary host

| Site-A
Secondary volume A1 | ← Replicated to Site-A ← | **Site-B primary host**

Site-B
Application2
Primary volume B1 |
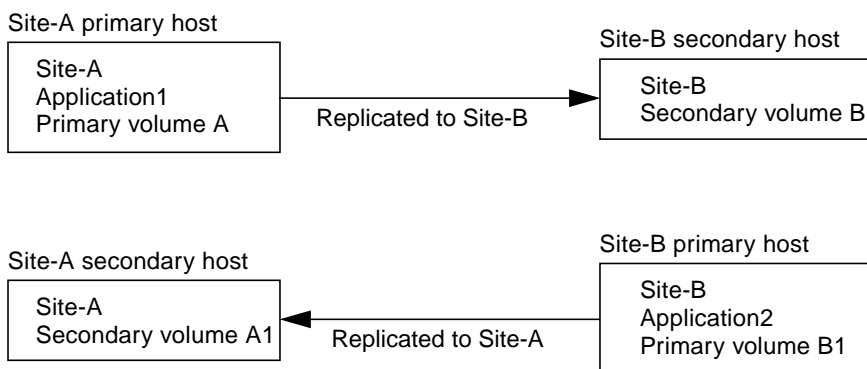
FIGURE 1-3   Volume Sets for Mutual Backup

VTOC Information

Note – This information is also in the *Sun StorEdge Availability Suite 3.2 Troubleshooting Guide.*

The Solaris system administrator must be knowledgeable about the virtual table of contents (VTOC) that is created on raw devices by the Solaris operating system.

The creation and updating of a physical disk's VTOC is a standard function of the Solaris operating system. Software applications like Availability Suite, the growth of storage virtualization, and the appearance of SAN-based controllers have made it easy for an uninformed Solaris system administrator to inadvertently allow a VTOC to become altered. Altering the VTOC increases the possibility of data loss.

Remember these points about the VTOC:

- A VTOC is a software generated virtual table of contents based on the geometry of a device and written to the first cylinder of that device by the Solaris `format`(1M) utility.

- Various software components such as `dd`(1M), backup utilities, point-in-time copy software, and remote mirror software can copy the VTOC of one volume to another volume if that volume includes cylinder 0 in its mapping.

- If the VTOC of the source and destination volumes are not 100 percent identical. Some type of data loss can occur.

This data loss might not be initially detectable, but can be detected later when other utilities are used, like fsck(1M).

When first configuring and validating volume replication, save copies of all affected devices' VTOCs using the prtvtoc(1M) utility. The fmthard(1M) utility can be used to restore them later, if necessary.

- When using volume managers like VxVM and SVM, copying between individual volumes created under these volume managers is safe. VTOC issues are avoided because the VTOC is excluded from volumes created by these volume managers.

- When formatting individual partitions on a raw device, for all partitions except the backup partition, make sure the partitions do not map cylinder 0, which contains the VTOC. When using raw partitions as volumes, you are the volume manager and you need to exclude the VTOC from partitions that you configure.

- When formatting the backup partition of a raw device, make sure that the physical geometries of the source and destination devices are identical. (Partition 2, by default, maps all cylinders under the backup partition.) If identical device sizing is not possible, make sure that the source backup partition is smaller than the destination partition, and that the destination partition does not map cylinder 0.

Replication and Synchronization Modes

This chapter describes the following:

Replication Modes

The remote mirror software supports two modes of data replication:

■ Synchronous replication

■ Asynchronous replication

The replication mode is a user-selectable parameter for each remote mirror volume set. The volumes can be updated synchronously in real time or asynchronously using a store-and-forward technique. Typically, a primary volume is first copied explicitly to a designated secondary volume to establish matching contents. As applications write to the primary volume, the software replicates changes to the secondary volume, keeping the two volumes consistent.

In the event of planned or unplanned outages, the software maintains per-device bitmap volumes that are marked to indicate changed blocks with a granularity of 32 Kbytes per segment. This technique optimizes resynchronization by allowing the software to resynchronize only the blocks that have changed since the last synchronization.

Use the `sndradm` **enable** command and select the volume set's `sync` or `async` parameter to select a replication mode. Use the `sndradm -R m` command to change the replication mode thereafter.

Synchronous Replication Mode

In synchronous replication mode, a write operation is not confirmed as complete until the remote volume has been updated. Synchronous mirroring forces the remote mirror software to wait until the primary volume receives an acknowledgment of the receipt of the data from the secondary volume before returning to the application. The application is not acknowledged until the write operation at the secondary site is complete.

The advantage of synchronous replication is that the primary volume and the secondary volume are in sync after an acknowledgment of a write by the secondary site. One disadvantage might be an increase in write response time, especially for large data sets or long-distance replication (where write operations can incur additional latency because of the time required to transfer data and return acknowledgments).

Asynchronous Replication Mode

In asynchronous replication mode, a write operation is confirmed as complete before the remote volume has been updated. Asynchronous replication enables the remote mirror software to return to the host as soon as the write operation has been completed on the primary volume and has been placed on a per-volume queue for the secondary site. The secondary site receives the queued requests in the order that they were queued. Once the I/O has been completed at the secondary site and the bitmap has been updated to reflect the volume state, notification is sent to the primary site.

The advantages of asynchronous replication are that it provides fast response and has the least impact on the response time of the primary application. A disadvantage is that data loss might occur at the secondary site after primary site or network failures.

For more information about asynchronous replication and adjusting the asynchronous queue, see "Setting the Asynchronous Queue" on page 104 and "Tuning the Asynchronous Queue" on page 105.

Synchronization Modes

The software synchronizes data in a *forward* (from the primary to the secondary) or *reverse* (from the secondary to the primary) direction. The software synchronizes data in four modes:

- Full synchronization
- Update resynchronization
- Reverse full synchronization
- Reverse update synchronization

Using one of the remote mirror synchronization modes ensures that the primary and secondary volumes contain the same data and that they are identical at a clearly defined time. Synchronization is driven by the software through the sndradm command and progresses to completion.

When a volume set is enabled using the sndradm –e command, you must synchronize the primary and secondary volumes in the set initially. (Use the sndradm –E command if the volumes are already identical.)

Once a volume set has been synchronized, the software ensures that the primary and secondary volumes contain the same data through *replication*. Replication is driven by user-layer application write operations. Remote mirror replication is an ongoing process.

Full Synchronization

Full synchronization starts a full copy operation from the primary volume to the secondary volume. It also enables replication concurrently from the primary volume to the secondary volume so that any new write operations to the primary volume are also replicated to the secondary volume. After the operation is complete, the remote mirror software maintains the normal replication mode for the volume: either synchronous or asynchronous replication.

Note – Volumes can be made identical using methods other than a full synchronization. When network latencies justify it, you can perform the initial synchronization by backing up a source or primary volume on magnetic tape on one site and then restoring the volume from the tape on the other site. During the period between when the backup is completed and the restore is started, place the source or primary volume in logging mode. Make sure that the backup copy is a physical copy (for example, one made using the dd(1M) command) and not a logical copy (for example, one made using the tar(1M) or cpio(1M) commands). The copies must have identical blocks, not just identical files.

FIGURE 2-1 shows the full forward synchronization process.

1. The remote mirror software on the primary system (host1) requests disk blocks from the active primary volume. The data might already be resident in the primary system data cache or might require a local disk access.

2. The software transmits the disk blocks, with destaging instructions, over the connection to a cache region on the secondary system.

3. The software on the secondary system updates its remote volume and acknowledges the update to the primary system.
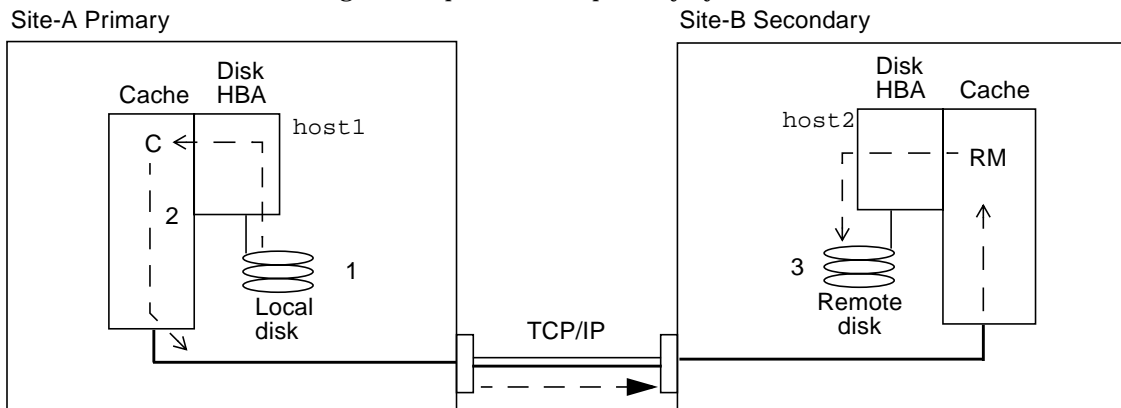
FIGURE 2-1   Full Synchronization (Volume-to-Volume Copy)

Update Resynchronization

An update resynchronization initiates replication of only the changed primary site volume data to the secondary site, based on the bitmap. Only the blocks marked "dirty" (changed) in the bitmap are copied to the target volume. After the update is complete, the software maintains the normal replicating mode. The software can also be placed in logging mode. See "Logging" on page 27.

Logging and update resynchronization serve as a safety net if one of your replication processes is interrupted. The software monitors the network connections between the primary and secondary hosts. Any link failures or remote system failures are detected by the transport interface and are passed to the remote mirror software.

The software resynchronizes the secondary volume from the primary volume. It uses the changes reported in the log files during the interruption to update the secondary volume. It also enables concurrent replication between the primary and secondary volumes so that any new write operations to the primary volumes are also replicated on the secondary volumes.

If the interruption lasts many hours and the updates are widespread, the advantages of logging and performing the update resynchronization diminish. As time passes, the proportion of bits set to true in the bitmap volumes of a volume set might reach 100 percent. The overhead of logging and update resynchronization must then be balanced against that of a full synchronization.

FIGURE 2-2 shows an update resynchronization from the primary system to its secondary system when the secondary volumes are stale from the interruption.

1. The remote mirror software on host1 examines a bitmap from the primary and secondary hosts for the volumes.

2. The software on host1 requests the blocks that were updated during the interruption from the up-to-date volume. The data might already reside in the host1 data cache or on the local disk.

3. The software on host1 transmits the update blocks 3R to the host2 remote mirror software.

4. The software on host2 refreshes its stale image with the updated blocks and acknowledges the action to host1.

5. The software revises the bitmap to track the remote update.

All steps repeat until the remote replicated image is updated.
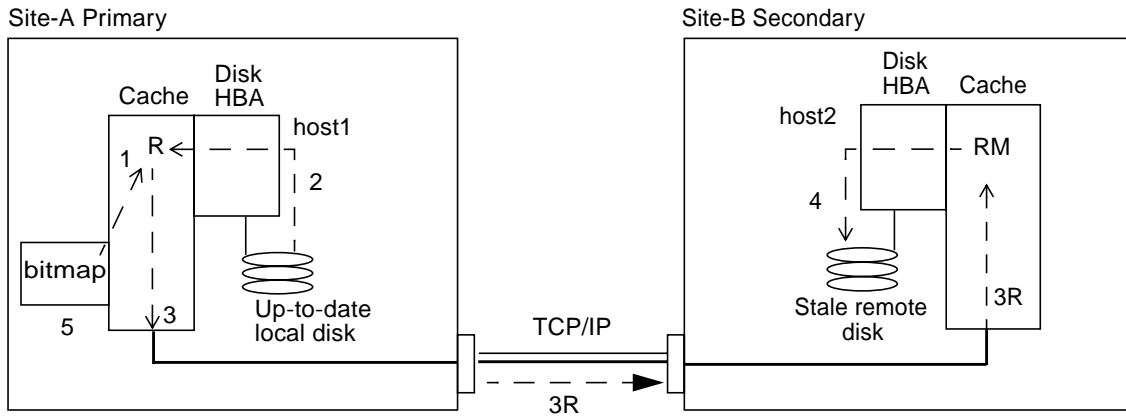
FIGURE 2-2   Update Resynchronization of a Secondary Volume Set

Reverse Full Synchronization

During reverse full synchronization, the remote mirror software replicates the volume data from the secondary site to the primary site. After you issue the `sndradm -m -r` command, the software starts a full reverse copy operation from the secondary volume to the primary volume. It also enables concurrent replication from the primary volume to the secondary volume so that any new write operations to the primary volume are also replicated to the secondary volume.

You can use the primary volume during reverse synchronization. The primary volume shows a consistent volume image of the latest data as soon as the reverse synchronization starts. If your application had been writing to the secondary volume as part of a failure or disaster rehearsal, move the application to the primary volume when the reverse synchronization starts.

FIGURE 2-3 shows the full reverse synchronization process.

1. The data might already be resident in `host2` data cache, or it might require a secondary disk access. If so, the remote mirror software on `host1` requests blocks from the up-to-date secondary volume on `host2`.

2. The software on `host2` transmits the cache blocks 2R over the link to a software region on `host1` with destaging instructions.

3. The software on `host1` updates its disk.
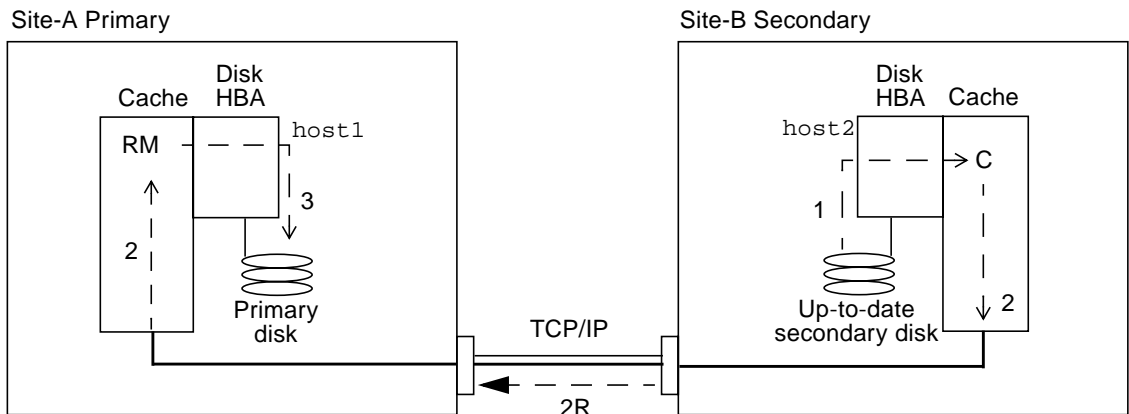
FIGURE 2-3   Reverse Full Synchronization

Reverse Update Synchronization

During reverse update synchronization, the remote mirror software compares the bitmaps between the primary and secondary sites and replicates only the changed blocks from the secondary site to the primary site.

The software resynchronizes the primary volume from the secondary volume. It uses the changes reported in the log files while replication was interrupted to update the primary volume. It also enables concurrent replication between the primary volume and secondary volumes so that any new write operations to the primary volume are also replicated on the secondary volumes.

You can use the primary volume during reverse update synchronization. The primary volume shows a consistent volume image of the latest data as soon as the reverse update synchronization starts. If your application had been writing to the secondary volume as part of a failure or disaster rehearsal, move the application to the primary volume when the reverse update synchronization starts.

FIGURE 2-4 shows a reverse update resynchronization from the secondary system to its primary system.

1. The remote mirror software on `host1` retrieves the secondary bitmap 1R from `host2` for one of the volumes affected by the interruption.

2. The software on `host1` requests the blocks updated during the interruption from the up-to-date secondary volume of `host2`. The data might already be resident in `host2`'s data cache or it might require secondary disk access.

3. The software on `host2` transmits the updated blocks 3R to `host1` remote mirror software region of cache using the intersite link.

4. The software on `host1` refreshes its stale image with the updated blocks.

5. The software on `host1` revises the bitmap to track the remote update.

All steps repeat until the primary volume is up-to-date.

Site-A Primary
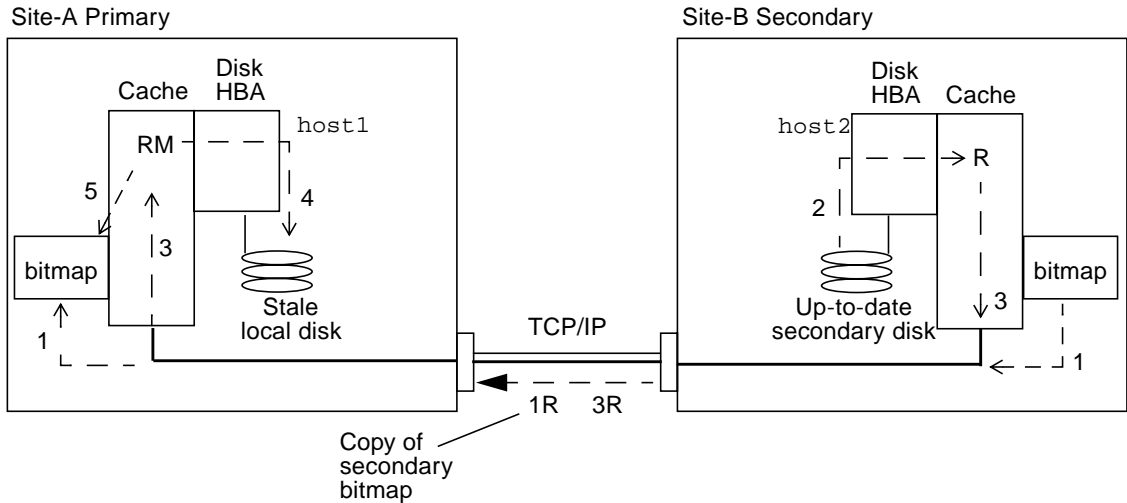
Site-B Secondary

FIGURE 2-4   Reverse Update Synchronization

Logging

During logging, the remote mirror software updates only the bitmaps at the primary site. No replication occurs. At a later time, the bitmaps at the primary and secondary sites are compared and the changed blocks in the primary site volume are mirrored by update resynchronization to the secondary site. You can use logging to save on telecommunications or connection costs. However, you risk the cost of data loss. If you lose the primary site, you do not have the data at the secondary site that was written to the primary site.

If all volume sets in an I/O group are replicating (that is, each secondary volume contains a valid point–in–time copy of the corresponding primary volume) and one volume set enters logging mode, all other sets in the I/O group enter logging mode automatically. This scheme ensures that each secondary volume contains a valid copy of the data.

You can also perform logging on the secondary site before a failover. You can then update the primary site using a reverse synchronization or reverse update synchronization command.

With synchronous and asynchronous replication, the remote mirror software automatically switches to logging mode at the primary site if a break in the network occurs or if the primary site is down. The secondary site does not automatically switch to logging mode if a break in the network occurs or if the primary site is down. Instead, it changes to a state called "need sync" to identify the condition

when you issue the `sndradm -P` command. This protects the data at the secondary site from unwanted write operations. An administrator must cause the software to actually fail over to the secondary site by issuing the `sndradm -l` command on the secondary site.

To resume remote mirror software operations after using the `sndradm -l` logging command, use the `sndradm -m` command to perform a full resynchronization or the `sndradm -u` command to perform an update resynchronization. When issued from the secondary host, the `sndradm -l` command does not work on the secondary volume for any volume that is currently synchronizing.

Choosing Update or Full Resynchronization

An update resynchronization updates the secondary site with changes that occurred at the primary site while replication was suspended. The primary site can also be updated from the secondary, if desired.

A full synchronization is a complete disk-to-disk copy. This operation is the most time-consuming of the synchronization operations. A full synchronization is typically performed only when the remote mirror software volume set is:

- Initially established
- Damaged due to a disaster
- Has questionable data integrity

The volume set data might be in question if, for example, a double disk failure on a RAID 5 set occurred or if the remote mirror software was shut down manually and write operations occurred to either the primary or secondary volumes when logging was not active. If the integrity of the volume data is questioned, the only way to get the volume to a synchronized set is to perform a full volume copy. The full copy can be performed from the primary site to the secondary site or, if appropriate, from the secondary to the primary.

Tip – If a synchronization initiated by the `sndradm -m` or `sndradm -m -r` commands is interrupted, use the `sndradm -u` update command to complete the synchronization.

Choosing Automatic or Manual Resynchronization

The remote mirror software provides two methods for synchronization after a scheduled or unscheduled link failure:

- Automatic (also known as autosynchronization): Synchronization occurs when the link is reestablished. Set this feature for each volume set using the `sndradm –a` command.
- Manual: Synchronization results from an update or full resynchronization. Issue one of the `sndradm` commands.

Caution – Autosynchronization is discouraged if the interruption might be the warning of a larger, rolling disaster. Keep the secondary site in a dated but consistent state rather than risk a disastrous interruption that leaves the secondary site inconsistent and difficult to recover. For this reason, autosynchronization is disabled by default.

See also "Using the Remote Mirror Software With the Point-in-Time Copy Software" on page 29. Before you start a resynchronization operation, ensure that you have an appropriate point-in-time copy of the target volume.

In a Sun Cluster environment, consider the following when using autosynchronization:

- If you want automatic resynchronization to occur in the event of a cluster failover, turn on the autosync feature. With this feature enabled, any cluster failover puts the remote mirror volume sets back into replication mode after an update occurs.
- If you want to force clusters to fail over, be sure to allow all remote mirror components to fully recognize the condition, including the remote host that is not in the cluster. Wait at least 30 seconds after the completion of the `scswitch` command before starting an update sync. This gives time for Sun Cluster to complete its logical host interface reconfiguration.

Using the Remote Mirror Software With the Point-in-Time Copy Software

To ensure a high level of data integrity on both sites during normal operations or data recovery, use the remote mirror software with the point-in-time copy software. Use the point-in-time copy software just before you perform a resynchronization to ensure that a consistent copy of the data exists. If a failure occurs, you have the point-in-time data copy to restore your data.

Note – Place the related remote mirror volume set in logging mode to perform an enable, copy, update, or reset operation on a remote mirror volume. If the volume set is not in logging mode, the point-in-time copy operation fails and the remote mirror software reports that the operation is denied.

During the resynchronization process of updating the local and remote sites, the data on a secondary remote mirror volume is temporarily inconsistent with the primary volume. The secondary volume cannot be relied on for data recovery. Consistency is restored when the resynchronization is complete. To ensure data integrity, use point-in-time copy software regularly to create a point-in-time copy of data at both sites. See the point-in-time copy documentation listed in "Related Documentation" on page xvi.

Volume Sizing

The software returns the following error message in several circumstances when the shadow volume of a point-in-time copy volume set is not the same size as the master volume.

```
Another package would not allow target to be changed at this moment
```

Whenever a point-in-time copy snapshot is taken, the volume that is used to create the snapshot, the shadow volume, is made to look exactly like the master volume, including matching the number of blocks. If the master volume is larger or smaller than the shadow volume's physical size, the shadow volume appears to be resized at the moment the snapshot is taken. Physically, the shadow volume has not changed size, but the point-in-time copy kernel module always reports its size to be the same size as the master volume. This can present several problems with the remote mirror software, which does not expect the size of the volume to change:

- If you plan to use a remote mirror primary or secondary volume as a point-in-time copy shadow volume, the master and the shadow of that point-in-time copy volume set must be exactly the same size. If they are different, you get an error when attempting to enable the point-in-time copy volume set with the `iiadm -e` command.

- You might also see this error during a disable `iiadm -d` command. If the shadow volume of the existing point-in-time copy volume set is not the same size as the master volume and a remote mirror volume set was enabled using that shadow volume as the primary or secondary, the point-in-time copy software stops you from performing the disable operation.

Autosynchronization

The `/usr/opt/SUNWrdc/lib/sndrsyncd` daemon automates update resynchronization after a network link or machine failure. If the point-in-time copy software is also installed and you have added point-in-time copy volume groups, the daemon invokes point-in-time copies when necessary to protect the data volumes being updated during a resynchronization.

When a network link in use by the remote mirror software becomes unavailable, the daemon attempts to execute remote mirror software update commands to resynchronize all volume sets that have autosynchronization enabled and are using the network link.

Use the `sndradm -I` command to create configuration entries marked with the `ndr_ii` key. The `ndr_ii` entries contain an additional state field that the kernel uses to determine when point-in-time copies must be made. The kernel notifies the remote mirror software synchronization daemon on the target system whenever a synchronization is started and waits for `sndrsyncd` to perform any necessary copies before allowing the synchronization to proceed.

The daemon is also notified when any remote mirror software resynchronization starts or finishes. The daemon performs point-in-time copy operations on the secondary or target host, if configured.

On a secondary host, the daemon checks whether a file system is currently mounted on the secondary volume and informs the kernel not to allow the synchronization to start if the file system is currently mounted.

See the command description for "Adding and Deleting Point-in-Time Copy Software Volumes" on page 79.

Replication Scenarios

This section describes three example scenarios:

One-to-Many Replication

Note – Consistency groups across interfaces are not supported in a one-to-many configuration because write ordering cannot be guaranteed.

In a one-to-many volume set, you can replicate data from one primary volume to many secondary volumes residing on one or more hosts. The primary volume plus each secondary site volume forms a single volume set. Each volume set requires its own unique bitmap volume.

With one primary and three secondary host volumes, you need to configure three volume sets: primary A and secondary B1, primary A and secondary B2, and primary A and secondary B3. FIGURE 2-5 shows one primary and three secondary host volumes and therefore three volume sets: A and B1, A and B2, and A and B3.

Caution – When creating volume sets, do not create secondary or bitmap volumes using partitions that include cylinder 0. Data loss might occur. See "VTOC Information" on page 17.

Consider the following:

- Each volume set on each host requires its own unique bitmap volume. For example, the volume set configurations shown in FIGURE 2-5 require three unique bitmap volumes at each host: one for each volume set (A+B1, A+B2, and A+B3).
- To forward or reverse synchronize one volume set or all volume sets, issue a separate command for each set. You update the primary volume by using a specific secondary volume.

Note – Do not reverse synchronize the primary volume from more than one secondary volume or host at a time.

- You can group one-to-many sets that share a common primary volume into a single I/O group to forward synchronize all sets simultaneously. You cannot use this technique to reverse synchronize volume sets, however. In this case, you must issue a separate command for each set and reverse update the primary volume by using a specific secondary volume.

When one-to-many replication is performed in synchronous mode, I/O from the primary volume is sent to the first secondary volume in the configuration (A+B1). The software does not wait for any I/O acknowledgment before starting to send the I/O to the second secondary volume in the configuration (B2). Writes are queued and processed in parallel. This pattern is repeated until I/O is acknowledged on all secondary volumes in the one-to-many configuration.

In a synchronous one-to-many configuration, the latency at the primary host is the combined I/O latency for each connection to a secondary host and for each disk access on the secondary hosts.

When one-to-many replication is performed in asynchronous mode, I/O is queued at the primary host for later transmission and acknowledgment for every secondary host. This scheme allows replication to proceed in parallel during one-to-many asynchronous replications.
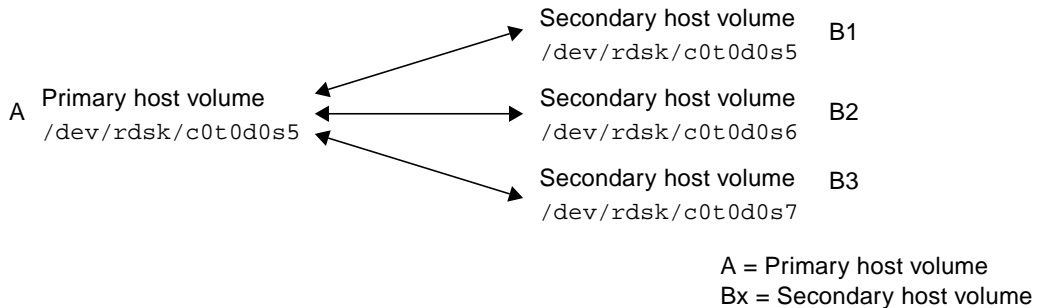
FIGURE 2-5   One-to-Many Volume Set Replication

Many-to-One Replication

The remote mirror software also supports the replication of volumes located on different hosts to volumes on a single host. The terminology differs from the one-to-many configuration terminology, where the "one" and the "many" referred to are volumes. Many-to-one configuration refers to the ability to replicate volumes across more than two hosts through more than one network connection. An example of a many-to-one configuration is shown in FIGURE 2-6.

FIGURE 2-6 shows a simple use of the many-to-one configuration. Host A backs up volumes on both Host B and Host C. As the remote mirror software does not place restrictions on many-to-one configurations, Host A can be configured as the primary host for some replicated volumes and as the secondary host for others.
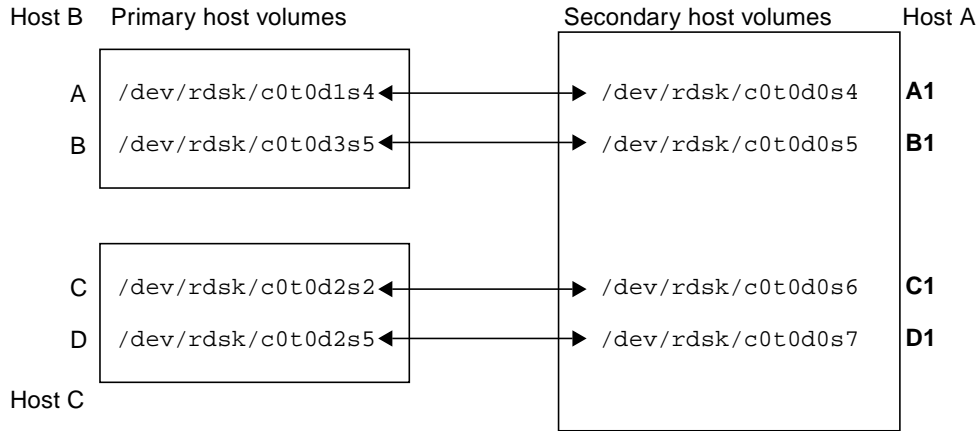
FIGURE 2-6   Many-to-One Volume Set Replication

Multihop Replication

In a multihop set, the secondary host volume of one volume set can act as the primary host volume of another volume set (it is still the secondary volume of the first volume set). In the case of one primary A and one secondary host volume B, the secondary host volume B appears as primary host volume A1 to the secondary host volume B1.

FIGURE 2-7 shows one primary volume A and its secondary host volume B. The secondary host volume B also acts as the primary host volume A1 to the secondary host volume B1.
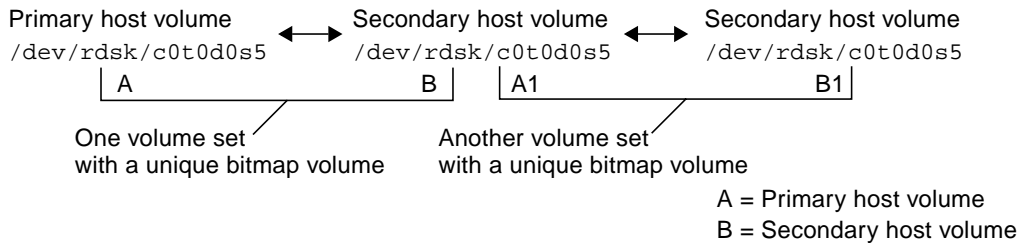
FIGURE 2-7   Multihop Volume Set Replication

Multihop configurations can become complex and the use and administration of multihop sets must be carefully considered. Consider what happens if resynchronization operations for every volume set in the multihop chain are performed in synchronous mode. The I/O proceeds along each link of the chain and the I/O acknowledgment is not confirmed until the last link is reached, at which point the process is complete.

If both sets were configured to replicate synchronously in the example in FIGURE 2-7:

■ The I/O acknowledgment from B1 would be received at A1

■ The I/O acknowledgment at B would be received at A

In a multihop configuration where every set in the chain is configured to replicate synchronously, the I/O latency at the primary node (assuming a forward replication) is the combined latency of every link and disk access along the chain.

Conversely, when volume sets are part of a multihop configuration where all sets replicate asynchronously, the contents of any given nonprimary volume is unpredictable with respect to its neighbor until the resynchronization is complete on all nodes.

These examples are for illustration only. The remote mirror software does not place any restrictions on the configurations between sets along the chain. A mix of synchronous and asynchronous sets is most useful.

As another example, configure the A+B volume set as a synchronous set running over a dark fiber in the same room (to ensure a consistent copy of the volume without adversely affecting performance on the primary site). Make the A1+B1 volume set an asynchronous set, running across a network to a remote location (to replicate the volume to a remote location at a comparatively fast rate).

Multihop configurations can be expanded and the performance of these configurations improved when the point-in-time copy software and the remote mirror software are used together.

Preparing to Use the Software

This chapter provides information about how to set up the software before you first use it. The topics include the following:

- "Increasing the Default Number of Volumes Allowed" on page 38
- "Setting Up Bitmap Volumes" on page 41
- "Customizing Volume Sets" on page 42
- "Commands and I/O Group Operations" on page 45
- "Commands and Sun Cluster Operations" on page 45
- "`dsbitmap` Bitmap Sizing Utility" on page 45

Increasing the Default Number of Volumes Allowed

The following sections describe how to change the default number of volumes you can use with the software.

- "Using More Than 64 Volume Sets" on page 39

  The default number of *remote mirror volume sets* you can enable is 64. Follow these procedures to increase this number.

- "Increasing the Storage Volume Device Limit" on page 40

  The default number of storage volume (SV) driver devices you can configure is 4096. This number of devices are divided for use between the remote mirror and point-in-time copy software. Follow these procedures to increase this number.

Note – After editing the files in this section, shut down and restart your server using the `shutdown` command for changes to take effect. Also, if you edit the `rdc.conf` file to use more than 64 volume sets, ensure that you have enough system resources.

Using More Than 64 Volume Sets

If you configure more than 64 volume sets, edit the `rdc_max_sets` field in the `/usr/kernel/drv/rdc.conf` file on each machine running the remote mirror software. The default number of configured volume sets is 64. For example, to use 128 sets, change the file as follows:

```
#
# rdc_max_sets
# - Configure the maximum number of RDC sets that can be enabled on
# this host. The actual maximum number of sets that can be
# enabled will be the minimum of this value and nsc_max_devices
#(see nsctl.conf) at the time the rdc kernel module is loaded.
#
rdc_max_sets=128;
```

Be sure to include the semicolon character (;) at the end of the `rdc_max_sets` field.

Save and close this file.

Change the number of storage volume (SV) driver devices, as described in "Increasing the Storage Volume Device Limit" on page 40

Increasing the Storage Volume Device Limit

The default number of SV driver devices (that is, volumes) you can configure is 4096, as set by the `nsc_max_devices` setting in the `nsctl.conf` file. The number of allowed volumes is divided for use between the remote mirror and point-in-time copy software. If you use the remote mirror and point-in-time copy software products together, the storage devices are divided between the two products.

The following procedure describes how to increase this default limit.

▼ To Increase the Storage Volume Limit

Caution – Increasing this limit causes more memory to be consumed. You might have to adjust the default `nsc_global_pages` value of 2 in the `/usr/kernel/drv/mc_rms.conf` file. Only an experienced system administrator should make these changes.

1. **Log in as superuser.**

2. **Open the** `/usr/kernel/drv/nsctl.conf` **file using a text editor.**

3. **Search for the** `nsc_max_devices` **field.**

4. **Edit the number in this field to increase your volume limit.**
   The default number is 4096.

5. **Save and exit the file.**

6. **Reboot your server using the** `shutdown` **command.**

Setting Up Bitmap Volumes

The remote mirror version 3.2 software does not support bitmap files. The remote mirror software uses raw devices to store bitmaps.

These raw devices must be stored on a disk separate from the disk that contains the data from the replicated volumes. Configure RAID (such as mirrored partitions) for these bitmap devices and ensure that you mirror the bitmap to another disk in a different array. The bitmap must not be stored on the same disk as the replicated volumes.

⚠ **Caution –** When creating volume sets, do not create secondary or bitmap volumes using partitions that include cylinder 0. Data loss might occur. See "VTOC Information" on page 17.

If the bitmap and the replicated volumes reside on the same disk or array, a single point of failure exists. In case of a disk or array failure, you risk a greater chance of data loss exists. The bitmap might become corrupted.

In a clustered environment, a bitmap must reside only on a volume and cannot be a file. The bitmap volume in this case must be part of the same disk group or cluster resource group as the corresponding primary or secondary data volume.

The bitmap size can be calculated using the following formula:

■ 1 Kbyte + 4 Kbyte per Gbyte of device storage space

For example, a 2-Gbyte data device requires a bitmap size of 9 Kbyte. (You can create bitmaps that are larger than the calculated size.)

See "dsbitmap Bitmap Sizing Utility" on page 45 for information about a utility that provides correct sizes for bitmap volumes.

Customizing Volume Sets

Before you start creating volume sets, see the following topics:

- "Restricted Access To Volume Sets" on page 42
- "Setting Up a Volume Set File" on page 43

See also "Reconfiguring or Modifying a Volume Set" on page 89.

Restricted Access To Volume Sets

Caution – In a clustered environment, only one system administrator or root user at a time is allowed to create and configure Sun StorEdge volume sets. This restriction helps avoid creating an inconsistent Sun StorEdge Availability Suite volume set configuration.

The operations that access the configuration include, but are not limited to:

- Creating and deleting volume sets
- Adding and removing volume sets from I/O groups
- Assigning new bitmap volumes to a volume set
- Updating the disk device group or resource name
- Any operation that changes the Sun StorEdge Availability Suite software and related volume set configuration

Note – When configuring a volume set, do not use the same volume set as a point-in-time copy shadow volume and as a remote mirror secondary volume. If you attempt to configure a volume set for two purposes, the data contained on the volume might not be valid for the application that accesses the volume.

Setting Up a Volume Set File

When you enable the remote mirror software, you can specify an optional *volume set file* containing information about the volume set: volumes, primary and secondary hosts, bitmaps, operating mode, and so on. Use the sndradm –f *volset-file* option when you use a volume set file.

You can also type information about each volume set from the command line, but it might be more convenient to put this information in a file when you have multiple volume sets.

One advantage when using volume set files is that you can operate on specific volume sets and exclude other sets from the operation. Unlike adding the volume sets to an I/O group, you can mix replication modes in a volume set file.

The fields for the volume set file specified using the –f *volset-file* option are:

phost pdev pbitmap shost sdev sbitmap ip {sync|async} [g *io–groupname*] [C *tag*] –q *qdev*

An example file entry is as follows:

```
atm10 /dev/vx/rdsk/oracle816/oratest /dev/vx/rdsk/oracle816/oratest_bm \
atm20 /dev/vx/rdsk/oracle816/oratest /dev/vx/rdsk/oracle816/oratest_bm \
ip sync g oragroup
```

See TABLE 3-1 for descriptions of the format fields. See the rdc.cf man page for more information about the volume set file format.

TABLE 3-1   Volume Set File Format Fields

Field	Meaning	Description
phost	Primary host	Server on which the primary volume resides.
pdev	Primary device	Primary volume partition. Specify full path names only (for example, `/dev/rdsk/c0t1d0s4`).
pbitmap	Primary bitmap	Volume partition in which the bitmap of the primary partition is stored. Specify full path names only.
shost	Secondary host	Server on which the secondary volume resides.
sdev	Secondary device	Secondary volume partition. Specify full path names only.
sbitmap	Secondary bitmap	Volume partition in which the bitmap of the secondary partition is stored. Specify full path names only.
ip	Network transfer protocol	Specify `ip`.
sync | async	Operating mode	`sync` is the mode in which the I/O operation is confirmed as complete only when the remote volume has been updated. `async` is the mode in which the primary host I/O operation is confirmed as complete before updating the remote volume.
g *io-groupname*	I/O group name	An I/O group name can be specified using the `g` character. In the preceding example, it is `oragroup`.
qdev	Disk queue	Volume to use for a disk-based queue.

Commands and I/O Group Operations

Adding the remote mirror software volume sets to an I/O group enables you to issue a single command that operates on all volume sets in the specified I/O group, and excludes all other volume sets. Most commands allow for group operations and perform them when you include -g *io–groupname* in the command syntax.

The operations performed are independent of each other. Operations performed on I/O group A, volume set 1 are independent of operations performed on I/O group A, volume set 2.

Failed Operations in an I/O Group

If an operation fails on one or more volume sets in an I/O group, the state of the data on the failed volumes in the volume sets is unknown. To correct this:

1. Correct any known problems with the failing sets

2. Reissue the command on the I/O group

Commands and Sun Cluster Operations

You use the C *tag* and –C *tag* options described in Chapter 5 in Sun Cluster 3.0 Update 3 and Sun Cluster 3.1 environments only. If you accidentally use these options in a noncluster environment, the remote mirror operation does not execute.

`dsbitmap` Bitmap Sizing Utility

The `dsbitmap` utility is installed with the Sun StorEdge Availability Suite 3.2 software. Use it to calculate the required size of a bitmap for a point-in-time copy shadow volume set or a remote mirror volume set.

The `dsbitmap` utility is typically used by the system administrator during the initial stages of configuring Sun StorEdge Availability Suite software. The utility determines the required bitmap volume size and then verifies that the bitmap volumes are suitable.

```
dsbitmap
```

This utility enables you to determine the size of the bitmap volume that is required for a remote mirror bitmap or a point-in-time copy bitmap. If you include a proposed bitmap volume in the command, dsbitmap tests its suitability as a bitmap volume for the proposed data volume.

Syntax

To obtain the size of a point-in-time copy bitmap, use this command:

```
dsbitmap -p datavolume [bitmap_volume]
```

To obtain the size of a remote mirror bitmap, use this command:

```
dsbitmap -r datavolume [bitmap_volume]
```

Usage for dsbitmap

```
# dsbitmap -h
usage: dsbitmap -h
       dsbitmap { -p | -r } data_volume [bitmap_volume]
       -h : This usage message
       -p : Calculate size of Point in Time bitmap
       -r : Calculate size of Remote Mirror bitmap
```

Examples for dsbitmap

Remote mirror volumes display both memory and disk queue sizes:

```
# dsbitmap -r /dev/md/rdsk/d100
Remote Mirror bitmap sizing

Data volume (/dev/md/rdsk/d100) size: 2064384 blocks
Required bitmap volume size:
  Sync replication: 9 blocks
  Async replication with memory queue: 9 blocks
  Async replication with disk queue: 73 blocks
```

- Point-in-time copy volumes display sizes for independent, dependent and compact dependent sizes:

```
# dsbitmap -p /dev/md/rdsk/d100
Point in Time bitmap sizing

Data volume (/dev/md/rdsk/d100) size: 2064384 blocks
Required bitmap volume size:
  Independent shadow: 62 blocks
  Full size dependent shadow: 62 blocks
  Compact dependent shadow: 798 blocks
```

Using the Remote Mirror Software

This chapter includes examples of how to use the remote mirror software command `sndradm`. The topics in this chapter include:

TABLE 4-1 lists the names used in the examples in this chapter:

TABLE 4-1   Example Names and Devices Used in This Chapter

Primary host name	`rmshost1`
Primary volume	`/dev/vx/rdsk/c0t117d0s3`
Primary bitmap	`/dev/vx/rdsk/bmap/bm1`
Secondary host name	`rmshost2`
Secondary volume	`/dev/vx/rdsk/c0t117d0s5`
Secondary bitmap	`/dev/vx/rdsk/bmap/bm2`
Set name (assigned by the software)	`rmshost2:/dev/vx/rdsk/c0t117d0s5`

Depending on the example, either site can be the primary or secondary host of the remote copy operation. As shown in TABLE 5-4, you must perform all synchronization operations from the primary host session.

To monitor the operation of the remote mirror software, use the `/usr/opt/SUNWscm/sbin/dsstat` command described in Appendix A.

Getting Started

Note – Use the same disk management method (software volume management or raw disk) with the remote mirror software and the point-in-time copy software on the primary site and the secondary site volumes. The Sun StorEdge Availability Suite software replicates data at the block level, and the block counts for a given size are different for a disk slice and a volume. Therefore, a remote mirror replication or reverse sync or a point-in-time full independent copy or reverse copy, might fail if the target size is smaller than the source size.

The following sections describe the initial steps for using the remote mirror software:

- "Enabling the Volume Sets" on page 50
- "Establishing Volume Copies for the First Time" on page 52
- "Updating the Secondary Volume" on page 53

Enabling the Volume Sets

The first step to using the remote mirror software is to enable the software on volume sets. Make sure you perform this step on both the primary and secondary hosts. A common user error is enabling volume sets on only one host.

Caution – When creating volume sets, do not create secondary or bitmap volumes using partitions that include cylinder 0. Data loss might occur. See "VTOC Information" on page 17.

In this example, enable the hosts in asynchronous mode. The example also shows that the volumes need to be synchronized and do not match (in this case, use the `sndradm -e` command).

▼ To Enable the Volume Sets

1. Log in to the primary host `rmshost1` **as the superuser.**

2. Enable the volume sets:

```
rmshost1# sndradm -e rmshost1 /dev/vx/rdsk/c0t117d0s3 /dev/vx/rdsk/bmap/bm1 \
rmshost2 /dev/vx/rdsk/c0t117d0s5 /dev/vx/rdsk/bmap/bm2 ip async
```

3. Log in to the secondary host `rmshost2` **as the superuser.**

4. Enable the volume sets:

```
rmshost2# sndradm -e rmshost1 /dev/vx/rdsk/c0t117d0s3 /dev/vx/rdsk/bmap/bm1 \
rmshost2 /dev/vx/rdsk/c0t117d0s5 /dev/vx/rdsk/bmap/bm2 ip async
```

The following events occur:

- The software defines the volume set name as
  `rmshost2:/dev/vx/rdsk/c0t117d0s5`. You can now use this name with the
  `sndradm` command instead of specifying each item in the volume set each time.

- The bitmaps are set to 1, indicating that a full synchronization between volumes
  is required.

Establishing Volume Copies for the First Time

The next step is to perform a full forward synchronization to copy the primary volume contents to the secondary volume. The remote mirror software performs this initial primary-volume-to-secondary-volume copy while also forwarding any new primary volume updates to the secondary volume.

The volumes can be made identical using methods other than a full synchronization. When network latencies justify it, you can perform the initial synchronization of a volume set by backing up a source or primary volume on magnetic tape on one site, then restoring the volume from the tape on the other site. During the period between when the backup is finished and the restore is started, place the source or primary volume in logging mode. Make sure that the backup copy is a physical copy (for example, by using the dd(1M) command) and not a logical copy (for example, one made using the tar(1M) or cpio(1M) commands). The copies must have identical blocks, not just identical files. In this case, use the sndradm -E command instead of sndradm -e to enable the volume sets.

▼ To Synchronize the Volumes During Update

1. **Log in to the primary host** rmshost1 **as superuser.**

2. **Unmount the secondary volume. You can keep the primary volume mounted.**

3. **Synchronize the volumes:**

   ```
   rmshost1# sndradm -m rmshost2:/dev/vx/rdsk/c0t117d0s5
   ```

4. **Check the synchronization progress:**

   ```
   rmshost1# dsstat -m sndr
   ```

After the synchronization is complete, the remote mirror software continues to replicate any primary volume changes to the secondary volume. Keep the secondary volume unmounted during replication or until you are ready to permit your application to write to it. When you are ready to permit write operations to the secondary volume, place the volume sets into logging mode and mount the volume. The software continues tracking changes through the bitmap until you are ready to update or resynchronize the volumes.

Updating the Secondary Volume

This section describes the commands to use when you are ready to update the secondary volume to resynchronize volumes.

Optionally, you can also use the Sun StorEdge Availability Suite 3.2 point-in-time copy software to ensure data consistency at the primary and secondary volumes. Using this software helps guarantee a known good copy of your data in case of a network link failure during the synchronization. See the *Sun StorEdge Availability Suite 3.2 Point-in-Time Copy Software Administration and Operations Guide* for details about the `iiadm` command.

Note – You must place the related remote mirror volume set in logging mode for the point-in-time copy software to successfully perform an enable, copy, update, or reset operation on a remote mirror volume. If the volume set is not in logging mode, the point-in-time copy operation fails and the remote mirror software reports that the operation is denied

---

▼ To Resynchronize Primary and Secondary Volumes

1. **Log in to the primary host** `rmshost1` **as the superuser.**

2. **Quiesce any applications writing to the primary volume.**

Note – You are not necessarily required to quiesce your applications but doing so helps guarantee a consistent copy of your data. Quiescing applications also helps guarantee a consistent copy for the point-in-time copy software. If you choose not to quiesce your applications and not to use the point-in-time copy software, the remote mirror software still forwards any data updates to the secondary volume during replication.

---

3. **(Optional) Take a point-in-time copy snapshot of the primary volume. Ensure that the primary volume is in logging mode (`sndradm -l`) and then use the `iiadm` command.**

4. **(Optional) Take a point-in-time copy snapshot of the secondary volume. Ensure that the secondary volume is in logging mode and then use the `iiadm` command.**

5. **Copy only the changed data from the primary volume to the secondary volume:**

```
rmshost1# sndradm -u rmshost2:/dev/vx/rdsk/c0t117d0s5
```

6. **Check the synchronization progress:**

```
rmshost1# dsstat -m sndr
```

When the update synchronization is completed, the secondary volume is a block-for-block copy of the primary volume and the bitmaps are cleared to 0. See "If a Network Link Fails" on page 54.

If a Network Link Fails

The remote mirror software uses a periodic signal to monitor the health of primary and secondary systems. If the software cannot sense a health monitor signal, it assumes an interruption in the remote mirror software service.

The remote mirror software then places all volume sets at the primary site into logging mode. During logging mode, the software updates only the primary volume bitmap. (The software assumes that the secondary volume is not mounted and not being written to.) See "Logging" on page 27 and "Stopping Replication and Starting Logging" on page 81.

Note – The secondary site does not automatically switch to logging mode if a break in the network occurs or if the primary site is down. This is to protect the data at the secondary site from unwanted writes. An administrator must cause the software to actually fail over to the secondary site by issuing the sndradm -l command on the secondary.

You can introduce interruptions intentionally to exercise remote failure strategies, for example, during the disaster recovery rehearsals described in "Rehearsing Disaster Recovery" on page 56.

When Not To Resynchronize Volumes

Resynchronization is discouraged if the interruption might be the warning of a larger, rolling disaster. Maintain the secondary site in a dated but consistent state rather than risk a disastrous interruption that leaves the secondary site inconsistent and difficult to recover from. The autosynchronization option is disabled by default for this reason. See "Choosing Automatic or Manual Resynchronization" on page 28.

Autosynchronization

The autosynchronizatin feature is designed to synchronize the primary and secondary volumes after a network link failure has been restored. When autosynchronizatin is enabled, the operation occurs only when replication occurs. For example, when you enable autosynchronization for a set, the software attempts to synchronize the primary and secondary volumes only when replication is performed. When you put the set into logging mode, the software does not synchronize the primary and secondary volumes. However, putting the set into logging mode does not disable autosynchronization. When a new synchronization request is issued, for example using the sndradm -u command, the autosynchronization feature becomes active again.

To enable or disable autosynchronization, use the sndradm -a command described in "Enabling or Disabling Autosynchronization" on page 103. See also "Choosing Automatic or Manual Resynchronization" on page 28 and "Autosynchronization" on page 30.

Resynchronizing Volumes Manually

Caution – While resynchronization is occurring, the secondary volume data is temporarily inconsistent and cannot be relied on for recovery. Consistency is restored when the resynchronization finishes executing. To ensure data integrity, use the point-in-time copy software regularly to create a snapshot of data at both sites.

Typically, interruptions in the remote mirror software services are infrequent.

If the secondary volume state is unknown because of system or disk failure, make full volume copies to reestablish matching remote mirror software volume sets. In this case, use the sndradm -m command to fully update the secondary volume set.

Perform the procedures in "To Synchronize the Volumes During Update" on page 52.

Rehearsing Disaster Recovery

The remote mirror software enables you to perform disaster rehearsals, encouraging verification of your disaster plans. Perform rehearsals regularly and refine them whenever a significant change is made to the primary or secondary host environments.

When rehearsing disasters or in an actual disk disaster or failure, keep the failed volumes under control of the remote mirror software. Do not disable the software. The remote mirror software marks the device as failed when the software is unable to read from or write to the device. In the case of a primary failure, for example, the remote mirror software continues to provide read and write services to the host application using the secondary volume at the remote site.

▼ To Rehearse a Primary Volume or Site Failure

1. **Simulate a primary volume or site disaster by using one of the following methods:**

   ■ Break the network link between the primary and secondary sites by disconnecting the network cable or otherwise shutting down the network link. In this case, the remote mirror software places the volumes at the primary site into logging mode after detecting the break.

   ■ Place the primary and secondary volumes into logging mode. See "Stopping Replication and Starting Logging" on page 81.

2. **When the data is destaged, mount the secondary volume in read-write mode so your application can write to it.**

3. **Configure your application to read and write to the secondary volume.**

   The secondary bitmap volume tracks the volume changes.

4. **Fix the "failure" at the primary volume by using one of the following methods:**

   ■ Reconnect the network.

   ■ Disable logging mode by issuing a resynchronization command. Quiesce your application and unmount the secondary volume if you choose this method.

Note – If the autosynchronization feature is enabled, the remote mirror software resynchronizes the primary volume from the secondary volume when the link is re-established. If you also install and configure the point-in-time copy software, this software takes a snapshot copy of your secondary volume data before performing the reverse update synchronization. Consider whether this approach is appropriate for your disaster recovery plan.

You can now choose to resynchronize your volumes:

5. **Perform an update by choosing one of the following methods:**

   - Roll back the secondary volume to its original predisaster state to match the primary volume. Use the `sndradm -m` copy or `sndradm -u` update commands.

   - Keep the changes from the updated secondary volume and resynchronize so that both volumes match. Use the `sndradm -m r` reverse copy or `sndradm -u r` reverse update commands.

▼ To Rehearse a Secondary Volume or Site Failure

1. **Simulate a secondary volume or site disaster by using one of the following methods:**

   - Break the network link between the primary and secondary sites by disconnecting the network cable or otherwise shutting down the network link. In this case, the remote mirror software places the primary volumes into logging mode after detecting the break.

   - Place the primary and secondary volumes into logging mode. See "Stopping Replication and Starting Logging" on page 81.

2. **Fix the failure at the primary volume by using one of the following methods:**

   - Reconnect the network.

   - Disable logging mode by issuing a resynchronization command. Unmount the secondary volume if you choose this method.

Note – If the autosynchronization feature is enabled, the remote mirror software resynchronizes the secondary volume from the primary volume when the link is re-established. If you also install and configure the point-in-time copy software, this software takes a snapshot copy of your secondary volume data before performing the reverse update synchronization. Consider whether this approach is appropriate for your disaster recovery plan.

You can now choose to resynchronize your volumes.

3. **Perform the update by choosing one of the following methods:**

   - Update the secondary volume to match the primary volume. Use the `sndradm -u` update command.

   - If you are unsure about the secondary volume contents, perform a full forward synchronization to match the primary volume. Use the `sndradm -m` synchronization command.

Handling Primary Volume Failures

Note – Keep the failed volumes under control of the remote mirror software. Do not disable the software. The remote mirror software marks the device as failed when the software is unable to read to or write from the device. The remote mirror software continues to provide read and write services to the host application using the secondary volume at the remote site.

The remote mirror software provides continuous data access during primary volume failures. The remote mirror software high-availability features are a superset of RAID 1 and RAID 5 storage protection that can be optionally configured for the primary volumes. The remote mirror software's remote volume access features start only after the disk protection schemes on the primary system are unable to provide data access to the local devices.

In the linear and striped (RAID 0) cases, failure of a single disk storing the primary volume triggers the remote mirror software to transparently redirect disk reads and writes to the remote storage system.

If the primary logical volume is locally mirrored (RAID 1) across two physical disks on the same system, a single disk failure results in its local mirror disk handling all requests for cache staging on a read miss and cache destaging. The remote mirror software relies on the remote site secondary devices only if both local mirrors fail.

If the primary volume is RAID 5 protected, its contents are striped across several physical disks. The local system considers the primary volume inaccessible and yields to the remote mirror software's remote volume access only when two or more of the disks in the RAID 5 stripe fail.

Recovering From a Primary Site Disaster

The remote mirror software minimizes the effects of a disaster at the primary site by enabling you to keep the secondary storage images updated. Although the secondary remote mirror software cache contains the latest writes issued on the primary before the disaster, that data might not have been destaged to the secondary disks yet. After detecting an interruption in the remote mirror software service, the remote mirror software automatically destages the secondary remote mirror software cache to its corresponding secondary volumes.

After all the secondary volumes have been updated with the latest remote mirror software cache images, the secondary volumes can be accessed by the secondary hosts. The dsstat command shows information confirming that destaging is complete. Run application-level recovery procedures to ensure a well-known state at the secondary site. The workload can then be switched to the secondary hosts for continued business operation.

Until the extent of the primary failure is understood, keep the remote mirror software enabled at the secondary site to track disk areas that are being modified.

Restoring a Primary Site From the Secondary Site

If the primary host becomes inoperative and primary data on the primary disks is lost, update logs at the secondary systems have little value. You must flush the cache and perform a full reverse synchronization on the repaired or replaced primary host. In other words, volume-to-volume copies from the secondary to the primary are required for all remote mirror software-managed volumes. This reverse synchronization process ensures that only the latest data is deposited on the primary disks. See "Rehearsing Disaster Recovery" on page 56.

Disabling Remote Replication

> **Caution –** Disable remote replication *only* when the primary and secondary volumes will no longer be associated.

Disabling the remote mirror software breaks the connection between primary and secondary volumes, discards any bitmap information, and removes the host and volume information from the Sun StorEdge configuration. After disabling the remote mirror software, enable and full synchronization (full volume copy) operations are necessary to re-establish the remote mirror software relationship and to ensure that the contents of each volume match. See "Enabling and Disabling Volume Sets" on page 76 and "Establishing Volume Copies for the First Time" on page 52.

Swapping the Remote Mirror Hosts

Caution – Before performing this procedure, ensure that I/O operations are not occurring to the volumes at the primary and secondary hosts. Data corruption will result if I/O operations continue.

In case of disaster recovery or link failure situations, you can also swap the remote mirror host roles to provide access to your critical data. That is, the primary host can become the secondary host and the secondary host can become the primary host. This alternate scheme enables you to recover the old primary host and, if you choose, switch back to the original roles.

The basic steps to swap host roles are as follows:

1. Quiesce the application accessing the primary volume. Unmount the volume if necessary.

2. Disable the remote mirror software on the primary site (Site-A). This step also discards the primary bitmap. A full copy is required when the set is enabled.

3. Disable the remote mirror software on the secondary site (Site-B).

4. Enable the remote mirror software on the new primary site (Site-B) with Site-B specified as the primary site.

5. Enable the remote mirror software on the new secondary site (Site-A) with Site-A specified as the secondary site.

6. At the new primary site (Site-B), synchronize the volumes from the primary to the secondary.

7. Perform any modifications or recovery procedures required by your application. For example, if you are using a database application, you might have to copy data and control files to the new secondary host after the synchronization.

8. Restart the application on the new primary site (Site-B). Mount the volumes if necessary.

Note – It may be useful to use a volume set file for ease of administration. See "Setting Up a Volume Set File" on page 43.

Sample `rdc.cf` File

The following is an example of a `/etc/opt/SUNWrdc/rdc.cf` volume set file. See also "Setting Up a Volume Set File" on page 43.

```
rmshost1 /dev/vx/rdsk/c0t117d0s3 /dev/vx/rdsk/bmap/bm1 \
rmshost2 /dev/vx/rdsk/c0t117d0s5 /dev/vx/rdsk/bmap/bm2 ip sync
```

Your actual volume set file can be any name. The `rdc.cf` file name is used as an example here.

TABLE 4-2   Example Names and Devices Used in This Procedure

Primary host name (Site-A)	`rmshost1`
Primary volume	`/dev/vx/rdsk/c0t117d0s3`
Primary bitmap	`/dev/vx/rdsk/bmap/bm1`
Secondary host (Site-B)	`rmshost2`
Secondary volume	`/dev/vx/rdsk/c0t117d0s5`
Secondary bitmap	`/dev/vx/rdsk/bmap/bm2`
Transmission protocol	`ip`
Replication mode	`sync`
Set name (assigned by the software)	`rmshost2:/dev/vx/rdsk/c0t117d0s5`

▼ To Disable the Software at Site-A

⚠ **Caution –** Before performing this procedure, ensure that I/O operations are not occurring to the volumes at the primary and secondary hosts. Data corruption will result if I/O operations continue to occur. Before performing this procedure, quiesce the application writing to the remote mirror volumes and then unmount those volumes.

The following example assumes that the `/etc/opt/SUNWrdc/rdc.cf` volume set file is already created and the volumes specified in it are enabled.

1. **At Site-A, disable the remote mirror software and discard the remote mirror scoreboard bitmap:**

```
rmshost1# sndradm -dn -f /etc/opt/SUNWrdc/rdc.cf
```

2. **Edit the `rdc.cf` file to swap the Site-A primary host information and Site-B secondary host information.**

   For example, in the entry shown in "Sample `rdc.cf` File" on page 63, change `rmshost1` to `rmshost2` and `rmshost2` to `rmshost1`.

3. **If possible, unmount the remote mirror volumes:**

```
rmshost1# umount mount-point
```

▼ To Change the Site-B Secondary Host to the Primary Host

1. **At Site-B, disable the remote mirror software and discard the remote mirror scoreboard bitmap:**

```
rmshost2# sndradm -dn -f /etc/opt/SUNWrdc/rdc.cf
```

2. **Edit the `rdc.cf` file to swap the Site-A primary host information and Site-B secondary host information.**

   For example, in the entry shown in "Sample `rdc.cf` File" on page 63, change `rmshost1` to `rmshost2` and `rmshost2` to `rmshost1`.

3. **Enable the remote mirror software *at both hosts*:**

```
rmshost1# sndradm -En -f /etc/opt/SUNWrdc/rdc.cf
rmshost2# sndradm -En -f /etc/opt/SUNWrdc/rdc.cf
```

   Use the -E enable option to ensure that the bitmap contents are clear (0), indicating that a synchronization is not required.

4. **If desired, at Site-A, perform a full synchronization from Site-B to Site-A.**

```
rmshost1# sndradm -mn -f /etc/opt/SUNWrdc/rdc.cf
```

5. **Perform any modifications or recovery procedures required by your application.**

The `sndradm` Command

Note – To use the remote mirror software, you must be the root user.

This chapter describes the remote mirror software's
`/usr/opt/SUNWesm/sbin/sndradm` command and options.

The topics in this chapter include:

Log File

The `/var/opt/SUNWesm/ds.log` file contains operation log messages for the remote mirror commands. For example, these log messages show a volume set enabling and updating:

```
Oct 17 13:48:10 sndr: sndradm -e atm10 /dev/vx/rdsk/oracle816/oratest
/dev/vx/rdsk/oracle816/oratest_bm atm20 /dev/vx/rdsk/oracle816/oratest
/dev/vx/rdsk/oracle816/oratest_bm ip sync g oragroup
Successful
Oct 17 13:48:10 sv: enabled /dev/vx/rdsk/oracle816/oratest
Oct 17 13:48:10 sv: enabled /dev/vx/rdsk/oracle816/oratest_bm
Oct 18 10:23:54 sndr: sndradm -u atm10 /dev/vx/rdsk/oracle816/oratest
/dev/vx/rdsk/oracle816/oratest_bm atm20 /dev/vx/rdsk/oracle816/oratest
/dev/vx/rdsk/oracle816/oratest_bm ip sync g oragroup
Starting
```

Syntax Summary

Note – If you use a command without specifying any volume set arguments, the command operates on all remote mirror volume sets.

Syntax

`sndradm` *switches* [*options*]

For example, to update the secondary volume from the primary volume:

`sndradm -u` [`-g` *io-groupname*] [`-C` *tag*] [`-n`] [`-f` *volset-file* | *volset* | *set-name*]

The `/usr/opt/SUNWesm/sbin/sndradm` command uses the command switches shown in TABLE 5-1. TABLE 5-2 shows the commands sorted in alphabetical order and TABLE 5-3 shows the options.

Command Switches

TABLE 5-1 Command Switches Sorted By Function

Function	Command and Switch	See This Section
Enable and Disable Volume Sets		
Enable the software for the specified set (sync not required).	sndradm -E	"Use the following commands to enable and disable volume sets." on page 76
Enable the software for the specified set (sync required).	sndradm -e	"Use the following commands to enable and disable volume sets." on page 76
Disable the software for the specified set.	sndradm -d	"Disabling Volume Sets" on page 78
Add or delete a point-in-time copy software volume group.	sndradm -I	"Adding and Deleting Point-in-Time Copy Software Volumes" on page 79
Stop the software replication, enter logging mode, and track changes using bitmaps.	sndradm -l	"Stopping Replication and Starting Logging" on page 81
Synchronize or Update Volume Sets		
Copy the entire contents of the primary volume to the secondary volume; also known as *full synchronization.*	sndradm -m	"Starting a Full Synchronization" on page 84
Copy the entire contents of the secondary volume to the primary volume; also known as *full reverse synchronization.*	sndradm -m -r	"Starting a Full Synchronization" on page 84
Update only the changed data of the primary volume to the secondary volume; also known as *update synchronization* or *resynchronization.*	sndradm -u	"Starting an Update Resynchronization" on page 86
Update only the changed data of the secondary volume to the primary volume; also known as *reverse synchronization* or *reverse update.*	sndradm -u -r	"Starting an Update Resynchronization" on page 86
Wait for a synchronization operation to finish executing.	sndradm -w	"Waiting for Synchronization to Finish Executing" on page 88
Reconfigure or Change Volume Sets		
Change the bitmap of a volume set.	sndradm -R b	"Reconfiguring or Modifying a Volume Set" on page 89
Update or reconfigure the disk group name or cluster resource tag.	sndradm -R C	"Updating the Disk Cluster Tag Name" on page 97

TABLE 5-1   Command Switches Sorted By Function *(Continued)*

Function	Command and Switch	See This Section
Update or reconfigure the software volume sets.	`sndradm -R -f` *`volset-file`*	"Updating the Volume Set Information" on page 96
Move all the specified remote mirror sets to a different group.	`sndradm -R g`	"Moving a Volume Set to a Different I/O Group" on page 91
Reset the replication mode of a volume set.	`sndradm -R m` `{sync|async}`	"Changing the Replication Mode of a Volume Set" on page 94
Display Information		
Display the command syntax.	`sndradm -h`	See the display that results from entering this command at the command prompt
Display the version information.	`sndradm -v`	See the display that results from entering this command at the command prompt

Display the software status.	`sndradm -p`
	
	`sndradm -P`
		"Displaying Volume Set and I/O Group Status" on page 99

Display the software volume set and I/O group names.	`sndradm -i`	"Listing of Volume Sets and I/O Group Names" on page 100
Display the status of the link connecting systems running the software.	`sndradm -H`	"Displaying Link Status" on page 102
Disk Queue Commands		
Set disk queue blocking mode.	`sndradm -D`	"Managing Disk Queues" on page 109
Add disk queue to group.	`sndradm -g` *`group`* `-q a` *`vol`*	"Managing Disk Queues" on page 109
Remove disk queue from group.	`sndradm -g` *`group`* `-q d`	"Managing Disk Queues" on page 109
Replace disk queue for a group.	`sndradm -g` *`group`* `-q r` *`newvol`*	"Managing Disk Queues" on page 109
Add disk queue to a set.	`sndradm -q a` *`vol shost:sdev`*	"Managing Disk Queues" on page 109
Remove disk queue from a set.	`sndradm -q d` *`shost:sdev`*	"Managing Disk Queues" on page 109

TABLE 5-1   Command Switches Sorted By Function *(Continued)*

Function	Command and Switch	See This Section

Replace disk queue for a set.	`sndradm -q r`
	`newvol`
	`shost:sdev`
		"Managing Disk Queues" on page 109

Miscellaneous		
Set the number of asynchronous threads	`sndradm -A`	"Setting the Number of Asynchronous Threads" on page 110
Set and tune the asynchronous queue.	`sndradm -W` `sndradm -F`	"Setting the Asynchronous Queue" on page 104
		"Tuning the Asynchronous Queue" on page 105
Toggle the autosynchronization state.	`sndradm -a` `{on|off}`	"Enabling or Disabling Autosynchronization" on page 103

TABLE 5-2   Command Switches in Alphabetical Order

Command and Switch	Description	See This Section
`sndradm -a {on|off}`	Autosynchronization on or off.	"Enabling or Disabling Autosynchronization" on page 103
`sndradm -A`	Set the number of asynchronous threads.	
`sndradm -e`	Enable the software for the specified set (sync required).	"Use the following commands to enable and disable volume sets." on page 76
`sndradm -E`	Enable the software for the specified set (sync not required).	"Use the following commands to enable and disable volume sets." on page 76
`sndradm -d`	Disable the software.	"Disabling Volume Sets" on page 78
`sndradm -D`	Set disk queue blocking mode.	
`sndradm -g group -q a`	Add disk queue to a group.	
`sndradm -g group -q d`	Remove disk queue from a group.	
`sndradm -g group -q r`	Replace disk queue for a group.	
`sndradm -h`	Display the command syntax.	See the display that results from entering this command at the command prompt

TABLE 5-2 Command Switches in Alphabetical Order *(Continued)*

Command and Switch	Description	See This Section
sndradm -H	Display the status of the link connecting systems running the software.	"Displaying Link Status" on page 102
sndradm -i	Display the software volume set and I/O group names.	"Displaying Volume Set and I/O Group Status" on page 99
sndradm -I	Add or delete a Sun StorEdge point-in-time copy software volume group for use with the software.	"Adding and Deleting Point-in-Time Copy Software Volumes" on page 79
sndradm -l	Place the volume sets into logging mode.	"Stopping Replication and Starting Logging" on page 81
sndradm -m	Copy the entire contents of the primary volume to the secondary volume.	"Starting a Full Synchronization" on page 84
sndradm -m -r	Copy the entire contents of the secondary volume to the primary volume.	"Starting a Full Synchronization" on page 84
sndradm -p	Display software status.	"Listing of Volume Sets and I/O Group Names" on page 100
sndradm -P	Display software status.	"Listing of Volume Sets and I/O Group Names" on page 100
sndradm -q a	Add disk queue to a set.	
sndradm -q d	Remove disk queue from a set.	
sndradm -q r	Replace disk queue for a set.	
sndradm -R	Use after fixing a failed bitmap volume. It clears the BMF error and commits the current state of the remote mirror set to the bitmap header.	
sndradm -R b	Reconfigure the volume set by changing the bitmap volume.	"Changing the Bitmap Volume in a Volume Set" on page 90
sndradm -R C	Reconfigure the disk group name or cluster resource tag.	"Updating the Disk Cluster Tag Name" on page 97
sndradm -R g	Reconfigure the specified volume sets by moving them to a different group.	"Moving a Volume Set to a Different I/O Group" on page 91
sndradm -R -f *volset-file*	Reconfigure the volume sets specified in the *volset-file*.	"Updating the Volume Set Information" on page 96

TABLE 5-2 Command Switches in Alphabetical Order *(Continued)*

Command and Switch	Description	See This Section
`sndradm -R m {sync|async}`	Reconfigure the replication mode of a volume set.	"Changing the Replication Mode of a Volume Set" on page 94
`sndradm -u`	Update only the changed data of the primary volume to the secondary volume; also known as *update synchronization* or *resynchronization.*	"Starting an Update Resynchronization" on page 86
`sndradm -u -r`	Update only the changed data of the secondary volume to the primary volume; also known as reverse synchronization or reverse update.	"Starting an Update Resynchronization" on page 86
`sndradm -v`	Displays the version information.	See the display that results from entering this command at the command prompt
`sndradm -w`	Wait for a synchronization operation to finish executing.	"Waiting for Synchronization to Finish Executing" on page 88
`sndradm -W` `sndradm -F`	Set and tune the asynchronous queue.	"Setting the Asynchronous Queue" on page 104 "Tuning the Asynchronous Queue" on page 105

Command Options

TABLE 5-3 Command Options

Option	Definition
`-n`	Does not prompt the user after issuing a `sndradm` command. The default is to prompt and request a response. For example, after starting a full synchronization from the primary to secondary volumes, the remote mirror software prompts `Overwrite secondary with primary? (Y/N) [N]`. This option is useful when you include the commands in a script file.
`-f` *volset-file*	Specifies a volume set file defining the remote mirror software volume sets. If you do not specify a *volset-file*, *volset*, or *set-name*, the command acts on all volume sets.
volset	Specifies the complete remote mirror software set configuration information. If you do not specify a *volset-file*, *volset*, or *set-name*, the command acts on all volume sets. The *volset* format is as follows: *phost pdev pbitmap shost sdev sbitmap* ip {sync | async} [g *io-groupname*] [C *tag*]

TABLE 5-3   Command Options *(Continued)*

Option	Definition
	phost – Server on which the primary volume resides.
	pdev – Primary volume partition to be copied. Specify only full path names, for example, `/dev/rdsk/c0t1d0s4`.
	pbitmap – Volume partition where the bitmap of the primary partition is stored. Specify only full path names, for example, `/dev/rdsk/c0t1d0s6`.
	shost – Server on which the secondary volume resides.
	sdev – Secondary volume partition. Specify only full path names, for example, `/dev/rdsk/c0t1d0s7`.
	sbitmap – Volume partition in which the bitmap of the secondary partition is stored. Specify only full path names, for example, `/dev/rdsk/c0t1d0s8`.
	`ip` – Specifies the network protocol.
	`sync` | `async` – Specifies the remote mirror software operating mode. `sync` is the mode in which the I/O operation is not confirmed as complete until the remote volume has been updated. `async` is the mode in which the primary host I/O operation is confirmed as complete before updating the remote volume.
	`g` *io–groupname* – Specifies which I/O group the volume set belongs to.
	`C` *tag* - Specifies the disk group name or resource tag of the local data and bitmap volumes in cases where this is not implied by the name of the volume. (For example, `/dev/rdsk/md/dg/vol` and `/dev/vx/rdsk/dg/vol` both indicate a disk group name of `dg`.) You must ensure that the cluster tag specified to the remote mirror software matches the appropriate cluster resource group tag and to keep all installed data services updated in the event of cluster resource group reconfigurations. You can use the `C` tag and `-C` tag options in a Sun Cluster 3.0 Update 3 or Sun Cluster 3.1 environment. If you use these options in another environment, the remote mirror operation does not execute.
	qdev - Specifies a disk queue device.
set–name	Specifies the name of the volume set as assigned by the remote mirror software. The software assigns a default volume set name of *shost*:*sdev*, where *shost* is the secondary host name and *sdev* is the secondary volume partition name, separated by a colon (:).
`-g` *io–groupname*	Specifies the name of the I/O group containing the collection of remote mirror software volume sets. Specifying `-g` *io–groupname* limits any operation performed to the sets in the named *io–groupname*.
`-C` *tag*	You can use the `C` *tag* and `-C` *tag* options only in a Sun Cluster 3.0 Update 3 or Sun Cluster 3.1 environment. If you use these options in a noncluster environment, the remote mirror operation does not execute. Specifying `-C` *tag* limits any operation performed to the sets in the named *tag* cluster resource tag.

Which Host Do I Issue Commands From?

TABLE 5-4 describes the commands and synchronization situations which can be performed from a primary host, secondary host, or both.

TABLE 5-4   Which Host to Issue Commands From

Task	Where Command Is Issued	Comments
Assign a new bitmap to a volume set.	Primary and secondary host	Perform this command on the host where the new bitmap resides and is being assigned. Then perform it on the other host.
Disable the software.	Primary or secondary host	You can disable on one host, leave the other host enabled, and then re-enable the disabled host.
	Primary and secondary hosts	Perform this operation on both hosts if you are deleting a volume set.
Enable the software.	Primary and secondary hosts	When enabling the software for the first time, issue the command from both hosts.
Perform full forward or reverse synchronization (copy).	Primary host	Ensure that both hosts are enabled.
Perform forward or reverse synchronization (update).	Primary host	Ensure that both hosts are enabled.
Set logging mode.	Primary host	Perform on the primary host only if a synchronization is in progress.
	Secondary host	Perform on the secondary host if the primary host failed.
	Primary or secondary host	Perform on either host if no synchronization is in progress
Toggle the autosynchronization state.	Primary host	
Update an I/O group.	Primary host	

Enabling and Disabling Volume Sets

Use the following commands to enable and disable volume sets.

Function	Command	See This Section
Enable the software for the specified set (sync not required).	sndradm -E	"Use the following commands to enable and disable volume sets." on page 76
Enable the software for the specified set (sync required).	sndradm -e	"Use the following commands to enable and disable volume sets." on page 76
Disable the software for the specified set.	sndradm -d	"Disabling Volume Sets" on page 78
Add or delete a point-in-time copy software volume group.	sndradm -I	"Adding and Deleting Point-in-Time Copy Software Volumes" on page 79
Stop the software replication, enter logging mode, and track changes using bitmaps.	sndradm -l	"Stopping Replication and Starting Logging" on page 81

Enabling Volume Sets

Note – When you first enable the remote mirror software, you must enable it on the primary and secondary hosts. Always maintain the same volume set field order on primary and secondary hosts.

The sndradm -e and -E commands enable the remote mirror volume set (*volset*) and start logging changes to the volumes. Also use these commands to create one–to–many, many-to-one, and multihop volume sets. See "One-to-Many, Many-to-One, and Multihop Volume Sets" on page 14.

Caution – When creating volume sets, do not create secondary or bitmap volumes using partitions that include cylinder 0. Data loss might occur. See "VTOC Information" on page 17.

When you first enable a set, the software assigns a default name of *shost*:*sdev*, where *shost* is the secondary host name and *sdev* is the secondary volume partition name, separated by a colon (:). The volume set name is referred to in this document as *set‑name*.

After executing these commands, you can use the *shost*:*sdev* name for a volume set each time you issue a command, instead of specifying the complete primary and secondary host, volume, and bitmap information for a volume set.

sndradm -e

This command sets the bitmap to indicate that a full synchronization between volumes is required. It also enables logging and adds the local volume set (including the local bitmap volume) to the storage volume (SV) driver control.

Syntax

sndradm –e [–n] {–f *volset‑file* | *volset*}

sndradm -E

Caution – Before using sndradm -E, ensure that the volumes are already synchronized by means other than the remote mirror software (for example, restoring the volumes from tape or other media). Otherwise, the secondary or target volume will contain inconsistent data.

This command clears the bitmap logging to indicate that the specified volumes are fully synchronized and enables logging. It also adds the local volume set (including the local bitmap volume) to the storage volume (SV) driver control.

Syntax

sndradm –E [–n] {–f *volset‑file* | *volset*}

After you enable a volume set, you can synchronize or update the volumes. See "Starting a Full Synchronization" on page 84 and "Starting an Update Resynchronization" on page 86 for more information.

Disabling Volume Sets

Use the `sndradm -d` command when the primary and secondary volumes no longer need to be associated with each other as software volume sets.

This command discards any active logging information in the bitmap volumes. See "Enabling and Disabling Volume Sets" on page 76 to re-enable software replication and "Starting a Full Synchronization" on page 84 to re-establish identical replicated sets.

Note – This command also removes the information for the specified set from the Sun StorEdge software configuration.

---

sndradm -d

When you issue this command, the software terminates all replication services between the specified primary and secondary volumes and discontinues the relationship between these volume sets. It also discards any active bitmaps that track temporary differences between primary and secondary volumes.

Syntax

sndradm –d [–g *io-groupname*] [–C *tag*] [–n] [–f *volset-file* | *volset* | *set-name*]

Adding and Deleting Point-in-Time Copy Software Volumes

Note – Use the sndradm -e command to enable software volume sets before using this command. "Enabling and Disabling Volume Sets" on page 76 describes this command.

Use the sndradm -I command to add or delete point-in-time copy software volumes for use with enabled software volume sets. You can use this command on the secondary and primary hosts to offer data protection during reverse and forward synchronization update operations. You also use this command to offer data protection during an automatic or manual resynchronization.

The sndradm -I a command creates configuration entries marked with the ndr_ii key. The ndr_ii entries contain an additional state field that the kernel uses to determine when point-in-time copies must be made. The kernel notifies the remote mirror software synchronization daemon whenever a synchronization is started and waits for sndrsyncd to perform any necessary copies before allowing the synchronization to proceed.

The daemon is also notified when any remote mirror software resynchronization starts or finishes executing. The daemon performs point-in-time copy operations on the secondary or target host, if configured to do so.

Consider the following when configuring a volume set:

- The master volume is the remote mirror secondary volume.
- The shadow and bitmap volumes cannot be a volume currently in use by the point-in-time copy software or remote mirror software. If the volumes are in use, data loss can occur as the result of this misconfiguration.
- Before the reverse synchronization occurs, the point-in time copy software enables the volume set as a dependent set. If the synchronization fails for any reason and the secondary volume data is lost or corrupted, you can restore the remote mirror secondary volume (the master volume in this set) by updating it with the shadow volume contents.

See also:

- "Choosing Automatic or Manual Resynchronization" on page 28
- "Using the Remote Mirror Software With the Point-in-Time Copy Software" on page 29
- "Autosynchronization" on page 30.

```
sndradm -I a
```

This command adds a point-in-time copy volume group entry to the Sun StorEdge configuration.

Syntax

```
sndradm -I a master–vol shadow–vol bitmap–vol
```

where *master–vol*, *shadow–vol*, and *bitmap–vol* are the full path to the raw device nodes for configured point-in-time copy volumes. In this case, the *master-vol* is the remote mirror secondary volume.

Caution – The shadow and bitmap volumes cannot be a volume currently in use by the point-in-time copy or remote mirror software. If the volumes are in use, data loss can occur as the result of this misconfiguration.

See the *Sun StorEdge Availability Suite 3.2 Point-in-Time Copy Software Administration and Operations Guide* for more information about these volume sets.

Note – The `/usr/opt/SUNWrdc/lib/sndrsyncd` daemon creates a matching point-in-time copy volume group entry before it performs an update synchronization after a network link or machine failure. After the update finishes executing, you can choose to delete this entry with the `/usr/opt/SUNWscm/sbin/iiadm -d` command. However, if you do not delete this entry, the daemon can use it in subsequent resynchronizations involving the point-in-time copy software.

```
sndradm -I d
```

This command deletes a point-in-time copy volume group entry from the Sun StorEdge configuration.

Syntax

```
sndradm -I d master–vol shadow–vol bitmap–vol
```

where *master–vol*, *shadow–vol*, and *bitmap–vol* are the full path to the raw device nodes for configured point-in-time copy volumes.

Stopping Replication and Starting Logging

Use the `sndradm -l` command to start bitmap logging on enabled volume sets and I/O groups. Issue the `sndradm -l` command as follows:

- **From the primary host if a synchronization is in progress.**
  The primary host is in logging mode and the secondary hosts are in "need sync" state. This state is cleared by either performing a synchronization or by updating the secondary volumes.

- **From the secondary host if the primary host or volume has failed.**
  When you place the secondary host into logging mode, the primary host continues in replication mode.

- **From either host if there is no replication occurring.**

To resume remote mirror software operations after using the `sndradm -l` command, use the `sndradm -u` command to perform an update resynchronization or the `sndradm -m` command to perform a full resynchronization. Note also that, when issued from the secondary host, the `sndradm -l` command does not work on the secondary volume for any volume that is currently synchronizing.

Tip – Place the secondary volume into logging mode and then mount any file systems in read-write mode to enable your application to write to the secondary volume. The remote mirror software logs the changes in the bitmap volume.

---

sndradm -l

This command enables you to stop replication between primary and secondary volumes and to start logging data changes on these volumes. The remote mirror software continues logging even if the operations of a volume set or I/O group are interrupted.

If all volume sets in an I/O group are replicating (meaning that the secondary volumes contain a valid copy of the corresponding primary volumes), all sets in the group enter logging mode automatically whenever one volume set enters logging mode. This scheme ensures that the secondary volumes contain a valid copy.

Syntax

`sndradm -l [-g io-groupname] [-C tag] [-n] [-f volset-file | volset | set-name]`

Synchronizing and Updating Volume Sets

The following commands enable you to synchronize and update volume sets.

See also and .

Function	Command	See This Section
Copy the entire contents of the primary volume to the secondary volume; also known as *full synchronization.*	`sndradm -m`	"Starting a Full Synchronization" on page 84
Copy the entire contents of the secondary volume to the primary volume; also known as *full reverse synchronization.*	`sndradm -m -r`	"Starting a Full Synchronization" on page 84
Update only the changed data of the primary volume to the secondary volume; also known as *update synchronization* or *resynchronization.*	`sndradm -u`	"Starting an Update Resynchronization" on page 86
Update only the changed data of the secondary volume to the primary volume; also known as reverse synchronization or reverse update.	`sndradm -u -r`	"Starting an Update Resynchronization" on page 86
Wait for a synchronization operation to finish executing before executing other commands.	`sndradm -w`	"Waiting for Synchronization to Finish Executing" on page 88

Restarting an Initial Synchronization

If, after enabling the volume sets, the initial synchronization operation using the `sndradm -m` or `sndradm -u` commands is interrupted, the software restarts the synchronization, according to the conditions described in the following table:

If You Enabled the Volume Sets With This Command	And Used This Command to Synchronize Volume Sets	The `sndradm -u` Command Results In
`sndradm -e`	`sndradm -m` `sndradm -m -r`	A full synchronization operation, starting from the beginning.
`sndradm -E`	`sndradm -m` `sndradm -m -r`	A synchronization restarts from the point of interruption.

Performing a Reverse Copy or Update After a Primary Volume or Network Link Failure

As specified in "Rehearsing Disaster Recovery" on page 56, do the following tasks after the network link or disk failure is repaired:

1. Place the primary and secondary host volumes into logging mode if they are not already in logging mode.
   See "Stopping Replication and Starting Logging" on page 81.

Note – If you check the volume status at this point using the `dsstat -m,sndr` command, the volume status might report `VF` (volume failed) or `RN` (reverse synchronization needed). When you perform the reverse copy or update, this volume status is cleared.

---

2. Perform the reverse copy or update.

Starting a Full Synchronization

Use the `sndradm -m` command when all of the following conditions exist:

- The primary volume has the desired contents.
- The contents of the primary and secondary volumes may be inconsistent.
- No logging information exists to incrementally resynchronize the volumes.
- You want to completely overwrite the contents of the secondary volume.

Use the `sndradm -m -r` command when all of the following conditions exist:

- The secondary volume has the desired contents.
- The contents of the primary and secondary volumes may be inconsistent.
- No logging information exists to incrementally resynchronize the volumes.
- You want to completely overwrite the contents of the primary volume.

You can use the primary volume during reverse synchronization
(`sndradm -m -r`). The primary volume shows a consistent volume image of the
latest data as soon as the reverse synchronization starts. If your application had been
writing to the secondary volume as part of a failure or disaster rehearsal, you can
move the application back to the primary volume when the reverse synchronization
starts. See also "Rehearsing Disaster Recovery" on page 56.

If a synchronization initiated by the `sndradm -m` or `sndradm -m -r` commands
is interrupted, use the `sndradm -u` update command to finish executing the
synchronization. You can use the `sndradm -p`, `sndradm -P`, and `dsstat` commands
to view volume set status.

See also "Restarting an Initial Synchronization" on page 83 and "Performing a
Reverse Copy or Update After a Primary Volume or Network Link Failure" on
page 83

`sndradm -m`

> This command enables you to start a full copy operation from the primary volume to the secondary volume. It also enables replication concurrently from the primary volume to the secondary volume so that any new writes to the primary volume are also replicated to the secondary volume.
>
> **Syntax**
>
> `sndradm -m` [`-g` *io-groupname*] [`-C` *tag*] [`-n`] [`-f` *volset-file* | *volset* | *set-name*]

`sndradm -m -r`

> This command enables you to start a full reverse copy operation from the secondary volume to the primary volume. It also enables replication from the primary volume to the secondary volume so that any new writes to the primary volume are also replicated to the secondary volume.
>
> **Syntax**
>
> `sndradm -m -r` [`-g` *io-groupname*] [`-C` *tag*] [`-n`] [`-f` *volset-file* | *volset* | *set-name*]

Starting an Update Resynchronization

Use the sndradm -u command when all of the following conditions exist:

- The primary volume has the desired contents.
- Replication was stopped by using the sndradm -l command or the network link was interrupted and bitmap logging was activated automatically.
- You want to overwrite those segments of the secondary volume that are different from the primary volume as defined in the bitmaps.
- You intend to start replication from the primary to the secondary volume.

Use the sndradm -u -r command when all of the following conditions exist:

- The secondary volume has the desired contents.
- Replication was stopped using the sndradm -l command or the network link was interrupted and logging was activated automatically.
- You want to overwrite those segments of the primary volume that are different from the secondary volume as defined in the bitmaps.
- You intend to start replication from the primary to the secondary volume.

Note – You can use the primary volume during reverse synchronization (sndradm -u -r). The primary volume shows a consistent volume image of the latest data as soon as the reverse synchronization starts. If your application had been writing to the secondary volume as part of a failure or disaster rehearsal, you can move the application back to the primary volume when the reverse synchronization starts. See also "Rehearsing Disaster Recovery" on page 56.

If a synchronization initiated by the sndradm -m or sndradm -m -r commands is interrupted, use the sndradm -u update command to finish executing the synchronization. You can use the sndradm -p, sndradm -P, and dsstat commands to view volume set status.

`sndradm -u`

> This command resynchronizes the secondary volume from the primary volume. It updates the secondary volume according to changes based on the bitmaps maintained while replication was stopped. It also enables concurrent replication between the primary and secondary volumes so that any new write operations to the primary volume are also replicated to the secondary volume.

> **Syntax**

> `sndradm -u` [`-g` *io-groupname*] [`-C` *tag*] [`-n`] [`-f` *volset-file* | *volset* | *set-name*]

`sndradm -u -r`

> This command resynchronizes the primary volume from the secondary volume. It updates the primary volume according to changes based on the bitmaps maintained while replication was stopped. It also enables concurrent replication between the primary volume and secondary volumes so that any new write operations to the primary volume are also replicated to the secondary volume.

> ---
> **Tip –** You can also use this command to roll back volume changes on the primary volume from a point-in-time copy software image captured on the stopped secondary volume.
> ---

> **Syntax**

> `sndradm -u -r` [`-g` *io-groupname*] [`-C` *tag*] [`-n`] [`-f` *volset-file* | *volset* | *set-name*]

Waiting for Synchronization to Finish Executing

Use the `sndradm -w` command in one of the following situations:

- When you need to be sure that the copy or update operation is finished executing before you use a primary application (like a database application) or another software command on the specified volume set or I/O group
- Before disabling a volume set, so that any write or copy operations can finish executing
- As part of a script executing commands

sndradm -w

This command enables you to make the software wait until an in-progress copy or update synchronization finishes executing. It prevents another software command from executing on the specified volume set or I/O group.

Syntax

sndradm -w [-g *io-groupname*] [-C *tag*] [-n] [-f *volset-file* | *volset* | *set-name*]

Reconfiguring or Modifying a Volume Set

> **Note –** Before reconfiguring or changing volume sets (except for changing the volume set bitmap), place the volume sets in logging mode by using the `sndradm -l` command.

The following commands enable you to reconfigure or modify a volume set.

Function	Command	See This Section
Change the bitmap of a volume set.	`sndradm -R b`	"Changing the Bitmap Volume in a Volume Set" on page 90
Move all the specified remote mirror sets to a different group.	`sndradm -R g`	"Moving a Volume Set to a Different I/O Group" on page 91
Change the replication mode of a volume set.	`sndradm -R m {sync|async}`	"Changing the Replication Mode of a Volume Set" on page 94
Update or reconfigure the software volume sets.	`sndradm -R -f` *volset-file*	"Updating the Volume Set Information" on page 96
Update or reconfigure the disk group name or cluster resource tag.	`sndradm -R C`	"Updating the Disk Cluster Tag Name" on page 97

Changing the Bitmap Volume in a Volume Set

Use the `sndradm -R b` command when you need to assign a new bitmap to a volume set.

sndradm -R b

This command enables you to assign a new bitmap to an existing volume set. You can change the bitmaps on the primary or secondary host. This command copies any data from the old bitmap to the new bitmap.

Syntax

sndradm -R b {p | s} *new-bitmap-name* [-C *tag*] [-n] [-f *volset–file* | *volset* | *set-name*]

- To change a primary host bitmap, issue the command from the primary host.
- To change a secondary host bitmap, issue the command from the secondary host.
- Optionally, to ensure consistency for status reporting, issue the command from both hosts. For example, to change a secondary host bitmap, issue the command from the secondary host first and then issue it from the primary host.

Moving a Volume Set to a Different I/O Group

Note – Before reconfiguring or changing volume sets (except for changing the volume set bitmap), place the volume sets in logging mode by using the `sndradm -l` command. After you finish changing the sets, perform an update synchronization using the `sndradm -u` command.

Use the `sndradm -R g` command to move volume sets to a different I/O group.

To move multiple sets to a single *io-groupname*, keep the *io-groupname* the same and specify different *set-name*s on the command line.

See the following procedures:

- "To Remove Volume Sets From an I/O Group" on page 92
- "To Move Volume Sets to Another I/O Group" on page 93
- "To Remove Volume Sets From an I/O Group and Change the Replication Mode" on page 95

sndradm -R g

This command enables you to move volume sets to a different I/O group, thereby updating an existing I/O group. You must specify at least one *set-name*. Issue this command from the primary and secondary hosts.

Syntax

`sndradm -R g` *io-groupname* [`-g` *io-groupname*] [`-C` *tag*] [`-n`] [`-f` *volset-file* | *volset* | *set-name*]

▼ To Remove Volume Sets From an I/O Group

1. **Place the I/O group volume sets into logging mode:**

   ```
   # sndradm -g io-groupname -l
   ```

2. **Remove volume sets:**
   - To remove a volume set from an I/O group, use quotation marks to specify a null group as follows:

   ```
   # sndradm -R g "" {volset | set-name}
   ```

   - To all volume sets from an I/O group:

   ```
   # sndradm -R g "" -g io-groupname
   ```

   - To remove selected sets from an I/O group, enter the volume set information in a file and use the −f *volset-file* option:

   ```
   # sndradm -R g "" -g io-groupname -f volset-file
   ```

3. **After you finish changing the sets, perform an update synchronization using the** sndradm −u **command.**

▼ To Move Volume Sets to Another I/O Group

1. Place the I/O group volume sets into logging mode:

```
# sndradm -g io-groupname -l
```

2. Move volume sets:

■ To move all volume sets in one I/O group to another I/O group:

```
# sndradm -g io-groupname -R g new-io-groupname
```

■ To move selected sets to an I/O group, enter the volume set information in a file
and use the −f *volset-file* option with these commands.

```
# sndradm -R g new-io-groupname -f volset-file
```

3. After you finish changing the sets, perform an update synchronization using the
sndradm −u **command.**

Changing the Replication Mode of a Volume Set

Note – Before reconfiguring or changing volume sets (except for changing the volume set bitmap), place the volume sets in logging mode by using the `sndradm -l` command. After you finish changing the sets, perform an update synchronization using the `sndradm -u` command.

Use the `sndradm -R m` command to change the replication mode of a volume set.

Note – You can remove one volume set, selected volume sets, or all volume sets from a group using the `sndradm -R g` command. See and

sndradm -R m

This command enables you to reconfigure the replication or mirroring mode of a volume set.

Syntax

`sndradm -R m {sync|async} [-C tag] [-n] [-f volset-file | volset | set-name]`

Considerations

- **Mixed mode I/O group** – Creating a mixed I/O group, where mirroring modes are asynchronous for one set and synchronous for another set, is not allowed.
- **Volume set requirements** – All volume sets in the I/O group must have the same primary host, secondary host, and mirroring mode.

▼ To Remove Volume Sets From an I/O Group and Change the Replication Mode

1. Place the I/O group volume sets into logging mode:

```
# sndradm -g io-groupname -l
```

2. Choose one of the following commands:

- Remove the volume set from its group:

```
# sndradm -R g ""{volset | set-name}
```

- To remove all volume sets from an I/O group:

```
# sndradm -R g "" -g io-groupname
```

3. Change the volume set replication mode of each volume set:

```
# sndradm -R m {sync|async} set-name
```

4. If desired, add the volume sets you have changed to a different or the previous I/O group:

```
# sndradm -R g io-groupname [-f volset-file | volset | set-name]
```

See also

5. After you finish changing the sets, perform an update synchronization using the `sndradm -u` **command.**

Updating the Volume Set Information

Use the `sndradm -R -f` command to change the current volume sets being operated on by the remote mirror software. This command is useful when you use a volume set file and have made changes to it.

The fields for the volume set file specified using the `-f` option are similar to those in the *volset*, described in TABLE 5-3:

phost pdev pbitmap shost sdev sbitmap ip {sync|async} [g *io-groupname*] [C *tag*] [q *qdev*]

See "Setting Up a Volume Set File" on page 43 and the `rdc.cf` man page for more information.

sndradm -R -f

This command enables you to update or reconfigure the current volume set information from a specified volume set file.

Syntax

`sndradm -R -f` *volset-file* [-g *io-groupname*] [-C *tag*] [-n]

Updating the Disk Cluster Tag Name

Note – The version 3.2 software is cluster-aware in the Sun Cluster 3.0 Update 3 and Sun Cluster 3.1 environments and provides high-availability for the Sun StorEdge software.

Use the `sndradm -R C` command in cases where the disk group name or cluster resource tag are not indicated by the volume path name. This command does not affect remote volumes and cannot be used in nonclustered environments.

You can also move all volume sets in one cluster tag to another cluster tag. Use the command as follows:

`sndradm -C` *tag* `-R C` *new-tag*

sndradm -R C

This command enables you to update or reconfigure the current disk group name or cluster resource tag of the local volume in a volume set. *tag* here is defined as the disk group name or cluster resource tag.

Syntax

`sndradm -R C` *tag* [`-g` *io-groupname*] [`-C` *tag*] [`-n`] [`-f` *volset-file* | *volset* | *set-name*]

Displaying Volume Set Information

Use the following commands to display information about volume sets.

Function	Command	See This Section
Display the command syntax information from the command prompt.	sndradm -h	See the display that results from entering this command at the command prompt
Display the software version information from the command prompt.	sndradm -v	See the display that results from entering this command at the command prompt
Display the software status.	sndradm -p sndradm -P	"Displaying Volume Set and I/O Group Status" on page 99
Display the software volume set and I/O group names.	sndradm -i	"Listing of Volume Sets and I/O Group Names" on page 100
Display the status of the link connecting systems running the software.	sndradm -H	"Displaying Link Status" on page 102

Displaying Volume Set and I/O Group Status

Use the `sndradm -p` and `sndradm -P` commands to display status information. Use the *volset* | *set-name* options to view the status for a single volume set. Use the `-g` *io-groupname* and `-C` *tag* options to filter the output so that it includes only those volumes that are part of *io-groupname* or *tag*.

Note – Issue this command from the primary host to display information correctly.

---

sndradm -p

This command displays a brief version of the remote mirror software status for volume sets and I/O groups. The following is example output of this command on the primary host:

```
# sndradm -p
/dev/rdsk/c4t96d0s1                      ->
fast7:/dev/rdsk/c2t0d0s1
/dev/rdsk/c4t97d0s1                      ->
fast7:/dev/rdsk/c2t0d0s1
/dev/rdsk/c4t98d0s1                      ->
fast7:/dev/rdsk/c2t0d0s1
```

Syntax

sndradm -p [*volset* | *set-name*] [-g *io-groupname*] [-C *tag*]

sndradm -P

This command displays a detailed version of the remote mirror software status for volume sets and I/O groups. The following is example output of this command on the primary host:

```
# sndradm -P

/dev/rdsk/c4t96d0s1                      ->
fast7:/dev/rdsk/c2t0d0s1
autosync: off, max q writes: 4194304, max q fbas: 16384, mode: sync

/dev/rdsk/c4t98d0s1                      ->
fast7:/dev/rdsk/c2t0d0s1
autosync: off, max q writes: 4194304, max q fbas: 16384, mode: sync
```

The reported status also includes information on the state of the set:

`volume failed`	The volume has failed.
`bitmap failed`	The bitmap has failed.
`disk queue failed`	The disk queue has failed.
`need sync`	Logging is in progress but an earlier synchronization did not complete. The secondary volume might be in an inconsistent state. Clear the "need sync" state by either performing a synchronization or by updating the secondary volume with a point-in-time copy to restore an earlier, valid snapshot.
`need rev sync`	Logging is in progress but an earlier reverse synchronization did not complete. The primary volume might be in an inconsistent state. Clear the "need rev sync" state by either performing a reverse synchronization or by updating the primary volume with a point-in-time copy to restore an earlier, valid snapshot.
`queuing`	Logging is in progress but data is still being added to the queue on the disk queue. An update synchronization clears this state.
`logging`	Data is not being replicated. Changes are being recorded to the scoreboard.
`reverse syncing`	A synchronization from the secondary volume to the primary volume is in progress.
`syncing`	A synchronization from the primary volume to the secondary volume is in progress.
`replicating`	Data written to the primary volume is being replicated to the secondary volume.

Syntax

sndradm -P [*volset* | *set-name*] [-g *io-groupname*] [-C *tag*]

Listing of Volume Sets and I/O Group Names

Use the sndradm -i command to display volume set and I/O group names. This command displays volume set information in a volume set file format. See Section , "Setting Up a Volume Set File" on page 3-43.

Tip – You can also use the output of this command to create a volume set file suitable for editing and using with the sndradm -R -f *volset-file* command.

```
sndradm -i
```

This command enables you to display a list of the current volume set and I/O group names in use. The following is an example of the output.

Note that the output is formatted as it would be entered in a *volset–file*:
phost pdev pbitmap shost sdev sbitmap `ip` {`sync`|`async`}.

```
# sndradm -i

fast7 /dev/rdsk/c2t0d0s1 /dev/rdsk/c2t1d0s0 fast8 /dev/rdsk/c4t96d0s1
/dev/rdsk/c6t0d0s4 ip sync

fast7 /dev/rdsk/c2t0d0s1 /dev/rdsk/c2t1d0s3 fast8 /dev/rdsk/c4t97d0s1
/dev/rdsk/c6t0d0s6 ip sync

fast7 /dev/rdsk/c2t0d0s1 /dev/rdsk/c2t1d0s4 fast8 /dev/rdsk/c4t98d0s1
/dev/rdsk/c6t0d0s8 ip async
```

Syntax

sndradm –i [*volset* | *set–name*] [–g *io–groupname*] [–C *tag*]

Use the –g *io–groupname* and –C *tag* options to filter the output so that it includes only those volumes that are part of *io-groupname* or *tag*.

Displaying Link Status

Use the sndradm -H command to check the status of the link connecting primary and secondary hosts.

Note – Use the ping(1M) command to supplement the link status shown by sndradm -H.

---

sndradm -H

This command displays the currently configured primary and secondary host machine names, volumes, and bitmap volumes for the selected sets. It also shows whether the link between machines is active or inactive. For example:

```
# sndradm -H atm-fred:/dev/vx/rdsk/freddg/sndr_vol01
Report SNDR link health? (Y/N) [N]: y

SNDR: atm-ethel /dev/rdsk/c3t9d0s3 /dev/rdsk/c6t0d0s4
atm-fred /dev/vx/rdsk/freddg/sndr_vol01 /dev/rdsk/c6t0d0s6
Inactive
```

Active indicates that replication or synchronization operations are occurring on the link used by the specified volume set or that all sets are in logging mode. Inactive indicates that the network link might be down.

Syntax

sndradm -H *shost*:*sdev*

where *shost* is the secondary host name and *sdev* is the secondary volume partition name, separated by a colon (:). When you first enable a set, the remote mirror software assigns a default volume set name of *shost*:*sdev*.

If you specify no arguments, the remote mirror software displays all configured volume sets. However, for best display results, specify a volume set using *shost:sdev*.

Enabling or Disabling Autosynchronization

Use the sndradm -a command to enable or disable autosynchronization. See "Choosing Automatic or Manual Resynchronization" on page 28 for more information about autosynchronization.

Note – The sndradm -P command does not report whether autosynchronization is enable or disabled for a set. It reports whether autosynchronization is active at the time the sndradm -P command is issued.

---

sndradm -a

This command enables or disables the remote mirror software autosynchronization. If remote mirror software autosynchronization is enabled on the primary host, the synchronization daemon attempts to resynchronize volume sets if the system reboots or link failures occur.

The default state is disabled.

Syntax

sndradm -a {on|off} [-g *io-groupname*] [-C *tag*] [-n] [-f *volset-file* | *volset* | *set-name*]

Using Autosynchronization with Sun Cluster

Consider the following when using autosynchronization with Sun Cluster:

- If you want automatic resynchronization to occur in the event of a cluster failover, turn on the autosync feature. With this feature enabled, any cluster failover will automatically put the remote mirror volume sets back into replication mode after an update occurs.
  However, in a two-node cluster with a one-to-one configuration and with autosynchronization enabled, an interrupted reverse synchronization does not restart automatically. Issue the following command to restart the reverse synchronization process:

    # sndradm -n -r -m

- If you want to manually force clusters to failover, make sure you allow all remote mirror components to fully recognize the condition, including the remote host that is not in the cluster. In practice, this means that you should not attempt an immediate update sync after the failover. You should wait at least thirty seconds after the completion of the scswitch command and before starting an update sync to allow time for Sun Cluster to complete its logical host interface reconfiguration.

Setting the Asynchronous Queue

These commands apply only to memory-based queues. They have no effect on disk queues.

Note – Ensure that you do not configure the queue size to be larger than the available cache. As a rule of thumb, reserve 5 percent of the total cache.

Use the following commands to set the maximum number of writes or 512-byte blocks that can be queued in asynchronous mode.

See "Tuning the Asynchronous Queue" on page 105.

sndradm -W

This command enables you to set the maximum number of write operations to queue for asynchronous mode volume sets. The default *value* number is 4194304. For example, set this value to 1000 to ensure that the secondary volume is never more than 1000 write operations behind the primary volume.

Syntax

sndradm -W *value* [-g *io-groupname*] [-C *tag*] [-n] [-f *volset-file* | *volset* | *set-name*]

sndradm -F

This command enables you to set the maximum number of 512-byte blocks in the asynchronous queue. The default number is 16384, allowing about 8 Mbytes of data per queue.

Syntax

sndradm -F *value* [-g *io-groupname*] [-C *tag*] [-n] [-f *volset-file* | *volset* | *set-name*]

Tuning the Asynchronous Queue

These commands apply only to memory-based queues. They have no effect on disk queues.

Note – Ensure that you do not configure the queue size to be larger than available cache. As a rule of thumb, reserve 5 percent of the total cache.

The queues containing the remote mirror software asynchronous write operations are stored in system kernel memory. You can tune or adjust these queues so that they are correctly sized for your remote mirror software use and volume data set size. Left untuned, the queues can appear to slow performance and consume more system memory than is desirable.

Consider the following, in order, when tuning asynchronous queues:

1. **Number of volume sets and whether a group contains a large number of volume sets in asynchronous mode (such a group shares a single queue and might affect performance)**

2. **Physical system memory**

3. **Network latency**

The number of volume sets is most important. Large queue settings matched with a small number of volume sets ensures better performance than large queue settings with a large number of volume sets. However, try not to size queues so large that you exhaust kernel memory and degrade your system's performance.

This section includes the following topics:

- "Asynchronous Mode and the Queue" on page 106
- "Sample Output For a Correctly-Sized Queue" on page 108
- "Sample Output for an Incorrectly-Sized Queue" on page 109

Asynchronous Mode and the Queue

The remote mirror software's asynchronous mode helps when your network has sufficient bandwidth to handle many write transactions or when your replicating application (such as a database) performs burst-writes at a rate higher than the network limits allow.

Generally, you use asynchronous mode for a volume set when the I/O write rate is less than the throughput of the network you are using for replication. This mode provides low latency rates to the software, resulting in faster response at the host (see "Asynchronous Replication Mode" on page 21).

As the number of write operations in the queue increases, the transactions can exceed the ability of the queue and network to process them (depending on network bandwidth and latency). The remote mirror software appears slow and reports less than optimal write transaction performance. Unless you tune (or size) the queue correctly, asynchronous mode performance might appear to be the same as or worse than synchronous mode, depending on the volume data set size.

Consider changing the software's default queue size if:

■ The amount of system memory cannot support the number of asynchronous mode volume sets in the queue.

■ The replicating application writes more data than the queue can store, affecting performance.

■ You do not want the application's write operations to become "out of sync" with the primary host, where pending write operations try to queue faster than the queue or network can handle them.

■ You do not care how "out of sync" the secondary host is from the primary host, you care only about response time on the primary host.

You can change the asynchronous queue size by using the `sndradm -W` and `sndradm -F` commands. The remote mirror software version 3.2 has set the following default sizes for the queue:

Default maximum number of write operations in the queue (default for the `sndradm  -W` command)	4194304
Default maximum number of 512-byte data blocks (default for the `sndradm  -F` command)	16384

▼ To Display the Current Queue Size

1. Type the following to display the current queue size:

```
# sndradm -P

/dev/vx/rdsk/rootdg/ds4-clone   ->      nws:/dev/vx/rdsk/rootdg/ds4-sndr-s
autosync: on, max q writes: 4194304, max q fbas: 16384, mode: async
```

The kstat command also displays queue information:

2. Perform any of the following kstat(1M) commands to show more information.

■ To show all sets, type:

```
# kstat sndr:::maxqfbas
```

■ To show the first instance (0) in the queue, type:

```
# kstat sndr:0::maxqfbas
```

■ To show more information, type:

```
# kstat sndr:::maxqitems
# kstat sndr:::async_throttle_delay
```

Sample Output For a Correctly-Sized Queue

The following kstat(1M) kernel statistics output shows information about the asynchronous queue. In this example, the queue is sized correctly.

```
# kstat sndr:0:setinfo
module: sndr                          instance: 0
name:   setinfo                       class:    storedge
        async_block_hwm               878
        async_item_hwm                483
        async_throttle_delay          0
        maxqfbas                      16384
        maxqitems                     4194304
        primary_host                  regina
        primary_vol                   rootdg/ds4-clone
```

This example shows only a portion of the command output needed for this section; the kstat command actually displays more information. The following table describes shows the settings and statistics shown in the example.

maxqfbas 16384	Default maximum number of 512-byte blocks in the asynchronous queue. Set this using sndradm -F
	The default setting allows about 8 Mbyte of data per queue.
maxqitems 4194304	Default maximum number of write operations to queue for asynchronous mode volume sets. Set this using sndradm -W.
	According to this setting, each write consumes 2 Kbyte for the 8 Mbyte of data.
async_block_hwm 878	A total of 878 512-bytes blocks (approximately 439 Kbytes) have been put into the queue.
async_item_hwm 483	A total of 483 write transactions have been put into the queue.[1]
async_throttle_delay 0	No delay in the queue. The queue is not yet full.

1 async_block_hwm and async_item_hwm show the maximum number that has been put into the queue since replication started. They do not show the current number in the queue.

Sample Output for an Incorrectly-Sized Queue

The following kstat(1M) kernel statistics output shows information about the asynchronous queue, which is incorrectly sized.

```
# kstat sndr:4:setinfo
module: sndr                              instance: 4
name:   setinfo                            class:    storedge
        async_block_hwm              16380
        async_item_hwm               2045
        async_throttle_delay         16497
        maxqfbas                     16384
        maxqitems                    4194304
        primary_host                 andrea
        primary_vol                  rootdg/ds-forall
```

This example shows only a portion of the command output needed for this section; the kstat command actually displays more information.

This example shows the default queue settings, but the application is writing more data than the queue can handle. The async_block_hwm value of 16380 indicates that the application is approaching the limit for 512-byte blocks. The next few I/O operations might not be placed into the queue.

The value of async_throttle_delay indicates that the application has been delayed 16497 x 2 milliseconds. In this example, if the system has sufficient memory, increase the size of maxqfbas using the sndradm -F command.

Managing Disk Queues

You can have memory-based or disk-based queues. Disk-based queues have some advantages over memory-based queues. See "Disk-Based Asynchronous Queues" on page 4 for further information. You can configure a disk queue when you first enable a set or you can use these commands later.

You can add, remove, and replace disk queues for sets and groups of sets with the disk queue CLI commands. You can also set each queue for blocking or non-blocking mode. See "Blocking Mode" on page 4 and "Nonblocking Mode" on page 4 for more information.

Note – The sndradm -q command is valid only in logging mode.

```
sndradm -q
```

This command enables you to add, remove, or replace a disk queue to a set or to a group.

Syntax

```
sndradm -g io-groupname -q a vol
```

```
sndradm -g io-groupname -q d
```

```
sndradm -g io-groupname -q r newvol
```

```
sndradm -q a vol shost:sdev
```

```
sndradm -q d shost:sdev
```

```
sndradm -q a newvol shost:sdev
```

For example, this is how to add a disk queue to a remote mirror set:

```
# sndradm -q a /dev/md/diskqdg/rdsk/d2 ns-81:/dev/md/rdsk/d221
```

```
sndradm -D
```

This command enables you to set the disk queue blocking mode for a set or group.

```
sndradm -D {block | noblock} set
```

For example, to change from blocking mode to nonblocking mode, type:

```
# sndradm -D noblock ns-81:/dev/md/rdsk/d221
Change SNDR tunable? (Y/N) [N]: y
```

Setting the Number of Asynchronous Threads

You can change the number of asynchronous threads from the default value of 2. By changing the number of threads to 1, you can simulate the behavior of Sun StorEdge Availability Suite 3.1 remote mirror software. You can set the number of threads to a higher number limited only by practical considerations.

```
sndradm -A
```

This command enables you to set the number of asynchronous threads that drain the asynchronous queues.

```
sndradm -A asyncthreads set
```

See "Multiple Asynchronous Flusher Threads" on page 5 for more details.

dsstat and Storage Cache Statistics

This section describes the dssat utility that enables you to gather statistics about the Sun StorEdge software and cache:

The /usr/opt/SUNWscm/sbin/dsstat command displays I/O statistics for the remote mirror, point-in-time copy, and cache software. See the dsstat man page for more information.

Unless you specify otherwise, all fields displayed show per-second averages based on the data collected during the interval. For example, if you specify 5-second intervals, the data displayed is a per-second average of the data collected during the last 5-second interval.

Syntax

dsstat –m *mode* [–r *report–options*][–d *display-options*][–s *volume-sets*] [–f |–F] [–z] [*interval* [*count*]]

See TABLE A-1 for the command options.

TABLE A-1 dsstat Options

Option	Description

–m *mode*	Specifies the statistics reporting mode, where *mode* is:
	• cache
	• ii
	• sndr
	This option determines the options you can use with the –r *report-option*. You cannot use the –r *report-option* if you specify –m cache.
	If –m is not specified, the default is for the statistics for all modes to be displayed.
	You can enter a comma-separated list of modes as well as multiple –m switches. The following commands are equivalent:
	`# dsstat -m sndr,ii`
	`# dsstat -m sndr -m ii`

–r *report–options*	No *report-options* are available for the –m cache mode. • If you specify –m ii, you can use this option as follows. If you do not specify –r *report-options*, the default display is –r msbo. -r m show master volume information -r s show shadow volume information -r b show bitmap volume information -r o show overflow volume information (if this volume type is attached) • If you specify –m sndr, you can use this option as follows. If you do not specify –r *report-options*, the default display is –r bn. -r b show bitmap volume information (this command shows only the bitmap of the host you are issuing the command from) -r n show network volume information
–d *display-options*	Specifies the types of information to show. You can specify one or more of the following options on the command line. (If you specify -d t, then the r and w options are ignored.) • If you specify –m cache, the available options are as follows. If you do not specify this option, the default display is -d sf. -d r detailed read information -d w detailed write information -d d destaged data information -d c cancelled write information -d s summary of information -d f cache behavior flags

	• If you specify –m sndr or –m ii, the available options are as follows. If you do not specify this option, the default display for -m ii is –d sf and for –m sndr is -d spf.
	–d r detailed read information
	–d w detailed write information
	–d t timing information
	–d s summary of information
	–d f volume type and status flags
	–d p (–m sndr only) percentage of volume requiring a synchronization
	–d a (–m sndr only) shows the secondary volume names
–s *volume-set*	Shows information about the specified *volume-set* only. *volume-set* can also be a comma-delimited list of volume names.
-f -F	Displays a field header for every report cycle. Displays a field header once, when reporting begins.
–z	Do not report or display lines that have a value of zero (indicates no activity)
interval [*count*]	*interval* specifies how often the report is updated. The default interval is one second. *count* specifies how many reports to display. If *count* is not specified, report output continues at the specified *interval* time until you interrupt it by a break (^C). If you do not specify *interval* and *count*, the command shows a single report for a one-second interval. Unless you specify otherwise, all fields displayed show per-second averages based on the data collected during the interval. For example, if you specify 5-second intervals, the data displayed is a per-second average of the data collected during the last 5-second interval.

Examples

TABLE A-2 describes the fields displayed.

- Report cache statistics (`dsstat -m cache`)
- Show detailed information for read and write statistics (`-d rw`)
- Show information for the volume `c1t35d0s6` (`-s /dev/rdsk/c1t35d0s6`)
- Generate reports at five-second intervals (5)

Note – Only the trailing sixteen characters of a volume name appear in a report. For example: `/dev/rdsk/c1t1d0s0` is displayed as `ev/rdsk/c1t1d0s0`.

```
# dsstat -m cache -d rw -s /dev/rdsk/c1t1d0s0 5

                       - read -              - write -
    volume          ckps    dkps    hit    ckps    dkps    hit
    ev/rdsk/c1t1d0s0    0       0   0.00       0       0   0.00
    ev/rdsk/c1t1d0s0    3    2396   0.13     983     763 100.00
    ev/rdsk/c1t1d0s0 2399     799  75.00    2815    2686 100.00
    ev/rdsk/c1t1d0s0 3200     800  80.00    2755    2908 100.00
    ev/rdsk/c1t1d0s0 3999     799  83.33    2809    2868 100.00
    ev/rdsk/c1t1d0s0 4800     800  85.71    2867    2931 100.00
```

- Report master and shadow volume point-in-time copy software statistics (`dsstat -m ii -r ms`)
- Generate reports at five-second intervals (5).

```
# dsstat -m ii -r ms 5

                         - master -        - shadow -
    set name        t  s    kps  tps  svt     kps  tps  svt
    ev/rdsk/c0t1d0s0 I  -      0    0    0       0    0    0
    ev/rdsk/c0t1d0s0 I  -   9047  219    3    9040  219    7
    ev/rdsk/c0t1d0s0 I  -  13548  317    3    9760  243    6
    ev/rdsk/c0t1d0s0 I  -   5946  155    3    9684  227    8
    ev/rdsk/c0t1d0s0 I  -  16539  417    2    9242  225    7
```

- Reports network statistics for the remote mirror software (-m sndr, -r n)
- Shows information for the secondary volume c1t35d0s6
  (-s /dev/rdsk/c1t35d0s6)
- Shows detailed information for read and write statistics, timing, and volume type
  or status flags (-d rwtf)
- Generates reports at five-second intervals (5)

```
# dsstat -m sndr -r n -d rwtpf -s /dev/rdsk/c1t1d0s0 5


                                    - network -
    set name         t  s      sn   rkps rtps   wkps wtps   svt
    ev/rdsk/c1t1d0s0  P  L   77.27      0    0      0    0     0
    ev/rdsk/c1t1d0s0  P SY   75.07      0    0   2317   72    11
    ev/rdsk/c1t1d0s0  P SY   71.67      0    0   3443  108     9
    ev/rdsk/c1t1d0s0  P SY   69.37      0    0   2426   76    12
    ev/rdsk/c1t1d0s0  P SY   66.68      0    0   2765   86    11
```

TABLE A-2   `dsstat` Display Fields

Fields	Description
set name	Volume set specified by -s option
t	Volume type for remote mirror volumes: P primary host volume S secondary host volume Volume type for the point-in-time copy volumes: I independent shadow volume set D dependent shadow volume set Valid cache modes C cache reads and writes D disk reads and writes
s	Volume status for the remote mirror software: L volume is in logging mode; changes are being logged R replicating changes to secondary SY forward synchronization is in progress RS reverse synchronization is in progress SN forward synchronization needed RN reverse synchronization needed VF volume failed BF bitmap failed Volume status for the point-in-time copy volumes: C copy in progress – no copy in progress
sn	Percentage of volume requiring sync
rtps	Number of reads
rkps	Kilobytes read
wtps	Number of writes
wkps	Kilobyes written
svt	Service time per operation
tps	Sum of rtps and wtps
kps	Sum of rkps and wkps
r	Cache read behavior

TABLE A-2 `dsstat` Display Fields

Fields	Description
w	Cache write behavior
ckps	Kilobytes read from cache
dkps	Kilobytes read from disk
hit	Read hits during interval
ds/s	Kilobytes destaged from cache
cn/s	Number of write cancellations

Glossary

asynchronous replication
(n.) A form of replication in which the remote mirror software confirms to the originating host that the primary I/O transaction is complete before updating the remote image. That is, completion of the I/O transaction is acknowledged to the host when the local write operation is finished and the remote write operation has been queued. Deferring the secondary copy removes long-distance propagation delays from the I/O response time.

configuration location
(n.) Network location where the Sun StorEdge Availability Suite software stores configuration information about all enabled volumes used by the software.

forward resynchronization
See Update synchronization.

full synchronization
(n.) A complete volume-to-volume copy operation, which is the most time-consuming of the synchronization operations. In most cases, a secondary volume is synchronized from its source primary volume. However, restoration of a failed primary disk might require reverse synchronization, using the surviving remote mirror as the source.

logging mode
(n.) Mode in which a bitmap tracks write operations to a disk rather than maintaining a running log of each I/O event. This method tracks disk updates that have not been remotely copied while the remote service is interrupted or impaired. The blocks that no longer match their remote sets are identified for each source volume. The software uses this log to re-establish a remote mirror through an optimized update synchronization rather than a complete volume-to-volume copy.

primary host, primary volume
(n.) The system or volume on which the host application is principally dependent. For example, this is where the production database is being accessed. This data is replicated to the secondary host by the software. Also known as the "local host" or "local volume."

Replication	(n.) The process of maintaining the same data on both the primary and secondary volumes. Replication is driven by user-layer application write operations so replication is an ongoing process.
Reverse synchronization	(n.) An operation during which the primary volume is updated with the contents of the secondary volume.
Secondary host, secondary volume	(n.) The remote counterpart of the primary host or primary volume, which data copies are written to and read from. Remote copies are transmitted without host intervention between peer servers. A server might act as primary storage for some volumes and secondary storage for others. Also known as "remote host" or "remote volume."
Synchronization	(n.) The process of establishing an identical copy of a source disk onto a target disk as a precondition for software mirroring.
Synchronous replication	(n.) Replication that is limited to short distances (tens of kilometers) because of the detrimental effect of propagation delay on I/O response times.
Update synchronization	(n.) Synchronization that copies only those disk blocks identified by logging, reducing the time required to restore remotely mirrored sets.
Volume set file	(n.) A text file containing information about specific volume sets. This text file is not the same as the configuration location, which contains information about all configured volume sets used by the remote mirror and point-in-time copy software.

Index