Best Practices for the Sun StorEdge™ 6920 System

Version 3.0

Sun Microsystems, Inc.
www.sun.com

Part No. 819-0122-10
October 2005

Submit comments about this document at: `http://www.sun.com/hwdocs/feedback`

Contents

Figures

Tables

Preface

This *Best Practices for the Sun StorEdge 6920 System* provides information about optimizing Sun StorEdge 6920 data services features to support the vision and mission of your organization. Refer to this guide after installation of your system is complete and you have completed the basic configuration tasks described in the *Sun StorEdge 6920 System Getting Started Guide.*

Audience

This *Best Practices for the Sun StorEdge 6920 System* is designed to support you in a variety of ways–as a guide to configuring your system to optimize features for your business practices, as a reference for concepts and recommended processes and procedures, as a tool for learning how system features can benefit your organization, and more.

To fully use the information in this document, you should be trained and certified by Sun to install and service Sun StorEdge products, and you should be familiar and experienced with these subjects:

- Storage area network (SAN) implementation and configuration
- Sun StorEdge 6020 and Sun StorEdge 6120 storage arrays
- Solaris™ Operating System (Solaris OS)
- Operating systems that operate with the Solaris OS (Microsoft Windows, Linux, and other UNIX® operating systems)
- Multipathing and failover
- Storage virtualization
- Command-line interfaces (CLIs)
- Browser interfaces

You should also be familiar with the physical and virtual elements of the Sun StorEdge 6920 system described in the *Sun StorEdge 6920 System Getting Started Guide* and in the Sun StorEdge 6920 Configuration Service online help.

How This Document Is Organized

This book contains the following chapters:

Chapter 1 gives a brief overview of the Sun StorEdge 6920 system, describes general system administration, provides licensing information, and describes key features.

Chapter 2 focuses on the logical components of the system and provides storage design information to consider before configuring the Sun StorEdge 6920 system, including provisioning storage, global spares, and hot-spares; multipathing and redundancy; and the Sun StorEdge Remote Response service.

Chapter 3 defines volumes as they are used by the Sun StorEdge 6920 system, and describes prerequisites for volume creation procedures.

Chapter 4 discusses mirroring features, including actions you can perform on a mirror, prerequisites for mirror creation, and how to manage mirror components.

Chapter 5 explains snapshot features, best practices for optimizing snapshot reserve space, using snapshots for backups, performing snapshot rollbacks, and other tasks.

Chapter 6 describes data replication architecture, how to plan and manage replication tasks, and examples for synchronous and asynchronous replication.

Chapter 7 discusses external storage, data migration, and how to manage a variety of external storage tasks.

Chapter 8 describes the Storage Automated Diagnostic Environment software component of the Sun StorEdge 6920 system, setting site information for diagnostics and monitoring, and using alarms and events features.

Chapter 9 provides information to help you maintain and tune the performance of your system, including online transaction processing (OLTP), upgradable components, and bottleneck solutions.

Appendix A lists the maximum values for elements of the Sun StorEdge 6920 system.

Typographic Conventions

Typeface*	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	Read Chapter 6 in the *User's Guide*. These are called *class* options. You *must* be superuser to do this.
	Command-line variable; replace with a real name or value	To delete a file, type `rm` *filename*.

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

Related Documentation

The following table lists the documentation for the Sun StorEdge 6920 system and related products.

Product	Title	Part Number
Late-breaking news	• *Sun StorEdge 6920 System 1.0 Release Notes*	819-0120
Sun StorEdge 6920 system series information	• *Sun StorEdge 6920 System Getting Started Guide*	819-0117
	• *Sun StorEdge 6920 System Site Preparation Guide*	819-0118
	• *Sun StorEdge 6920 Hardware Quick Setup Guide (poster)*	817-5226
	• *Sun StorEdge 6920 System Administration Guide for the Browser Interface*	819-0123
	• *Sun StorEdge 6920 Regulatory and Safety Compliance Manual*	819-0119
Best Practices, version 2.0	• *White Paper – Best Practices for the Sun StorEdge 6920 System*	817-6985
Sun StorEdge 6120 and 6020 arrays	• *Sun StorEdge 6120 Array Start Here*	817-0198
	• *Sun StorEdge 6120 Array Release Notes*	817-0201
	• *Sun StorEdge 6120 Array Site Preparation Guide*	817-0960
	• *Sun StorEdge 6120 Array Installation Guide*	817-0199
	• *Sun StorEdge 6020 and 6120 Arrays System Manual*	817-0200
Sun StorEdge Expansion Cabinet	• *Sun StorEdge Expansion Cabinet Installation and Service Manual*	805-3067
Remote monitoring services	• *Storage Automated Diagnostic Environment 2.4 Release Notes*	819-0432
Sun StorEdge network FC switch-8 and switch-16 switches	• *Sun StorEdge SAN Foundation Software 4.4 Installation Guide*	817-3671
	• *Sun StorEdge SAN Foundation Software 4.4 Configuration Guide*	817-3672
	• *Sun StorEdge SAN Foundation Software 4.4 Release Notes*	817-3673
3COM Ethernet hubs	• *SuperStack 3 Baseline Hub 12-Port TP User Guide*	3C16440A
	• *SuperStack 3 Baseline Hub 24-Port TP User Guide*	3C16441A
SANbox-8/16 Segmented Loop FC Switch	• *SANbox-8/16 Segmented Loop Fibre Channel Switch Management User's Manual*	875-3060
	• *SANbox-8 Segmented Loop Fibre Channel Switch Installer's/User's Manual*	875-1881
	• *SANbox-16 Segmented Loop Fibre Channel Switch Installer's/User's Manual*	875-3059

Accessing Sun Documentation

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

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

For Sun StorEdge 6920 system documentation, go to:

`http://www.sun.com/products-n-solutions/hardware/docs/Network_Storage_solutions/Midrange/6920/index.html`

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 this document with your feedback:

Best Practices for the Sun StorEdge 6920 System, part number 819-0122-10.

CHAPTER 1

Introduction to the Sun StorEdge 6920 System

Sun Microsystems is a leader in delivering next-generation solutions that address companies' plans for tiered storage and more robust business continuity. The features of the Sun StorEdge 6920 system provide a solid foundation for deploying highly scalable, manageable, and reliable tiered storage for your enterprise. The purpose of this guide is to help you implement Sun StorEdge 6920 system capabilities, to support the vision and business goals of your organization.

This chapter contains the following sections:

Architectural Overview

The Sun StorEdge 6920 system has a modular architecture with integrated system-wide manageability. The Sun StorEdge 6920 system features include:

- Storage virtualization and pooling to simplify storage management.
- Scaling and aggregation of both capacity and performance. System capacity scales from 504 gigabytes to 65 terabytes and capacity can be added while the system is online.
- Centralized management and monitoring through a browser interface and remote scripting client.
- High availability and failover capability through redundant hardware components.
- Serviceability of components that can be replaced while the system is online.
- Diagnostic monitoring to enhance reliability, availability, and serviceability (RAS) of the system through the Storage Automated Diagnostic Environment.
- Remote monitoring, troubleshooting, and servicing by Sun-trained personnel through Sun StorEdge Remote Response service.
- Virtualization of heterogeneous external storage to facilitate data migration from existing storage devices.
- Centralized data services including Sun StorEdge Storage Pool Manager, Sun StorEdge Data Snapshot, Sun StorEdge Data Mirror, and Sun StorEdge Data Replicator software.

With the Sun StorEdge 6920 system at the center of the storage network, all storage functions can be served from a single storage pool, easing the overall administrative burden. The system not only assumes the critical functions of external network and storage appliances, but also integrates intelligent services, such as storage virtualization, snapshot, remote replication, and secure storage provisioning, to the storage network.

The Sun StorEdge 6920 system enables you to manage two types of storage:

- Internal storage, provided by the arrays within the base cabinet (managed by the system)
- External storage, provided by the external storage devices (externally managed by customers)

FIGURE 1-1 illustrates the Sun StorEdge 6920 system environment.

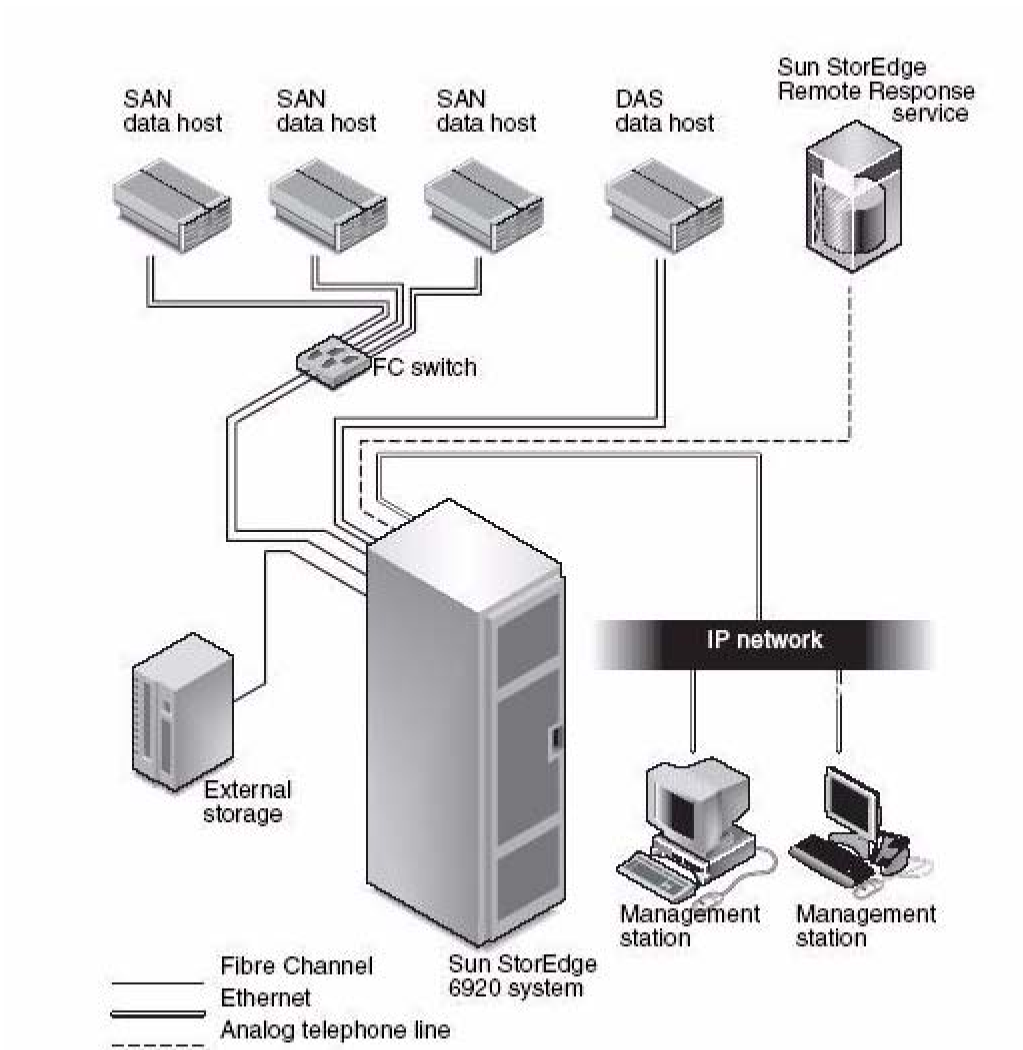

FIGURE 1-1 Sun StorEdge 6920 System Environment

When you have completed the basic configuration tasks described in the *Sun StorEdge 6920 System Getting Started Guide*, your system has the following characteristics:

- The Ethernet connection is cabled to a host or a gateway on the local area network (LAN).
- A service console is connected.
- The system is powered on.
- The IP address and netmask are configured.
- Initial setup of the system has been completed.
- The DEFAULT storage domain uses one storage pool named Default.
- All storage ports are assigned to the Default storage pool.

If you are upgrading an existing system, the configuration is dependent on the system's prior configuration history. The Default storage profile, which the Default storage pool uses, has the specifications described in TABLE 1-1.

TABLE 1-1 Default Storage Profile Specifications

Storage Attribute	Value
RAID level	RAID-5
Segment size	64 KB
Dedicated hot-spare	Not enabled
Read-ahead mode	Enabled
Number of drives	Variable
Array type	Best available match - bandwidth
Virtualization strategy	Stripe
Stripe size	1 Mbyte

Your system ships from the factory with the configuration described in TABLE 1-2.

TABLE 1-2 Default Storage Configuration

Default Configuration	Description	Configuration Options
Storage domains	All storage elements are in one storage domain (Default), which can be accessed by all hosts (initiators). This storage domain contains the Default storage pool with its associated initiators and virtual disks.	You can choose either of the following options: • use the existing default configuration (as is) • create a custom storage domain
Storage profiles	The Default storage pool uses the default profile to specify its storage parameters.	You can choose any of the following options: • use the existing default configuration (as is) • select one of the predefined storage profiles shipped with the system • create a custom storage profile **Note:** You cannot delete or modify any of the factory-provided profiles, including the Default profile, nor can you delete or modify a profile that is in use and associated with a storage pool.
Storage pools	All virtual disks are in one storage pool (Default).	You can choose either of the following options: • use the existing default configuration (as is) • create a new storage pool
Virtual disks	Virtual disks are preconfigured at the factory, based on the number of drives ordered for your system. The virtual disks are configured as RAID-5. If a tray has 7 disks, the factory creates one 5+1 RAID-5 virtual disk and configures one global hot spare. If a tray ships with 14 disks, the factory configures one 5+1 RAID-5 virtual disk, one 6+1 RAID-5 virtual disk, and one global hot spare	You can choose any of the following options: • use the existing default configuration (as is) • move an existing, unused virtual disk to a different storage pool with similar attributes • delete the existing RAID-5 virtual disk and reconfigure the recovered disks into a new virtual disk in a storage pool with a different storage profile

You will learn more about configuring these virtual components of the system in "Storage Design Considerations" on page 15.

Data Services Platform (DSP)

At a high level, the Data Services Platform (DSP) of the Sun StorEdge 6920 system separates the physical structure of the storage devices (such as arrays and virtual disks) from the presentation layer (LUNs presented to initiators). By having the physical layout separate from the presentation layout, storage and system administrators are able to work with the storage more effectively. This model also enables the deployment of advanced storage area network (SAN) storage and data management applications (for example, point-in-time data copies).

The DSP is a high-performance, six-slot, carrier-class, storage-utility controller component. It employs a crossbar switch architecture and enables storage administrators to create scalable, reliable, highly manageable, and cost-effective storage networks by providing the following:

- A virtualization engine that manages attached storage resources
- Virtual volumes that serve as disk devices to hosts

In the Sun StorEdge 6920 system, the DSP is included in the cabinet configuration, as shown in FIGURE 1-2.

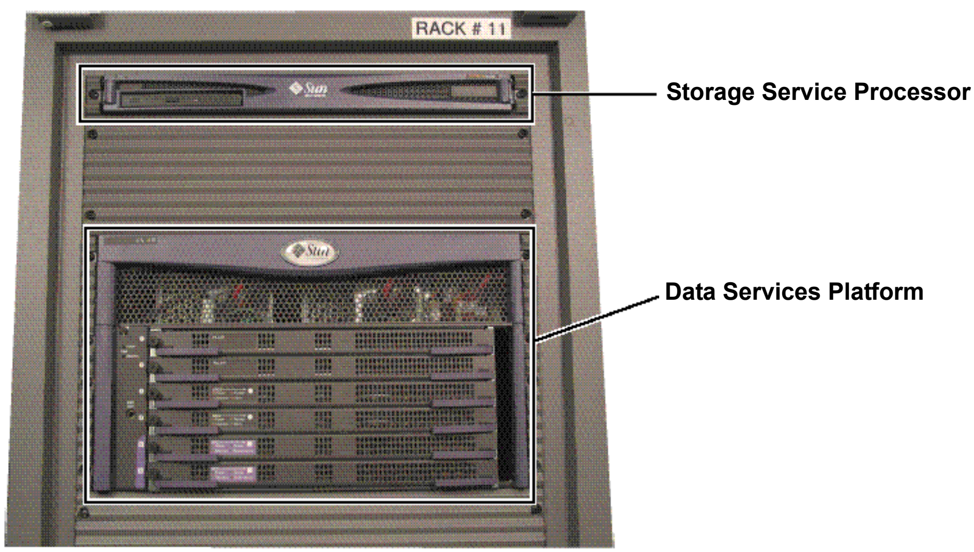

FIGURE 1-2 Data Services Platform

Much of the information in this guide pertains to the Sun StorEdge 6920 system virtualization services that you manage through the DSP to optimize the functionality of volumes, snapshots, mirrors, external storage, data replication, and more.

System Administration

The system is managed through two basic interfaces:

- A remote scripting command-line interface (CLI) client
- A browser interface

Remote Scripting CLI Client

The remote scripting CLI client communicates with the Storage Service Processor over the out-of-band Ethernet connection.

For you to use the CLI, the remote scripting client must be installed on the host. Once installed, the CLI provides two capabilities:

- An interactive command console for configuring and managing the device
- A secure scripting capability for building scripts to manage system and application interaction

The client communicates with the system over SSL, so all the commands that travel across the LAN are encrypted. For security and system integrity reasons, the user cannot access the CLIs of the system components directly, nor use Telnet to access the Storage Service Processor.

The commands run in a shell on the selected host or hosts. The program that runs the client is invoked from `/opt/se6x20/cli/bin/sscs`. The `sscs` command enables the user to manage all aspects of the system.

For Windows clients, the script is located in `Program Files/Sun Microsystems/SSCS/bin`.

Note – The remote scripting CLI client for the Solaris Operating System is provided on the Host Installation Software CD that is delivered with the Sun StorEdge 6920 system, installed here: `/opt/se6x20/cli/man`. Host software for other operating systems is available from the Sun Download Center (SDLC). For information about installing the remote scripting CLI client, see the *Sun StorEdge 6920 System Getting Started Guide*. For information about managing the system using the CLI, see the `sscs`(1M) man page.

Browser Interface

The browser interface for the Sun StorEdge 6920 system is designed to mask the complexity that is typically involved in configuring a storage system. The browser interface is invoked by launching of a browser that points to the IP address assigned to the Storage Service Processor. The sessions are encrypted through SSL. The browser interface also includes extensive event management capabilities including the Sun StorEdge 6920 Configuration Service management software as well as an embedded version of the Sun Storage Automated Diagnostic Environment monitoring and diagnostic software.

Supported Operating Systems

The following operating systems are supported by the Sun StorEdge 6920 system:

- Solaris 8 Operating System (update 4) or higher
- AIX 5.2
- Hewlett-Packard HP-UX 11.0 and 11.i
- IBM AIX 5.1 (64-bit)
- Microsoft Windows 2000 Server and Advanced Server SP4
- Microsoft Windows 2003 Server Standard, Enterprise, and Web Editions
- Novell Netware 6.5
- Red Hat Enterprise Linux 3.0
- SuSe Linux 8
- VMware 2.5.1

Supported browsers are the following:

- Mozilla 1.2.1 and above
- Netscape Navigator™ 7.0 and above
- Microsoft Internet Explorer™ 5.0 and above

Sun StorEdge 6920 System Features

Your Sun StorEdge 6920 system might include only some of the features described in this guide, depending upon your particular license agreements. For features not pertaining to your organization at this time, this guide can serve as a reference should you become interested in the future.

This section provides an overview of licensing and describes three key features of the Sun StorEdge 6920 system:

- Mirrors
- Snapshots
- Data replication

More information about concepts and best practice applications is provided in feature-specific chapters later in this guide.

Feature Licensing

Before you can use premium data services features on your system, you must obtain a feature license key and register feature licenses for each feature that you plan to use. When you purchase a feature license, you receive a feature license key that entitles you to use that feature. These feature licenses are capacity-based licenses. If necessary, you can increase the capacity of the feature license at a later time.

Displaying a Feature License Key

To display a feature license key for a registered feature:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Administration > Licensing.

   The Feature License Summary page is displayed.

2. Click the name of the feature whose feature license key you want to see.

   The License Details page for the selected feature is displayed.

3. See the License Key field for the feature license key.

For details about licenses for the features described in this guide, see the online help.

For information about purchasing feature licenses, contact your Sun sales representative.

Displaying Feature License Compliance

To display summary and detailed information about features that require feature licenses, and to audit feature license compliance of your system:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Administration > Licensing.

   The Feature License Summary page is displayed.

2. Click the name of the feature for which you want to see feature license information.

   The License Details page for the selected feature is displayed.

Mirrors

A mirror is a special type of volume consisting of up to four separate yet equal mirror components that you can access independently and use to track changes to and update your data set.

Mirroring can be an important part of your data migration strategy. Suppose you want to migrate data from legacy to internally managed storage on the Sun StorEdge 6920 system. You can mirror the data on a legacy volume to a normal volume, during which time the system can continue to access the legacy volume. When mirroring is complete, the mirror components are resilvered (synchronized) so that the legacy data also exists as an independent copy on the new mirrored volume. You can access your data on the new mirrored volume, so you might choose to delete the legacy volume.

Suppose you want to add a fourth mirror component to your three-component mirror, as part of a strategy in which you rotate mirror components. Each day you add a mirror component, wait for it to resilver so that each mirror component is identical to the other, and then perform a split operation to save a picture of the data at the time of the split.

Another scenario in which you might want to add a mirror component to a mirror is when you have performed a force break operation and need to add a new mirror component to replace the one you that you removed from the mirror.

For more information, see Chapter 4.

Snapshots

A snapshot is an instantaneous copy of volume data at a specific point in time. Snapshots reside in the same storage domain as the volume for which they are taken, called the parent volume.

Sun StorEdge Data Snapshot software enables you to create a dependent copy (snapshot) of a parent volume. You can work with a snapshot as you would with any other volume, with the exceptions that you cannot take a snapshot of a snapshot, mirror a snapshot, or replicate a snapshot. Each snapshot can be accessed independently by other applications. Available to servers on the storage area network (SAN) as a unique volume, a snapshot always retains an association with the parent volume.

You can use snapshots for the following purposes:

- To reduce downtime for backup operations

  Rather than using online data, back up your snapshots to enable critical transactions to continue during the backup process. Mount the snapshot to the backup server, and then back up the snapshot's data to tape.

- To perform data analysis and test applications with actual, current data

  Use the most recent snapshot for data analysis or testing, instead of working with operating data that might interfere with critical transactions.

- To restart applications from the snapshot

  If an application problem causes questionable data to be written to the parent volume, restart the application with the last known good snapshot until it is convenient to perform a full recovery or rollback.

The system stores snapshots in an area associated with the parent volume called snapshot reserve space. Snapshot reserve space can be a simple partition or a concatenated group of virtual disks. For each parent volume that will contain snapshots, only one area of snapshot reserve space is needed.

You determine the size of the snapshot reserve space appropriate for your business practice, based on two factors: the number of snapshots that will be taken of the parent volume, and the anticipated percentage of changed data on the parent volume for each snapshot.

For more information, see Chapter 5.

Data Replication

The Sun StorEdge Data Replicator software offers a volume-level replication mechanism to protect your data.You can use this software to replicate volumes between physically separate primary and secondary Sun StorEdge 6920 systems. The software is active while your applications access data volumes, and it continuously replicates data from the primary to the secondary site.

Data replication enables you to maintain a copy of data that is physically separate from the original. The Sun StorEdge 6920 system replication provides remote replication capability, such that a copy of volumes on one Sun StorEdge 6920 system can be maintained on a second Sun StorEdge 6920 system.

The Sun StorEdge Data Replicator software provides the following features:

- Replication based on Fibre Channel (FC) or Transmission Control Protocol/Internet Protocol (TCP/IP).
- Support for synchronous and asynchronous replication of up to 128 volumes.
- Potential for configuration of multiple volumes in a consistency group, enabling preservation of write ordering between remote associated volumes. This is for applications, such as databases, that typically spread data structures across multiple volumes.
- Role reversal

  When a primary site goes down, the secondary site assumes the role of the authoritative copy.

  For more information, see Chapter 6.

Getting Started With Configuration

Your system is shipped with a default configuration that simplifies storage provisioning. If you decide to use the default configuration that shipped with the system, you must do the following:

- Create volumes within the Default storage pool.
- Map each volume to one or more initiators. Typically, you will map to at least two initiators.

If you decide the default configuration does not meet your storage needs, you can customize it for your organization by doing any of the following:

- Create one or more storage domains

  For example, you might want to create a separate domain to isolate financial data and ensure that only specified hosts can access it.

- Create one or more storage pools

  For example, if you plan to create a legacy volume from external storage, you might want to create a custom legacy pool using the Legacy profile, rather than use the default legacy pool provided by the system.

- Move a virtual disk to another pool, or delete a virtual disk and create a new virtual disk with another configuration

  For example, although the virtual disks are configured as RAID-5 by default, you might want to use a RAID-1 profile. You would need to delete the virtual disk and create a new storage pool with a different profile.

- Access storage and user data that resides on an external storage array or system

  For example, you might want to migrate user data from an external storage device to internal storage on the system.

The information that follows in this book provides guidelines to help you optimize Sun StorEdge 6920 system functionality as you configure and then maintain your system for your organization's success.

Note – For detailed procedures and information on pages that are displayed in the browser interface, see the online help.

CHAPTER 2

Storage Design Considerations

There are a number of factors to consider before configuring your Sun StorEdge 6920 system. This chapter will help you understand how the system is designed so you can configure all elements for optimal performance, to support your organization's business needs.

This chapter contains the following sections:

Storage Design

In the default storage configuration, all data hosts are in one storage domain and all available storage is in one storage pool. Any host (initiator) in the domain can access any storage in the pool. However, your organization's needs determine how you customize the storage configuration. Instead of assigning hosts to specific physical storage, you provision the storage for your organization by creating domains consisting of initiators and pools. The system has physical and logical storage elements that you use to provision your storage. The relationships among the physical and logical storage elements is shown in FIGURE 2-1.

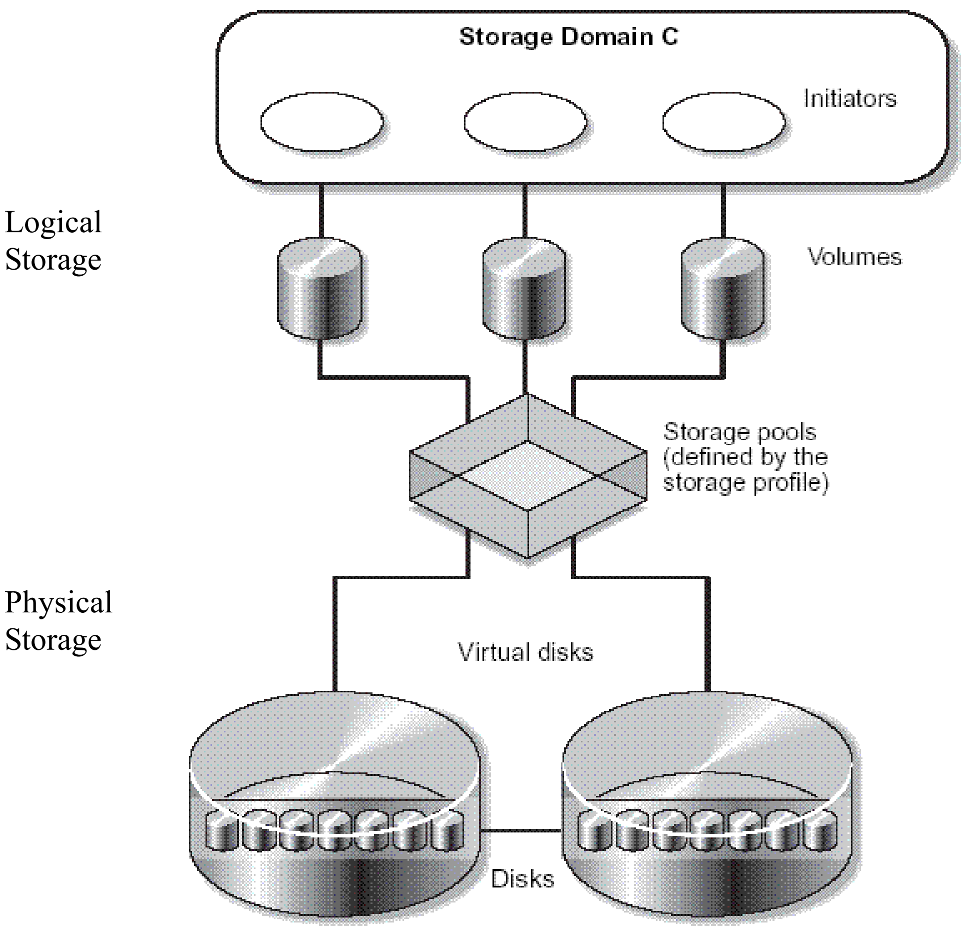

FIGURE 2-1 Relationship Between Logical and Physical Storage

Physical storage elements include arrays internal to the base system or expansion cabinet, external storage, and the associated storage trays, and disks. Logical storage elements include domains, profiles, pools, virtual disks, volumes, snapshots, and replication sets.

Storage Domains

One unique design element of the Sun StorEdge 6920 system is the concept of storage domains. A storage domains is a "container" of physical devices and virtual (logical) storage elements managed by the system. A storage domain includes a virtual (internal) network, connections to physical devices, and policies that govern the storage domain.

Storage domains isolate data so that data from one host or set of hosts is protected from access by other hosts. In its default configuration, the system is provisioned with one storage domain, called DEFAULT. This storage domain contains the Default storage pool and its associated virtual disks.

To group servers together and provision storage for that group of servers only, you redefine the DEFAULT storage domain into two or more storage domains. A configuration with multiple storage domains provides an environment in which multiple servers share storage without compromising client security, as shown in FIGURE 2-2.

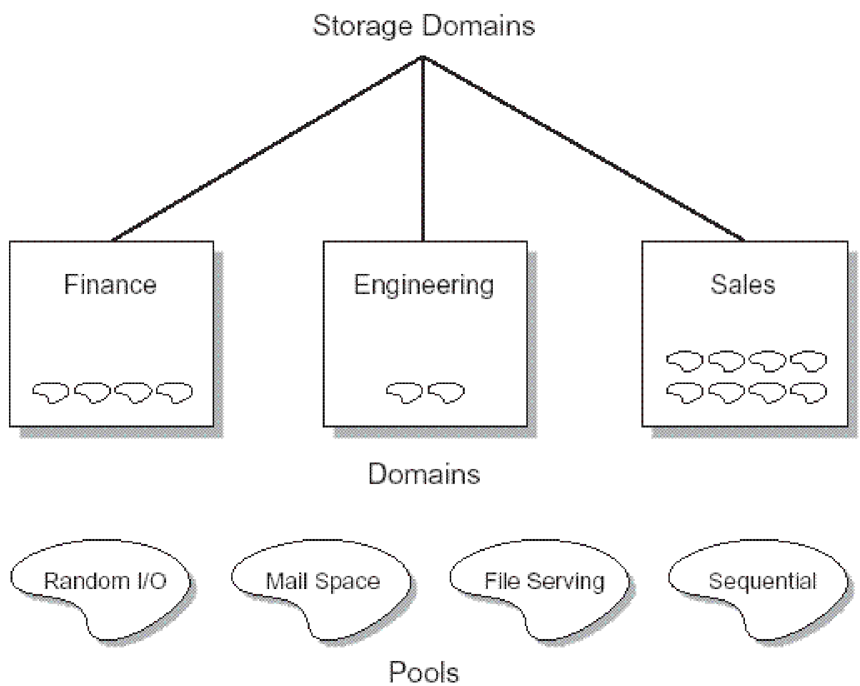

FIGURE 2-2 Storage Domains

You can create a maximum of 14 storage domains within one system. Within a storage domain, you can have multiple storage pools. However, each storage pool can be associated with only one storage domain.

For example, suppose the Finance domain contains the Random I/O and Mail Space pools, the Engineering domain contains the File Serving pool, and the Sales domain contains the Sequential pool. In this case, the Finance domain cannot be associated with the File Serving or Sequential pool, the Engineering domain cannot be associated with the Random I/O, Mail Space, or Sequential pool, and the Sales domain cannot be associated with the Random I/O, Mail Space, or File Serving pool.

Storage Pools

A storage pool is created from a storage profile, whether it is a predefined, factory-delivered storage profile or a customer-created storage profile. A storage pool can consist of one or more virtual disks. When you create a storage pool, you will select or create a storage profile that matches the requirements for your application. You may use one of the factory defined profiles or create one that matches the specific needs of your application.

The Sun StorEdge Storage Pool Manager enables virtualization and pooling of storage assets across applications. Licensing for up to 2 terabytes of capacity is provided with the system. Additional right-to-use (RTU) licensing is required for use of larger capacities.

Key to the Sun StorEdge 6920 system enterprise solution, the Storage Pool Manager is the focal point for managing all the storage capacity (internal and external) controlled by Data Services Platform (DSP) on the Sun StorEdge 6920 system. Rather than managing individual logical unit numbers (LUNs) or an expanding pool of very large disks, the Storage Pool Manager enables general IT administrators, not just storage experts, to manage pools of storage, each with unique attributes relative to size, performance, cost, or reliability, depending on the needs of the application. For example, your company might create three pools of storage:

- **Tier 1 storage** for a high-end online transaction processing (OLTP) application that resides on high-performance 15Krpm FC drives internal to the Sun StorEdge 6920 system
- **Tier 2 storage** for email on larger-capacity 10Krpm FC drives resident on both the Sun StorEdge 6920 system and an older back-end attached array
- **Tier 3 storage** for long-term email and file archiving on a back-end attached high-density array with serial SATA drives

IT managers can specify configuration-tuning parameters tied to each of these pools based on desired application workload characteristics. To make this capability more quickly useful to companies, the Sun StorEdge 6920 system is delivered with 15 pre-tested profiles for widely used business applications from vendors such as Oracle and SAP as well as for specific workloads such as OLTP and High-Performance Computing (HPC).

Another important feature of Storage Pool Manager is support for legacy volumes. With this capability, you can non-disruptively attach legacy arrays through the Sun StorEdge 6920 system without having to reconfigure or rename existing storage area network (SAN) volume names on the legacy arrays or on connected servers. At the same time, those arrays now have access to the advanced heterogeneous data replication functions on the Sun StorEdge 6920 system. The combination of legacy volumes and Sun StorEdge Mirroring software enables you to efficiently and nondisruptively migrate data from older arrays to new arrays.

Planning Storage Pools

The number of storage pools you need for a storage domain depends on your array configuration. Use care in determining your storage pool and virtual disk utilization. Disks are assigned to virtual disks which are then assigned to pools.

An internal storage tray can have a maximum of two virtual disks. Therefore, if seven disks in a tray are assigned to a single storage pool, and you add one disk and configure it for a second storage pool, you cannot later add six more disks to the empty disk slots without reconfiguring the virtual disk in the storage pool. In this situation, you cannot use the disks as virtual disks; however, you can use them as array hot-spares.

Before adding a new storage pool, determine the following:

- The storage domain with which you will associate the pool
- The storage profile with which you will associate the pool

  The attributes of the storage profile determine the storage characteristics of the virtual disks that you can configure for the new storage pool.

Note – You create virtual disks from storage trays (by adding storage to the pool) or from external (raw) storage.

- The number of disks and the target size for the pool, if you want to add storage to the pool when you create it

  The system provides a list of available capacities, based on the configuration method and target size you choose, from which you specify one of the following: blocks, kilobytes, megabytes, gigabytes, or terabytes.

Reconfiguring a storage pool consists of removing the storage pool and then creating a new storage pool. Storage pools can be expanded dynamically, so you can remove unused virtual disks and add new virtual disks.

Note – You cannot delete a storage pool until you delete all volumes created from space in the pool. If a pool is deleted that has virtual disks, those virtual disks will then be unassigned and will be available to be assigned to other pools that have profiles that match the characteristics of the unassigned virtual disks.

Note – You cannot move a storage pool into another storage domain.

Before you can allocate storage appropriately, consider the following requirements for your site:

- **Security** – By creating additional storage domains, you segregate initiators and storage. For example, the hosts that handle financial data store their data in a different domain from the domain used by hosts that handle research data.
- **I/O** – The Default storage profile specifies a general, balanced access to storage, but some parts of your organization might require an emphasis on frequent, small I/O, while other parts of the organization might require an emphasis on very large I/O. The system provides a set of profiles to meet various needs. You can also create custom profiles.
- **Performance** – In most cases, each host (server) will be connected to at least two ports to provide redundancy. Many applications do not require nearly the full bandwidth of a pair of ports, and in a SAN environment it is appropriate to connect more than one server to a pair of ports. To maintain performance, you can expand the number of ports by adding storage resource card (SRC) sets. For information about how to add an SRC set, log in to the management software, go to the Sun Java Web Console page, and click Sun Storage Automated Diagnostic Environment > Service Advisor > X-Options.

To increase performance beyond what a balanced configuration provides, you can configure the system to include more host ports if the system supports an OLTP application. Keep in mind that configuring more host ports that are required to support the application workload will not help performance.

For more information about OLTP, see "Online Transaction Processing (OLTP)" on page 187.

External Storage Devices

The Sun StorEdge 6920 system provides support for virtualization of data stored on external storage devices. This data can be preserved and added to the system as a legacy volume. You can also use external storage as a means of increasing storage capacity. In this case, the data on the external storage device is not preserved, and the capacity is added to a storage pool as one virtual disk per external volume.

For more information, see Chapter 7.

Storage Profiles

A storage profile is a set of storage allocation attributes that you select when creating a storage pool, instead of setting each attribute individually. The system has a predefined set of storage profiles, including the Default storage profile. You can choose a storage profile suitable for the application that is using the storage, or you can create a custom storage profile.

TABLE 2-1 describes the parameters for storage profile settings.

TABLE 2-1 Storage Profile Settings

Parameter	Value or Variable Type	Description
Name	Up to 32 characters	Unique identifier for the storage profile.
Description	Up to 256 characters	Description of the storage profile.
RAID Level	The RAID level of the virtual disk: • RAID-0 • RAID-1 • RAID-5	RAID level configured across all disks within a virtual disk.
Segment Size	4 KB, 8 KB, 16 KB, 32 KB, or 64 KB	The size of the array segment in kilobytes: 4 KB, 8 KB, 16 KB, 32 KB, or 64 KB. This is an array attribute and must be the same for all virtual disks within an array. Once a virtual disk exists on any of the array's trays (there can be only 0, 1, or 2 virtual disks on a tray), then all the virtual disks on that array must have the same segment size as that virtual disk. Be careful when allocating segment size to a profile, as you cannot have pools whose profiles differ in segment size reside in the same 6020 array. Use the same array. If the pools' profiles do not have the same segment size, you can configure a system with an array that has unused disks in which no virtual disks can be configured.

TABLE 2-1 Storage Profile Settings *(Continued)*

Parameter	Value or Variable Type	Description
Dedicated Hot-Spare	Enabled or Disabled	Indication of whether each virtual disk requires an additional disk from the same tray to be used as its dedicated hot-spare. The system has dedicated hot-spares (controlled by this attribute) and array hot-spares (controlled as an editable attribute on the Array Details page). • Dedicated hot-spare – There is a one-to-one correspondence between a dedicated hot-spare and a virtual disk (that is, the hot-spare is dedicated to a specific virtual disk). This profile attribute controls whether a dedicated hot-spare disk is assigned to the virtual disk when a virtual disk is created in a pool with this profile. • Array hot-spare – This is a disk in an array that can act as a hot-spare for any disk in the array. The number of drives and dedicated hot-spare attributes taken together define how many disks are consumed for each virtual disk in a pool.
Read Ahead	On (Enabled) or Off (Disabled)	Read-ahead mode of the array. This array attribute must be the same for all virtual disks within an array. Once a virtual disk exists on any of an array's trays (you can have only 0, 1, or 2 virtual disks on a tray), all the virtual disks on that array must have a Read-Ahead Enabled setting that matches that virtual disk's Read-Ahead Enabled setting. You must plan the use of the Read-Ahead Enabled setting carefully, because you cannot have pools whose profiles have different Read-Ahead Enabled settings on the same array. You can configure a system with unused disks in which no virtual disks can be configured if the pools' profiles do not match the Read-Ahead Enabled setting.
# Drives	Variable, 2 through 14	The number of drives associated with this profile. This is the number of disk drives that are used to construct the virtual disk in a pool using this profile. The number of drives can be 2 to 14, or Variable. If it is set to variable, you must specify the number when you create virtual disks (that is, when you add storage to the pool).

TABLE 2-1 Storage Profile Settings *(Continued)*

Parameter	Value or Variable Type	Description
Array Type	Exact Match: IOPS (2x4 array) Exact Match: Bandwidth (2x2 array) Exact Match: Capacity (2x6 array) Best Available Match: IOPS Best Available Match: Bandwidth Best Available Match: Capacity	The Exact Match options must be used only with the specified array configuration (2x2, 2x4, and 2x6) so, for a single system, only one of the Exact Match options is appropriate. If a profile defines an Exact Match option but is applied to the wrong array type, you cannot add storage to that pool. The Best Available options attempt to match the array type based on IOPS (input/output performance speed of transactions per second), bandwidth, or capacity. The following rules apply: Bandwidth: 2x2 > 2x4 > 2x6 Capacity: 2x6 > 2x4 > 2x2 IOPS: 2x4 > 2x2 > 2x6
Virtualization Strategy	Concatenate or Stripe	Strategy to be used for all volumes allocated from a storage pool. The Concatenate value: • Requires a minimum of one virtual disk • Is used in sequential environments • Implements algorithms to create volumes from “best fit” virtual disks to minimize overall partitioning of virtual disks The Stripe value: • Requires a minimum of two virtual disks • Is used in random environments • Stripes across multiple partitions in volumes
Stripe Size	128 kilobytes, 512 kilobytes, 1 megabyte, 2 megabytes, 4 megabytes, 8 megabytes, or 16 megabytes	The amount of space on a virtual disk to be allocated for the stripe used in the Stripe virtualization strategy. This field is valid when the Virtualization parameter is set to Stripe.

Standard Storage Profiles

TABLE 2-2 describes the storage profiles included in the Sun StorEdge 6920 system.

TABLE 2-2 Standard Storage Profiles in the Sun StorEdge 6920 System

Profile Name	RAID Level	Segment Size	Read Ahead Mode	Number of Drives	Array Type	Dedicated Hot-Spare	Virtualization	Stripe Size
Default	RAID-5	64 KB	On	Variable	Best Available Match: Bandwidth (2x2 > 2x4 > 2x6)	No	Stripe	1 MB
Sequential	RAID-5	64 KB	On	Variable	Exact Match: Bandwidth (2x2 array)	No	Concatenate	n/a
Random 5	RAID-5	64 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	2 MB
Random 1	RAID-1	64 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	2 MB
High Performance Computing	RAID-5	64 KB	On	Variable	Exact Match: Bandwidth (2x2 array)	No	Concatenate	n/a
Mail Spool	RAID-1	64 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	2 MB
NFS Striping	RAID-5	64 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	2 MB

TABLE 2-2 Standard Storage Profiles in the Sun StorEdge 6920 System *(Continued)*

Profile Name	RAID Level	Segment Size	Read Ahead Mode	Number of Drives	Array Type	Dedicated Hot-Spare	Virtualization	Stripe Size
NFS Mirroring	RAID-1	64 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	2 MB
Oracle OLTP	RAID-5	32 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	1 MB
Oracle OLTP HA	RAID-1	32 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	1 MB
Oracle DSS	RAID-5	64 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	2 MB
Sybase OLTP	RAID-5	32 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	1 MB
Sybase OLTP HA	RAID-1	32 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	1 MB
Sybase DSS	RAID-5	64 KB	On	Variable	Best Available Match: IOPS (2x4 > 2x2 > 2x6)	No	Stripe	2 MB

Multipathing and Redundancy

Multipathing is a design for redundancy that provides at least two physical paths to a target. Redundancy is duplication for the purpose of achieving fault tolerance. Redundancy refers to duplication or addition of components, data, and functions within the array.

This section discusses both of these topics as they relate to storage design considerations.

Multipathing Design Considerations

Host-based software is required for multipathing functionality to provide both redundancy and load-balancing. All hosts connected to the Sun StorEdge 6920 system must use one type of multipathing software. It is a best practice to install and enable Sun StorEdge Traffic Manager software on all servers connected (directly attached or via a SAN) to the Sun StorEdge 6920 system. Both Sun StorEdge Traffic Manager software and VERITAS Dynamic Multipathing (DMP) software can be loaded on the same host server. However, when they are, Sun StorEdge Traffic Manager software provides the multipathing functionality while VERITAS DMP software remains idle.

Multipath States

Sun StorEdge Traffic Manager assigns physical paths to one of two classes, primary or secondary. When hosts are directly attached to arrays, the primary class is assigned to the path attached to the array controller that is the primary owner of the LUN. The secondary class is assigned to the path attached to the other array controller. While Sun StorEdge Traffic Manager supports both primary and secondary paths, all paths between the Sun StorEdge 6920 system and initiators or host bus adapters (HBAs) are assigned to the primary class.

All these paths are found in the ONLINE state. The driver provides transitions through a number of distinct path states for both primary and secondary paths. The following states in the Sun StorEdge Traffic Manager are significant to the Sun StorEdge 6920 system:

- ONLINE – The path information node has been attached to a client device and is available for use by the system. This is the active path.
- OFFLINE – The path information node is offline. The path is not available for use by the system.

High-Availability Connectivity

To achieve high availability (HA) for this system, each of the following is recommended:

- Hosts connecting to the system must provide dual paths to the system for redundancy.
- Each host must use multipathing software for protection if a failure along one of the paths in the Sun StorEdge 6920 system.
- Dual paths from the host must be routed to two different storage I/O cards to avoid single points of failure.

Storage Array Configurations

The storage array design of the Sun StorEdge 6920 system is modular, with a variety of possible configurations. Each configuration option includes two controllers (also referred to as a controller pair) to provide redundancy and failover capabilities. Each storage array also has redundant FC data paths and four power supplies with an integral battery backup system. In the event of a total power failure, each array has sufficient power from the batteries to shut down the system in an orderly fashion. The system is available in three supported storage array configurations as described in TABLE 2-3.

TABLE 2-3 Storage Array Configuration Options

Option	Controllers x Trays	Number of Disk Drives	Minimum Capacity	Maximum Capacity
1	2 x 2	14 to 28	504 gigabytes	4 terabytes
2	2 x 4	28 to 56	1 terabyte	8 terabytes
3	2 x 6	42 to 84	1.5 terabytes	12 terabytes

Redundancy Configurations

The purpose of redundancy configuration is duplication in order to achieve fault tolerance. It refers to duplication or addition of components, data, and functions within the array.

RAID-0

RAID-0 is striping without parity or mirroring protection. Data is distributed evenly at the block level among disks for performance. No redundancy is provided, and the loss of a single disk causes the loss of data on all disks. As a best practice, this level should be used for high-speed streaming of large file reads (for example, video) of noncritical data that is easily available elsewhere within the organization.

RAID-1

RAID-1, or mirroring, addresses the potential for disk failure by writing every byte of data that is on one disk to a second disk. The Sun StorEdge 6920 system implements parallel writes for up to three disks. If one disk fails, the second disk continues to operate without interruption. The primary advantages of RAID-1 are data protection when a disk fails and increased read performance, given that the read I/O workload is spread across multiple disks.

RAID-5

When large storage capacities are needed, mirroring everything can become cost-prohibitive. Thus, it is best practice to strike a balance between the performance of striping and the redundancy of mirroring. RAID-5 solves this problem by combining disk striping and distributed parity data. Both data and parity information are evenly striped across the drives. Because of parity, if a single drive fails, data can be reconstructed from the remaining drives.

Two disk drive failures cause all data to be lost. Both data and parity are distributed evenly across all the disks in the array at the block level, so no single disk can compromise the integrity of the data. Not only does RAID-5 perform well, but it also yields good disk economy while maintaining a high level of data protection.

Combining RAID-0 and RAID-1

Although RAID-5 is a cost-effective way to provide redundancy in an array of disks, it offers little protection against multiple drive failures. For very crucial corporate data, this risk might be unacceptable. To combine high performance and high levels of redundancy, a combination of RAID levels is necessary. RAID-0's data striping is a cost-effective way to create high levels of performance in a disk array, and having multiple copies of data is the best way to create redundancy. By combining disk mirroring (RAID-1) with disk striping (RAID-0), you can take advantage of both features.

Choosing a Virtualization Strategy

Because of striping's lack of redundancy, it is best practice to use a virtualization strategy of striping only from storage pools containing RAID-1+0 or RAID-5 virtual disks for random I/O workloads. If you use striping in conjunction with storage pools of RAID-0 virtual disks, do so only for high-speed streaming of large file reads of noncritical data (like video) that is easily available elsewhere within the organization. Physical disk failures in a RAID-0 virtual disk cause loss of all data on the virtual disk. In the case of a stripe across multiple RAID-0 virtual disks, the failure of a single physical disk would cause the loss of all volumes created by striping across those vdisks.

Concatenation is used for applications with highly sequential workloads. It minimizes partitioning of virtual disks and involves as many virtual disks as possible. There are a large class of applications (databases, mail servers) that generate a combination of random and sequential workloads. Both databases and mail applications (specialized databases) have intent logging mechanisms that write sequentially to log files. The overall performance of these applications is in many cases limited by the rate that the application can write to the log file. Segregating this type of workload on separate physical disks is often an important part of maximizing performance. The best practice for these types of applications can be to create concat RAID 1+0 pools for log files and create striped RAID-5 pools for the rest of the application data that more typically sees random I/O patterns.

Concatenation is good for sequential workloads because those workloads require fewer physical disks to support the peak throughput, as compared with random workloads that are limited by the time it takes to move the disk heads. Random workloads need to be spread across as many physical disks as possible to support peak throughput from a volume.

For more information, see "Virtualization Strategies" on page 41.

Array Hot-Spares and Dedicated Hot-Spares

A hot-spare is a spare disk within an array that is used for failover when any of the active disks fail. The Sun StorEdge 6020 array supports two types of hot-spares: array hot-spares and dedicated hot-spares.

With array hot-spare capability, one or more disk drives are configured as part of an array hot-spare pool, and the array hot-spare pool serves all virtual disks in the array. This can reduce the amount of storage used for hot-spare capability.

With dedicated hot-spare capability, a disk drive serves as a hot-spare for a designated virtual disk and can serve only one virtual disk.

A maximum of eight array hot-spares can be combined to form an array hot-spare pool. An array hot-spare pool can be shared among all trays associated with a controller pair. A hot-spare, whether an array hot-spare or a dedicated hot-spare, must have a capacity greater than or equal to the largest capacity of any drive that it is serving.

Array Hot-Spare Pool

You can add drives to or delete drives from an array hot-spare pool at any time. Drives do not need to be allocated to an array hot-spare pool. Although an array hot-spare pool comes preconfigured from the factory, you can reconfigure these array hot-spare drives if you want a different configuration. You can also add hot-spare drives to an array hot-spare pool, up to a maximum of 8 per array.

Drives added to an array hot-spare pool are assigned from the right-most available drives. In addition, array hot-spares are allocated from expansion trays before they are allocated from controller trays. That is, the hot-spares should be balanced across slot 14 of each tray in the array before using any slot 13 of an array; the array hot-spare pool should attempt to span all expansion and controller trays before two array hot-spares reside in one tray.

Dedicated Hot-Spares

Dedicated hot-spare drives do not have to be used within the Sun StorEdge 6920 system. They do not need to reside in the same tray as the virtual disk and should have a capacity greater than or equal to the largest drive capacity of the virtual disk. In general, dedicated hot-spares are employed only in cases in which the user needs to ensure that a drive is available for the RAID pool when needed and when expense is not a concern. Dedicated hot-spares are not preconfigured from the factory.

Simultaneous Use of Dedicated and Array Hot-Spares

Both array hot-spare and dedicated hot-spare drives can be used simultaneously, which enhances high availability of the Sun StorEdge 6920 system at the expense of storage for data. If a drive in a virtual disk fails, and the virtual disk is associated with both an array hot-spare pool and a dedicated hot-spare, the dedicated hot-spare drive is used first. If the dedicated hot-spare is unavailable, a drive from the array hot-spare pool is then used.

Nonfloating Hot-Spare

A nonfloating hot-spare is a drive that reverts back to hot-spare status after a failed drive has been replaced and the mirror resilvered (synchronized) or the stripe rebuilt. The far right drive bay of each tray can optionally be used for a nonfloating hot-spare.

Dedicated Hot-Spare Drives and Virtual Disks

The maximum number of virtual disks per tray is two. Thus, controller x tray configurations of 2x2, 2x4, and 2x6 can contain a maximum of 4, 8, and 12 virtual disks, respectively. Virtual disks must consist of whole drives and must be contiguous within the tray. Drives must be filled from left to right, with the exception of hot-spares, which are placed in slot 14.

Backups

The growing amount of data in today's enterprise environment can be attributed to the Internet, faster computers, more powerful applications, and the conversion from analog to digital data. Data research companies suggest that the amount of data in an average enterprise will at least triple in the next two to three years. It is not uncommon today to have 10 terabytes of data on local file systems or distributed throughout an enterprise that must be backed up or restored quickly by backup applications.

In addition, a global economy does not allow data to be offline during nights or weekends for creation of full backups. Corporations cannot wait many hours or days to be back online and to continue business as usual in case of a disaster or hardware failure. This limits the time period available for backup and recovery and creates an even bigger challenge for current backup and recovery strategies.

Tape Library Implementation

The Tape Library Implementation service delivers a working Sun StorEdge robotic tape library backup system consisting of both hardware and backup and monitoring software components. This provides a platform that you can use to develop and implement your production backup and recovery policies.

Because backup of large amounts of data is time consuming, many companies can afford to make only one copy of data to tape. This makes protected data vulnerable to media errors, tape failures, and loss in a disaster recovery. To solve this issue, companies are making multiple copies of backup tapes, called backup clones. A backup clone can be created only from the original backup copy and is usually moved off-site into a vault. In case of a catastrophe and loss of the original backup media, the backup clones can be used to recover data. However, making clones multiplies the time it takes to create the additional backup sets and adds tremendous media cost and administration overhead.

Data Verification

Data verification means verifying that data moved to B is the same as the data at A. In a migration using a tape backup and restore methodology, verification is relatively simple, since the data does not change during data migration. Checksums or CRC checks are usually used to verify the integrity of the data volumes. Data transfer becomes more complex if data sets are part of a database or customer application using multiple data sets. Care must be taken that the integrity of the whole set remains intact. The process becomes even more complicated if data migration takes place while the data sets are being used, because there are no real points of comparison. In a data migration scenario in which the Sun StorEdge 6920 system is used, the point at which the legacy volumes are seen by the hosts can be taken as the point for verification.

Enterprise Strategy

The current enterprise backup strategy enables disparate enterprise-wide network clients to use a network to automatically move backup data to a tape drive connected to a backup server. Automated tape libraries with multiple tape drives allow multiple backup streams to be received from multiple backup clients in parallel. In a server backup implementation, data is backed up from the local file systems directly to local tape devices.

A common backup strategy is to deploy one or more backup servers throughout the enterprise. Backup clients are deployed on every system or workstation in the enterprise that needs backup. It is not uncommon for a company to have two to three backup servers and hundreds of backup clients. On a regular schedule, the backup clients send data, utilizing an incremental or full backup strategy, through a TCP/IP network directly to a tape drive connected to one of the backup servers.

Incremental Backup

To protect and restore files and file versions between full backup cycles, you might deploy an incremental backup strategy. With incremental backup, only new or changed files are copied to secondary storage devices. This strategy enables you to back up and protect newly created or changed files more efficiently, without having to complete a full backup of all files every day.

The incremental backup is usually performed on a daily basis, at times of no user activity. Each incremental backup creates its own daily restoration point for new and changed files. For an enterprise-wide backup, incremental backup also minimizes network traffic during the normal work week, because only the changed and new files must be moved through the backup network. After daily incremental backups are created during the normal work week, a new full backup is performed on the weekend.

For information on using snapshots for backups, see Chapter 5.

Sun StorEdge Remote Response Service

The Sun StorEdge Remote Response service enables Sun-trained personnel to remotely monitor, troubleshoot, diagnose, and service your Sun StorEdge 6920 system 24 hours a day, 7 days a week. This service offers fast turnaround when the monitoring and diagnostic software resident on the Storage Service Processor detects a serviceable action and subsequently dials up a Sun Solution Center. The software transmits pertinent information that enables Sun-trained personnel to resolve the problem, often without involving the customer or affecting data availability.

Note – The Sun StorEdge Remote Response service does not give Sun-trained personnel access to customer data stored within the system. Only system configuration information is available through the SSRR.

The Sun StorEdge Remote Response service is offered by Sun Services and is a no-charge service for products covered by a warranty or a SunSpectrum contract. The Sun StorEdge 6920 system ships complete with all hardware and agent software ready to enable activation. All that is required is connection to an analog phone line. Activation of the service enables lights-out, continuous monitoring of the unit. This offers the following benefits:

- Potential to increase availability of the Sun StorEdge 6920 system
- Faster problem resolution should a problem occur

How It Works

The Sun StorEdge 6920 system comes with the necessary hardware (located in the Storage Service Processor accessory tray) and software (loaded on the Storage Service Processor) to take advantage of the remote support feature of the Sun StorEdge Remote Response service.

The hardware consists of a network terminal concentrator (NTC), modem, firewall/router, and integrated power supply module. The NTC provides a modem connection point for the Sun StorEdge Remote Response service. Sun Solution Center engineers connect through the modem to the NTC to log in to the Storage Service Processor. A PPP connection is then established between the Storage Service Processor and remote support server using the NTC to pass only the data stream.

The modem enables the Sun StorEdge 6920 system to communicate with Sun Solution Center over a dial-up telephone connection. It is configured to dial up the local support center whenever the system either detects the need for a service action or periodically checks in with a heartbeat message. Sun service personnel dial back in to the Storage Service

Processor to obtain additional data about the alert or, in some cases, to fix it remotely. For a group of Sun StorEdge 6920 systems, a modem can be shared, with one system's modem designated to act as the modem proxy for the other system's modem.

The Sun StorEdge Remote Response service process is illustrated in FIGURE 2-3.

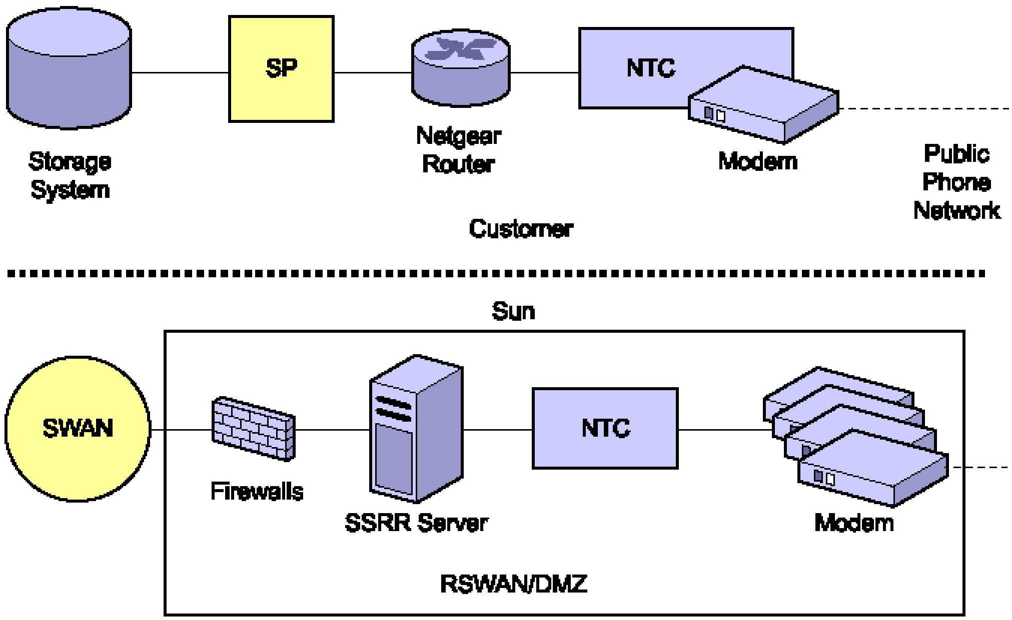

FIGURE 2-3 Sun StorEdge Remote Response Service

The firewall and router provide an additional layer of security by using network address translation (NAT) and disabling specific network services from the customer's local area network (LAN). When several Sun StorEdge 6920 systems are physically located within the same data center or site and the Sun StorEdge Remote Response service is employed, only one telephone line is needed to support the site.

Note – The USER LAN IP address should not overlap the COMPONENT_LAN IP address (192.168.0.X) or the SP LAN address (10.0.0.X). Therefore, the USER LAN IP address cannot be set to 192.168.0.X or 10.0.0.X.

Shared Telephone Lines

If a Sun StorEdge 6320 system is sharing a phone line with a Sun StorEdge 6920 system, the former must have the latest Remote Response firmware on the Storage Service Processor. The Remote Response service enables up to eight systems to be connected together to share a single telephone line for communication with the Sun Service Center and support teams. To connect systems together, connect the SP LAN OUT port of each system's service panel to the SP LAN IN port on the adjacent system's service panel. The first system in the chain must have nothing connected to its SP LAN IN port and the last system in the chain must have nothing connected to its SP LAN Out port. In addition, during the initial configuration of each system, a unique system ID must be assigned. One of the systems (preferably the system with the phone line connected) must be assigned a system ID of 0.

Security

Security is provided at four levels:

- A secure shell (`ssh`) is used between the Storage Service Processor and the modem, as well as between the Sun StorEdge Remote Response service server and the Storage Service Processor.
- The NTC requires authentication before allowing traffic from the modem to the Storage Service Processor LAN.
- A firewall/router prevents hackers from using standard network protocols to gain unauthorized access to the system.
- Management traffic to the system is encrypted.

Note – You can also add a firewall between your LAN and the Sun StorEdge 6920 system to provide an additional layer of security.

Customer Contact

Customers who reside in any of the following countries can enable or activate the Sun StorEdge Remote Response service: Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, China, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, India, Ireland, Italy, Japan, South Korea, Malaysia, Mexico, The Netherlands, New Zealand, Norway, Philippines, Poland, Portugal, Romania, Russia, Singapore, South Africa, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, U.K., and U.S.A.

Customers in other countries should do one of the following:

- Contact Sun or a licensed Sun reseller or service provider to activate the Sun StorEdge Remote Response service.
- Run the `setcountrycode` command as part of `sscs` to set up the modem properly, and then contact Sun or a licensed Sun reseller or service provider to activate the Sun StorEdge Remote Response service.

CHAPTER 3

Working With Volumes

A volume is a logically contiguous range of storage blocks presented by a disk array as a logical unit number (LUN). A volume can span the physical devices composing the array or be wholly contained within a single physical disk, depending on its virtualization strategy, size, and the array internal configuration. The array controller makes these details transparent to applications running on the attached server system.

This chapter provides instructions on working with volumes. It contains the following sections:

Understanding Volumes

You can create volumes to efficiently manage data for your storage system. Volumes act as containers into which applications, databases, and file systems can store data.

The Sun StorEdge 6920 system presents a volume as a logical entity that can be mapped to a host and allocated from a storage pool.

Note – Each volume must have a unique name, regardless of the domain in which it resides.

You manage the system's physical storage by placing virtual disks in a pool of storage space for creating volumes. Volumes are created from the virtual disks in storage pools. Based on your specifications, the system automatically allocates the storage from the appropriate pool to satisfy your volume configuration requirements.

When you create a volume, the system allocates capacity from a storage pool. You will most likely also map the volume to initiators. The system provides wizards to assist you in creating volumes and assigning them to initiators.

Although you should make every attempt to configure the sufficient capacity for volumes that you create on your system, you can extend volumes if you find that their configured capacity is not sufficient.

Note – Extending the capacity of a volume that has snapshot reserve space enlarges the reserve to match the new volume size. If you need to extend a volume with snapshots, you must first delete the snapshots.

The Sun StorEdge 6920 system can manage a maximum of 1024 volumes. The following objects count against that total, even when they are not presented to hosts as volumes:

- **Single volume** – A logically contiguous range of storage blocks corresponding to virtual disks in the system, not limited to a single array, and presented as a logical unit number (LUN). The details are transparent to applications running on the attached server system.
- **Legacy volume** – An entire LUN on an external storage array that you can manipulate as if it were a local volume, while preserving the user data on the external storage array. You can apply the system's data services to a legacy volume; however, you cannot extend a legacy volume. There is a limit of 128 legacy volumes that can be managed at one time.

  For more information, see "About Legacy Volumes" on page 148.
- **Mirror** – A special type of volume consisting of up to four separate yet equal mirror components that you can access independently to update and track changes to your data set and to manage your data migration strategy.

- **Mirror component** – One of up to four individual copies of the same data set that constitute a mirror in the Sun StorEdge 6920 system. When you perform an action on a mirror component, the resilvering process synchronizes the mirror so that each component is an equal yet separate copy of the same data set. You can perform a variety of operations on a mirror component.
- **Primary componen**t – The first component (mirrored volume) created to establish a mirror.
- **Split component** – A mirror component separated from the mirror, intended to be rejoined later. A split component is counted toward the limit of four mirror components per mirror and continues to be tracked by the system as part of the mirror.
- **Primary volume** – The volume that contains the original user data that the primary peer replicates to the secondary peer.
- Secondary volume – The remote counterpart of the primary volume. The secondary volume is the replicated copy of the primary volume. You can map or create a snapshot of a secondary volume. You cannot read from or write to a secondary volume unless it is in Suspend mode or you change its role to primary.
- **Snapshot** – Treated as a volume, a snapshot is an instantaneous image taken at a specific point in time which is identical to the volume at that point. Snapshots are useful for backups of some data and for a rotating data migration strategy, as well. They are volumes with images of their parent volume.
- **Parent volume** – The volume for which a snapshot is taken.
- **Virtual disk** – A volume presented to the Sun StorEdge 6920 system as a set of disk blocks in a range of consecutive numbers, with disk-like storage and I/O semantics. The virtual disk is the disk array object that most closely resembles a physical disk from the operating environment's viewpoint.

Note – There is a limit of 128 legacy volumes that can be managed at one time.

Preparing for Volume Creation

Sun StorEdge Storage Pool Manager enables virtualization and pooling of storage assets and flexible allocation of these storage assets based on application type. In general, the Sun StorEdge Storage Pool Manager enables you to create storage pools, from which volumes are carved. Storage pools are quickly and conveniently created from storage profiles. Fourteen predefined storage profiles are included with the Sun StorEdge 6920 systems. You can also create your own storage profiles. The storage profile of the pool you select determines the volume's storage characteristics.

Before you create a new volume, you need to make decisions about the following:

- The storage domain that will be associated with the volume
- The volume type – single or mirrored
- The storage pool in which the volume will be created
- The capacity of the volume – minimum 16 megabytes; maximum 2 terabytes
- The initiators you will map to the volume
- Whether you will use split mirror components or snapshots for pictures of your data at a specific time:
  - Split mirror components – When you want to take a picture of the data set represented in a mirror, you can split a mirror component from the mirror and save it as an independent copy or reference point of your data at that time. Although you cannot roll back mirrored snapshots, you can split a mirror component and later rejoin it to the mirror. You can even perform a reverse rejoin action on the previously split component if you want the other mirror components to synchronize to its data. The split mirror component has no snapshot reserve space limitations.

    For more information, see "Split Mirror Components Compared With Snapshots" on page 57.
  - Snapshots – The snapshot feature is most useful as a method of tracking data through precise, point-in-time pictures that reside in snapshot reserve space. You can modify the snapshot reserve space of a mirrored volume as you would of any other volume. The older a snapshot, the more likely that the data it represents has changed. If the volume is expected to change substantially between snapshots, using split mirror components might be more appropriate.

    For more information, see "Snapshots Compared With Split Mirror Components" on page 79.
- The snapshot reserve space needed if snapshots will be created for the volume
- The virtualization strategy of the volume – striped or concatenated

The following section describes factors involved in choosing a virtualization strategy.

Virtualization Strategies

You can perform virtualization using striping or concatenation. The strategy to be used for all volumes associated with a storage pool is determined by the pool's storage profile. Regardless of the virtualization strategy, the system can divide each virtual disk into at most 32 partitions.

Striping

When you create a striped volume, the system divides data into blocks and distributes (stripes) the data blocks evenly across several disk blocks. Striping data improves disk performance because it enables simultaneous disk read/write operations on multiple disks. If you do not select the "stripe all" option, the Sun StorEdge 6920 system distributes the volume space within the volume equally by using the smallest number of virtual disks possible to create a volume of appropriate size. When you select the "stripe all" option, the volume will be created by striping across as many vdisks in the pool as still have adequate space.

A pool using a stripe virtualization strategy must have a least two virtual disks added to it before volumes can be created. A volume created from a stripe pool with two virtual disks consumes one partition on each virtual disk.

Keep in mind the following parameters for striped volumes:

- Two virtual disks per tray
- Minimum of two partitions per volume
- Maximum of 32 partitions per virtual disk
- Maximum of 16 virtual disks
- Maximum of 16 controllers per stripe

The stripe size is set in the profile applied to the volume. The Sun StorEdge 6920 system supports the following stripe sizes:

- 128 kilobytes
- 512 kilobytes
- 1 megabyte
- 2 megabytes
- 4 megabytes
- 8 megabytes
- 16 megabytes

FIGURE 3-1 depicts a striped volume (Volume 1) with a size of 400 gigabytes inside a pool with four virtual disks. The system creates a 400-gigabyte striped volume by allocating 100 gigabytes of space equally across all the virtual disks.

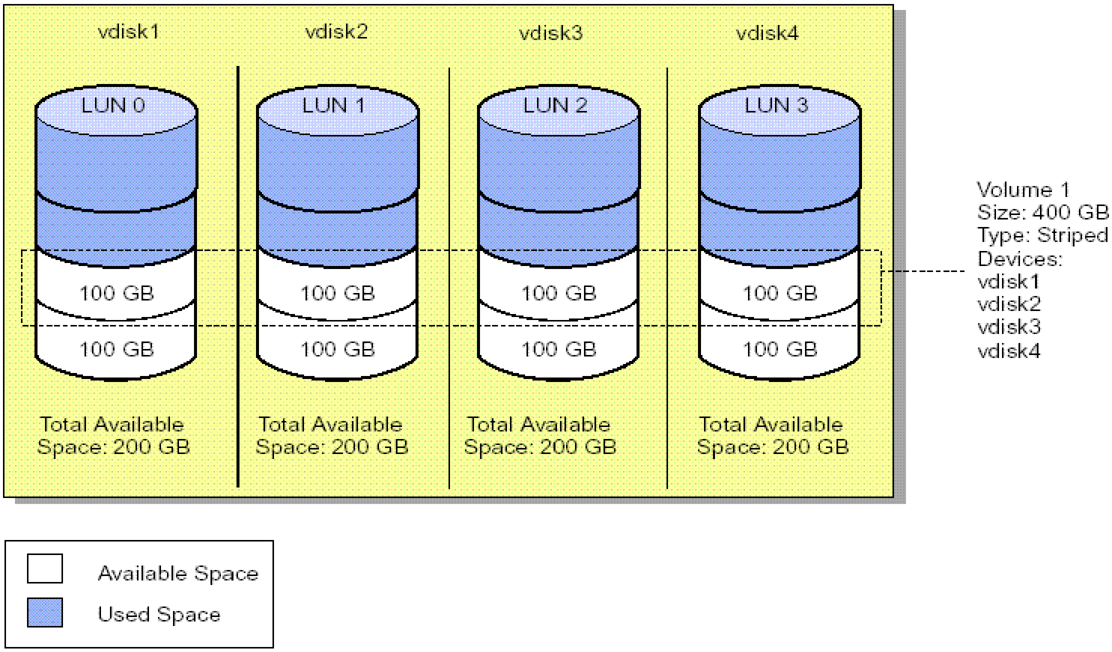

FIGURE 3-1 Striped Volume Example

The system is able to detect 400 gigabytes of total space for the volume using the smallest number of virtual disks. If a virtual disk is unavailable for use (for example, it has no free space on it or it is uninitialized), the system does not consider it for inclusion in the volume you are creating. If the pool has insufficient available free space, the system cannot create the volume.

Concatenation

When you create a concatenated volume, virtual disk partitions are sequentially filled with data. The second disk is written to when no space remains on the first, the third when no room remains on the second, and so on.

The system distributes the volume space across virtual disks, with the following conditions:

- By default, the system first uses the virtual disks with the smallest amount of available space.
- If you specify the maximum size available in the pool, the system first uses the virtual disks with the largest amount of space.

FIGURE 3-2 shows the creation of a concatenated volume (Volume 2) with a size of 400 gigabytes.

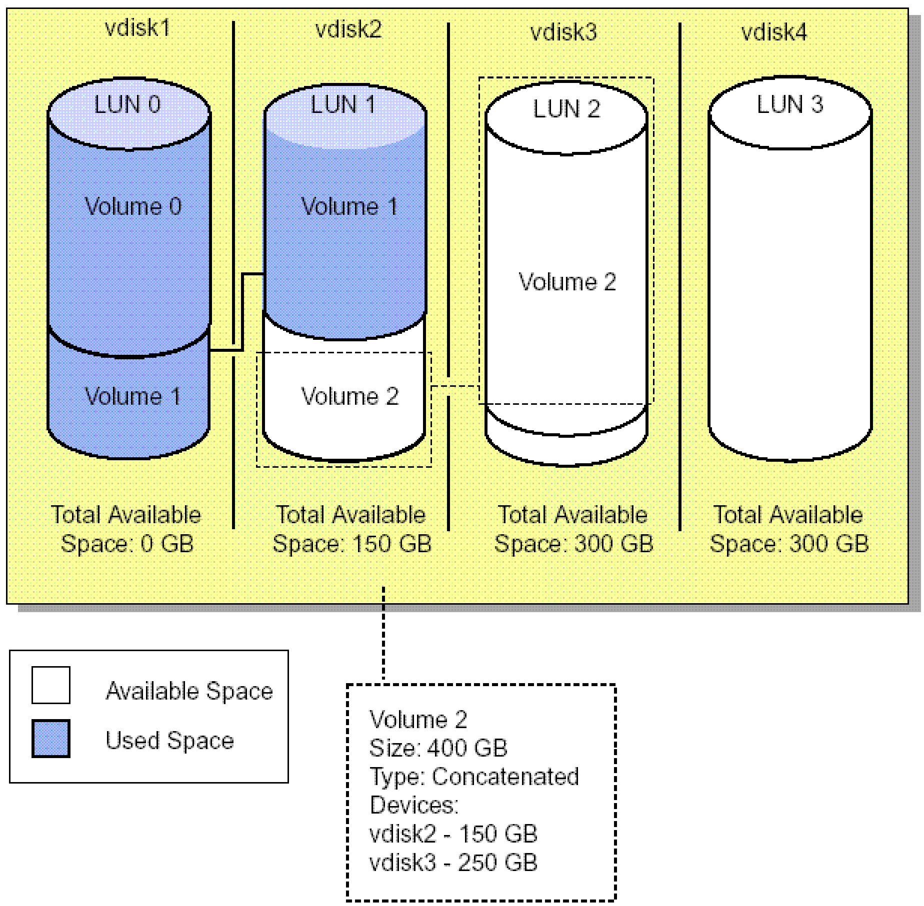

FIGURE 3-2 Concatenated Volume Example

Concatenated volumes are written serially, starting in the first virtual disk with available space. The virtual disk labeled vdisk1 has no available free space, since Volume 0 and part of Volume 1 have already been written there. The next concatenated volume starts where Volume 1 ends. The system uses up the space on vdisk2 (150 gigabytes of available space), and then uses 250 gigabytes from vdisk3 to allocate 400 gigabytes of total space for Volume 2. The next volume to be created first consumes the available space left on vdisk3, and then proceeds to write to vdisk4. The ratio of volumes to virtual disks for concatenated pools can be at most 32.

A pool using a concatenate virtualization strategy must have at least one virtual disk added to it before volumes can be created. A volume created from a concatenated pool with only one virtual disk consumes one partition.

Note these parameters for concatenated volumes:

- Minimum of one virtual disk per volume
- Maximum of two virtual disks per tray
- Maximum of 32 partitions per virtual disk

Presentation of Volumes to the Host

When volumes you create are presented to a host, you can use the secure LUN mapping feature to map volumes to an initiator on an FC server. The DSP uses the LUN mapping to selectively configure a LUN so that it is available to a specific FC device. LUN mapping matches FC initiators inside the storage domain to particular LUNs. This makes a given LUN visible to an FC target or initiator, while at the same time removing its visibility (or masking it) from another FC initiator. When HBAs are mapped, the hosts are then able to mount the file systems and partition the volumes to the requested sizes, just as with a direct attached storage (DAS) device.

When creating a volume that is to include snapshots, it is best practice that the volume be striped (rather than concatenated). Striping takes best advantage of the distributed nature of the snapshot function. For more information, see Chapter 5 "Working With Snapshots" on page 71.

Typically, storage tends to grow, and the most recent blocks are the ones that exhibit the most write activity. In a concatenated volume, these active blocks likely reside in a single concatenated element. This concentrates the snapshot workload and creates an unnecessary problem area, or hotspot. It is therefore better to configure the volume using striping, which attributes the snapshot workload more evenly among the available resources.

Note – The Sun StorEdge 6920 system supports 1024 objects. This number includes snapshots, so that 512 volumes with one snapshot each totals 1024, which is within system limits.

Best Practices: Using Volumes

The following examples describe volume best practices:

- "Example 3-1. Creating a Volume" on page 45
- "Example 3-2. Extending a Volume" on page 46

Example 3-1. Creating a Volume

You can create volumes to efficiently manage data for your storage system. Volumes act as containers into which applications, databases, and file systems can store data.

You manage the system's physical storage by placing virtual disks in a pool of storage space for creating volumes. Volumes are created from the virtual disks in storage pools. Based on your specifications, the system automatically allocates the storage from the appropriate virtual disks to satisfy your volume configuration requirements.

Notes

- Each volume must have a unique name, regardless of the domain in which it resides.
- The maximum capacity for a volume is 2 terabytes.
- The storage pool from which the volume is created defines the volume's performance characteristics. You can create a pool before creating the volume if a pool with suitable characteristics does not yet exist. For more information, see Chapter 7 and Chapter 9.

Procedure

To create a volume:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Logical Storage > Volumes.

   The Volume Summary page is displayed.

2. Click New.

   The New Volume wizard is displayed.

3. Follow the steps in the wizard.

   The New Volume wizard guides you through the steps to create and map a new volume. The New Volume wizard prompts you to:

- Enter a name and description for the new volume
- Select a storage domain in which you want to create the volume
- Specify the type of volume you want to create: single or mirrored volume
- Enter the capacity of the volume
- Reserve space for snapshots (optional)
- Map initiators to the volume (optional)

If you decide not to map the volume to a host initiator now, you can map a volume later using the Map Volume wizard from the Volume Detail page. For more information about creating volumes and mapping them to hosts, click Help.

Example 3-2. Extending a Volume

Although you should make every attempt to configure sufficient capacity for volumes that you create on your system, you can extend volumes if you find that their configured capacity is not sufficient.

Prerequisites

To extend the capacity of a volume that has snapshots associated with it, you must first delete the snapshots.

Notes

- When you extend one mirrored volume, the capacities of all other components in that mirror are equally increased, regardless of whether they are on different arrays.
- Extending the capacity of a volume that has snapshot reserve space increases the reserve to match the new volume size.
- Extending the capacity of a volume will involve the use of another partition of the virtual disk.
- Do not extend a legacy volume if it is the primary or secondary volume in a replication set. If a legacy volume is a secondary volume in a replication set, do not extend the primary volume beyond the size of the legacy volume.
- The maximum capacity for a volume is 2 terabytes.
- Be sure to fully back up a volume before you follow this procedure, to ensure your data is preserved if issues occur when you attempt to extend it.
- A volume can have a maximum of 16 partitions. When you reach this limit, you can no longer extend the volume.

Procedure

To extend the capacity of a volume:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Logical Storage > Volumes.

   The Volume Summary page is displayed.

2. Select the volume whose capacity you want to extend.

   The Volume Details page is displayed.

3. Click Extend.

   The Extend Volume page is displayed.

4. Select Maximum Capacity or specify the new capacity you want.

5. Click OK.

Displaying Volume Information

This section describes the following procedures:

- "Displaying Volume Information Using the Browser Interface" on page 47
- "Displaying Volumes Using the CLI" on page 48
- "Displaying Virtual Disks Within the Volume" on page 50

For other procedures regarding managing volumes, see the online help.

Displaying Volume Information Using the Browser Interface

You can display summary information and details about existing storage volumes. You can also view additional information about initiators, snapshots, and virtual disks associated with each volume.

To display information about volumes using the browser interface:

1. Click Sun StorEdge 6920 Configuration Service > Logical Storage > Volumes.

   The Volume Summary page is displayed.

2. Select a volume to see its detailed information.

The Volume Details page is displayed.

3. Go to Additional Information to see the number of mapped initiators, replication sets, snapshots, and virtual disks associated with the volume.

   Click an item to see its Summary page.

Displaying the Data Path Using the CLI

To complete the high level view of the data path from the CLI, display the initiators and the associated volumes, as shown in the following example:

```
[admin_host]# sscs list initiator
Initiator: host1_port1 Storage Domain: AppDomain
Initiator: host1_port2 Storage Domain: AppDomain
[admin_host]# sscs list initiator host1_port1
Initiator: host1_port1
Description: host1_port1
Storage Domain: AppDomain
WWN: 210000E08B0760B5
OS Type: Standard
State: Allocated
Status: Online
Associations:
Volume: r0vol0 LUN: 0 Permission: Write/Read Map State: Online
Volume: r0vol1 LUN: 1 Permission: Write/Read Map State: Online
```

Displaying Volumes Using the CLI

When you identify the World Wide Name (WWN) presented by Sun StorEdge Traffic Manager on the host, you can compare it to the WWN for the same volume as presented by the command-line interface (CLI). Use the `sscs` command to display volume information as follows:

```
[admin_host]# sscs list volume
Volume: r0vol0 Storage Domain: AppDomain
Volume: r0vol1 Storage Domain: AppDomain
[admin_host]#
```

To identify the GUID, you need to display a detailed listing of the volume shown as follows:

```
[admin_host]# sscs list -S AppDomain volume r0vol0
Volume: r0vol0
Description: raid0_vol0
Creation Date: Wed Aug 12 00:00:00 MST 1970
WWN: 600015D00020E1000000000000006709
Storage Domain: AppDomain
Pool: r0pool
Profile: r0profile
Size: 10.0000 gigabytes
State: Mapped
Condition: OK
Type: Concat
Snapshot Policy: Favor original volume over snapshots
Snapshot Pool:
Space Allocated For: No Snapshots
Expected Write Activity: Low(10%)
Snapshot Reserve Size: 0
Snapshot Reserve Percent Full: 0
Snapshots:
Associations:
Initiator: host1_hba1 LUN: 0 Permission: Write/Read Map State:
Online
Initiator: host1_hba2 LUN: 0 Permission: Write/Read Map State:
Online
[admin_host]#
```

Displaying Virtual Disks Within the Volume

You can identify virtual disks associated with the volume by using the SSCS CLI, as shown in the following example:

```
[admin_host]# sscs list vdisk
VDisk: disk/3/5/1/0/0 Storage Domain: DatanDomain Pool: r5pool
VDisk: disk/4/5/2/1/0 Storage Domain: DatanDomain Pool: r5pool
VDisk: disk/3/6/1/0/0 Storage Domain: DatanDomain Pool: r5pool
VDisk: disk/4/6/2/1/0 Storage Domain: DatanDomain Pool: r5pool
VDisk: disk/3/6/1/2/0 Storage Domain: AppDomain Pool: r0pool
VDisk: disk/4/6/2/3/0 Storage Domain: AppDomain Pool: r0pool
[admin_host]# sscs list vdisk disk/3/6/1/2/0
VDisk: disk/3/6/1/2/0
Storage Domain: AppDomain
Storage Pool: r0pool
Profile: r0profile
Tray: 0
Associated Volumes:
Volume: r0vol0
Volume: r0vol1
[admin_host]# sscs list vdisk disk/4/6/2/3/0
VDisk: disk/4/6/2/3/0
Storage Domain: AppDomain
Storage Pool: r0pool
Profile: r0profile
Tray: 1
Associated Volumes:
```

From the output, you can see that (virtual disk) `disk/3/6/1/2/0` supports the volume `r0vol0`. This identifies the path as slot 3 FC port 6 on the DSP connected to Array TargetID 1, LUN 2. The second path shows (virtual disk) `disk/4/6/2/3/0`, which is slot 4 FC port 6 on ArrayTargetID 2 LUN 3.

CHAPTER 4

Working With Mirrors

Mirroring is a form of storage whereby two or more identical copies of data are maintained on separate media. Also called RAID-1 and real-time copy, mirroring enables the cloning of data sets to provide redundancy for your system.

The Sun StorEdge 6920 system empowers you to create and manipulate a mirror with up to four mirror components that you can access independently. You can also update each independent copy of the mirror's data set by performing a variety of operations on any mirror component within the mirror.

This chapter provides instructions on working with mirrors. It contains the following sections:

Understanding Mirroring

Mirroring writes every byte of data on one disk to a second disk. As part of disaster recovery, if one disk fails, the second disk continues to operate without interruption. Mirroring's primary advantages are data protection when a disk fails and increased read performance. Read performance is slightly faster because the read I/O workload is spread across multiple disks.

In the Sun StorEdge 6920 system, a mirror is a special type of volume made up of as many as four mirror components that you can use for data redundancy protection, to track changes to your data set, and to migrate data.

A mirror is made up of independent mirror components that reflect the same data set. The first mirrored volume that you create (also called the primary component) establishes the initial data set for the mirror that will be copied to mirror components that you add.

Note – The Sun StorEdge 6920 system implements parallel writes to all components in the mirror.

How It Works

You manage a mirror through its mirror components. An action you perform on one mirror component can cause all other mirror components to synchronize with it, and vice versa. When you perform an operation that causes the data of one component or the entire mirror to change, the process is called resilvering.

Mirror components consist of data surfaces and logs. When you perform a write operation on one component's surface, the system executes parallel writes to the other surfaces in the mirror. The mirror log keeps track of actions you perform until resilvering is complete.

Mirror Actions

You can perform the following actions on mirror components:

- **Split** – You can separate a mirror component from the mirror, with the intent to rejoin it later. Because a split mirror component tracks changes, it counts toward the limit of four mirror components per mirror. You can use a split mirror component to create an image of your data at the time of the split.
- **Rejoin** – You can rejoin a split mirror component back into the mirror. Resilvering immediately begins, and the previously split mirror component becomes identical to its counterparts in the mirror. Because the split component maintains a record of changes made to the mirror (and the split component) after the split, resilvering takes little time. Only the changes are copied as the component is rejoined to the mirror.
- **Reverse rejoin** – You can perform a reverse rejoin action on a split mirror component when you want to use its data as the basis for the mirror as a whole. The result is that every mirror component in the mirror becomes identical to the previously split mirror component, and the data set represented in the mirror matches that of the previously split mirror component.
- **Break** – You can remove a mirror component from the mirror, voiding its association with the other mirror components. The broken mirror component becomes a standalone volume in the system, and synchronization with the other mirror components is not maintained. Because changes to the mirror are not tracked through this volume, the resilvering that occurs after you add a previously broken component back to a mirror takes longer than resilvering does with a split component, because more data has to be copied.

Note – I/O to split mirror components might load separate spindles, depending on where the volumes are presented; however, they incur overhead to track changes. When you break a mirror component, you potentially load separate spindles and you incur no overhead.

Note – If you break the last remaining component of a mirror, the result is a non-mirrored volume with the same name as the mirror. The name of the previous mirror component is no longer used.

- **Force break** – You can execute a force break operation to remove a mirror component that has not been completely resilvered and is therefore not synchronized with the other components in the mirror. You can use the force break operation to break a mirror component from the mirror, even if the component is inaccessible.

  Just as with a break operation, the mirror component for which you implement a force break becomes a standalone volume in the system; however, the condition of its data might be degraded and it might therefore be inaccessible after being removed from the mirror.

- **Repair** – When the Repair Mirror button is available, the mirror logs are broken or inaccessible and need repair. This can happen if the storage device containing the logs becomes unavailable or damaged. To create new mirror logs in an appropriate place to restore the mirror, click this button.

  The Repair mirror function does the following:

  - replaces inaccessible mirror logs
  - removes mirror components that are no longer associated with virtual disks

---

Note – A component in need of repair because one of its virtual disks is inaccessible can be removed using a force break operation. A component in need of repair because all its virtual disks are inaccessible must be removed using the Repair mirror function. Because it can be difficult to determine whether one or all virtual disks associated with a mirror component are missing, it is best practice to first issue the Repair mirror operation and then force break any mirror components that are inaccessible yet still attached to the mirror.

Mirror Logs

A mirror log is a disk space region that maintains status and metadata information about the mirror. The mirror log contains no user data but does contain pointers and bitmaps to track create, rejoin, reverse rejoin, and other operations. Sun StorEdge Data Mirror software automatically sets up and manages its own mirror log space. The mirror log is itself mirrored for redundancy.

When a write operation is executed on one surface of a mirror, the virtual mirror manager writes the data to all surfaces, and the mirror log tracks this. When the data is successfully written to all surfaces of a mirror (in parallel fashion), the mirror log clears automatically.

When a mirror component's surface is inaccessible, missing, or "dirty," you can use the mirror log to track changes. You can click the Repair Mirror button to create a new mirror log.

FIGURE 4-1 illustrates the Sun StorEdge 6920 system's mirroring functionality.

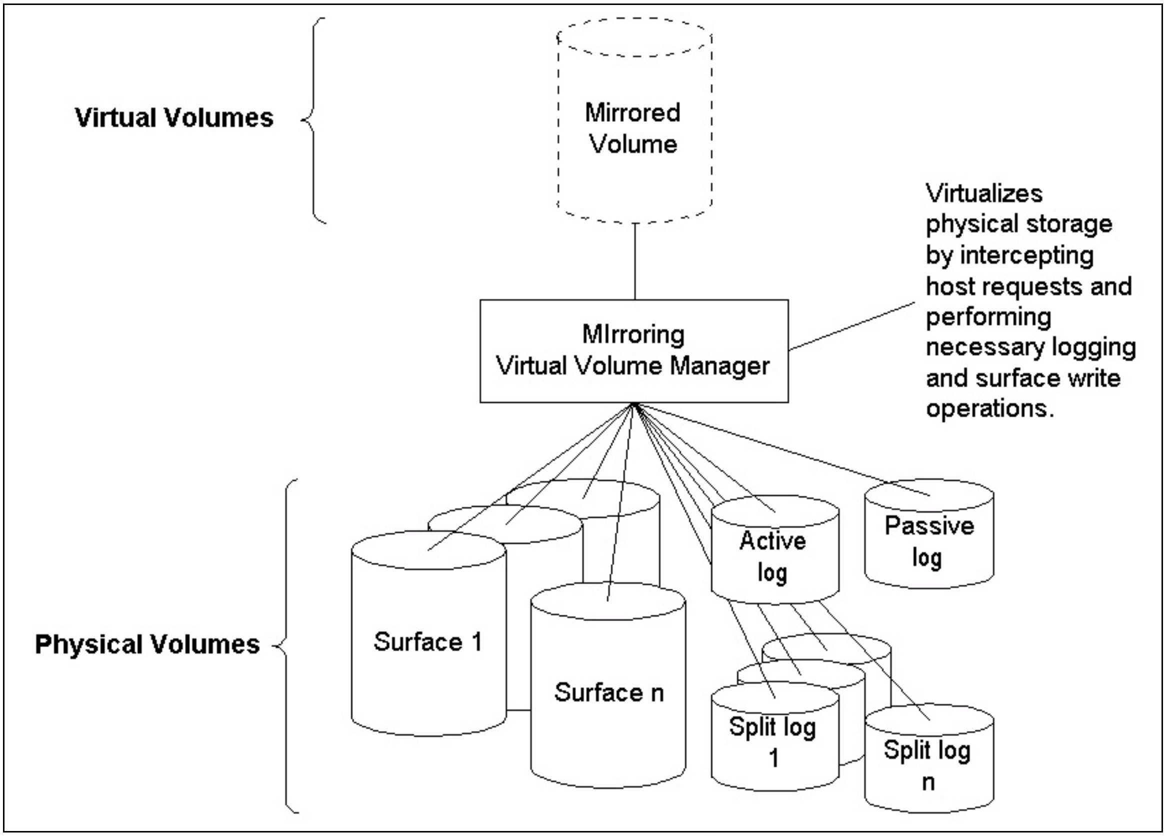

FIGURE 4-1 How Mirroring Works

Preparing for Mirror Creation

To create a mirror, you can create a new mirrored volume or mirror an existing volume and then add other volumes as equal mirror components.

Note – All elements of a mirror must reside in the same storage domain, though mirror components can reside in different storage pools.

Before starting the process, you must determine the following:

- The mirror components that will make up the mirror.
- The isolation policy you want for mirror components:
  - Required – Mirror components must reside on different arrays (default).

    When mirror components exist on different arrays, redundancy protection is at its highest for data preservation.
  - Optional – Mirror components can reside on the same array.

    If you know that a mirror component shares an array with other components in the mirror, you must set the isolation policy to Optional.

    When mirror components share an array, the redundancy factor is reduced for the data set associated with each mirror component should the array or access to the array fail.
- The resilver priority setting you want to set for the system:
  - High – The system attempts to resilver the mirror quickly, but this might noticeably slow down other system operations and host I/Os.
  - Medium – The system balances I/O activities with the resilvering process, yielding an incremental, steady pace for all jobs being executed.
  - Low – The system takes longer to resilver the mirror, but with less impact to the rest of the system.
- Whether you will use split mirror components or snapshots for copies of your data at a specific point in time.

Split Mirror Components Compared With Snapshots

If your organization has licensed both the Sun StorEdge Data Mirror software and the Sun StorEdge Data Snapshot software, there several reasons to consider whether to use split mirror components or snapshots as images of your data at a specific time.

Split Mirror Components

When you want an image of the data set that is represented in a mirror, you can split a mirror component from the mirror and save it as an independent copy of your data at that time. After you have analyzed the data in the split mirror component, you can rejoin it to the mirror, in which case it is synchronized to match the mirror's data set. Alternatively, if you want the other mirror components to be synchronized to match the data of the split component, you can reverse rejoin it to the mirror. Reverse rejoining a split mirror component is similar to rolling back a snapshot, because it causes the data set of the mirrored volume to match the data set of the point-in-time (PIT) image represented by the split component. One difference between using a split mirror component and using a snapshot rollback as an image of your data is that the mirroring process does not require snapshot reserve space.

You can rejoin the split component to the mirror, to synchronize it back to the mirror's data set, or you can perform a reverse rejoin action on the previously split component if you want the other mirror components to synchronize to its data. In effect, reverse rejoining a mirror component is similar to rolling back a snapshot. One difference in using a split mirror component as an image of your data is that it does not require snapshot reserve space. If the application will need to write to the split/snapshot volume, use a split component to avoid consuming snapshot reserve.

You can choose to use low-cost legacy storage devices for split mirror components when a mirror is created. Whereas a snapshot needs to read from the parent volume, a split component stands alone. If an application needs to write to the image, you can use a split component to avoid consuming snapshot reserve space.

One popular use for split mirror components is as part of a rotation process used to track changes in your data set over a period of time. Suppose that you want to add a fourth mirror component to your three-component mirror, as part of a strategy in which you rotate mirror components. Each day you add a mirror component, wait for it to resilver so that each mirror component is identical to the others, and then perform a split operation to save an image of the data at the time of the split.

Snapshots

The snapshot feature enables you to track data through precise, point-in-time images that reside in snapshot reserve space. A snapshot is dependent on the existence of the parent volume and the snapshot reserve space allocated within it. You can take up to eight snapshots for a parent volume. You can modify the snapshot reserve space of a mirrored volume as you would the snapshot reserve space of any other volume.

Because snapshots don't need to resilver to update, the resnap operation is quicker than splitting a mirror component, rejoining the mirror component, waiting for resilvering to complete, and then splitting the component once again.

One popular use for snapshots is reducing downtime for backup operations. Rather than backing up online data, you can back up snapshots and thereby enable critical transactions to continue during the backup process. You can also create frequent snapshots and, in the event of an error, execute a rollback operation to restore your data.

A snapshot uses less space than a split mirror component. The amount of space required by a snapshot is proportionate to the amount of data that changes after the snapshot is taken. The older a snapshot, the more likely the data it represents has changed; thus, the more space it requires. If the parent volume is expected to change substantially between snapshots, using split mirror components might be more appropriate because of the overhead of copy-on-writes.

Summary of Differences

The major differences between split mirror components and snapshots as images of data at a specific time are summarized in TABLE 4-1.

TABLE 4-1 Split Mirror Components Compared to Snapshots

Characteristic	Split Mirror Component	Snapshot
Point-in-time (PIT) image availability	A complete, independent copy of the data.	A copy of the data that is dependent on both the original volume component and the snapshot reserve space.
Capacity used	Each image requires storage for a complete copy of the original volume.	Storage required depends on how many changes have occurred on the parent volume since the snapshot was created.
Quantity of PIT images	Up to 3.	Up to 8.
Time until PIT access	The resilvering process must run before the mirror can be split. The first time a mirror component is added, the entire volume must be copied and this can take a long time. The length of time the resilvering process needs after a rejoin depends on how many changes have occurred on the mirror since the split.	Ready to access as soon as the snapshot is created. This takes only a few seconds.
Capacity reclaim	Components of a mirror can be broken off and deleted to reclaim space.	Snapshot reserve space is reclaimed after the oldest snapshot is deleted. For more information, see FIGURE 5-3 on page 78.

Best Practices: Using Mirroring

The following examples describe mirroring best practices:

Example 4-1. Using Mirroring to Ensure Availability of Data

This example describes the practice of using data mirroring to ensure data availability. Mirroring maintains multiple, identical copies of your data to provide data redundancy. If one of your copies becomes damaged due to a hardware failure, the other copy is still viable and the data remains available. Whereas RAID-1 redundancy protects against the failure of any single disk in a given array, data mirror redundancy protects against a single point of failure within an internal component array.

Prerequisites

Before mirroring your data, answer the following questions:

- What are the profile characteristics you want for the user data?
- Do you have a pool with those characteristics?
- Is there sufficient space available in the selected pool?
- Does the pool have a sufficient number of virtual disks (at least two) from different arrays?
- Are all elements of the mirror in the same storage domain?
- Do you have multiple, independent access paths from the host to the Sun StorEdge 6920 system?
- Will accessing the data from this destination create a performance conflict with other storage volumes?

Notes

- The Storage Pool Details page provides an array view that you can use to determine whether there are enough arrays in the pool. Each independent copy of your data is placed on a different array in the pool. For each array, verify that there is enough space among the virtual disks to hold a separate copy of your data.
- This practice maintains two independent copies of the data and thus requires two arrays to contribute virtual disks to the pool. If you want more copies, more arrays must contribute virtual disks to the pool.

Procedure

When you are comfortable with the construction of your selected pool, you can use the Sun StorEdge 6920 Data Mirror software to assure redundancy and availability.

To mirror your data:

1. Select a pool whose profile has the characteristics you want for your data.
2. Verify that two arrays contribute virtual disks to the pool. Each array's contribution must have enough space to hold an entire copy of your data.
3. If you need to add more space or more arrays to the pool, use the Add Storage To Pool wizard to add the necessary resource.
4. Use the New Volume wizard to create a new, mirrored volume in the pool. Specify that you want two components from the selected pool, and set the Mirror Isolation option to "required." This ensures that the two copies of your data are isolated on separate arrays in the pool.

   The system begins a resilvering task in the background.

Upon completion of the wizard, the volume is ready for host access. The system begins a background task to ensure that the two copies start out with identical data.

Result

The system maintains identical copies of your data in each of the independent mirror components. If you lose access to one of the copies due to a hardware failure, the data remains available through the other copy.

Example 4-2. Using Mirroring to Make Independent Copies of Data

This example describes the practice of making independent copies of a selected volume of user data. The copies have the same profile characteristics as the original data. The user data remains online and accessible throughout all phases of the process.

Prerequisites

Before making copies of a volume's data, consider the following questions:

- What are the characteristics of the selected volume?
- Is there sufficient space available in the pool containing the selected volume?
- Will the copy operation or subsequent access to copies create a performance conflict with the original or other storage volumes?

Notes

- The mirror assumes the World Wide Name (WWN) of the first mirrored volume (primary component) that you create to establish the mirror. Any components that you add to the mirror retain the WWN of the primary component. When a mirror component is broken from the mirror and becomes a standalone volume in the system, its original WWN is restored.
- The copy operation consumes system resources. Large volumes might require a long time to copy.
- Each copy of the data requires the same amount of space as the original volume.

Procedure

When you are ready to copy your user data, use the Sun StorEdge Data Mirror software to perform the operation.

To make independent copies of your data, follow these steps for as many mirror components as you want to add.

1. Select the volume that includes the data you want to copy, click Mirror, and follow the steps in the wizard to make it a mirrored volume.

   The system takes care of copying the data for you, using an internal copy mechanism. This is done without host I/O intervention or interference with access to your data from the host. It runs as a background job whose progress you can monitor on the Job Details page for that job.

2. When specifying the destination of the copies, select the same pool as the original volume or a different pool. Be sure that the destination pool has enough space for each complete copy of the data.
3. Select the mirror isolation property of “optional,” to specify that virtual disks are not required to reside on the same array.
4. Quiesce your application so that the data is in a consistent state, and then use the break operation to separate each copy from the original data.

When you have finished making copies of the data, break the last (original) component from the mirror. This turns the mirror into a simple volume, which consumes fewer system resources than a mirror.

Result

The copies can be mapped and accessed for data analysis or archiving. The host continues to access original data from the selected volume throughout the procedure.

Example 4-3. Using Mirroring to Migrate Data

Mirroring can be an important part of your data migration strategy. For example, you might choose to consolidate existing storage on new devices, creating multiple copies of data for archiving, off-site storage, or other uses such as data mining or new application deployment.

You can use mirroring to migrate data from legacy to internally managed storage on the Sun StorEdge 6920 system by mirroring the data on a legacy volume to a normal volume, during which time the system can continue to access the legacy volume. When mirroring is complete, the mirror components are resilvered so that the legacy data also exists as an independent copy on the new mirrored volume. Once you can access your data on the new mirrored volume, you might choose to delete the legacy volume.

Prerequisites

Before you can use mirroring to migrate from legacy to regular storage, the external storage must be added as a legacy volume that can be accessed by the Sun StorEdge 6920 system.

Before migrating a mirrored volume’s data, also answer the following questions:

- What are the characteristics of the destination volume?
- Is there sufficient space available on the destination volume?
- Do you have a data protection and access strategy in place such as multipathing and high-availability (HA) array?

- Will accessing the data from this destination create a performance conflict with other storage volumes?
- Is the pool to which you are migrating/mirroring in the same storage domain as the primary component (mirrored volume)?

Notes

- The mirror assumes the WWN of the primary component (mirrored volume) you create to establish the mirror. Any existing volumes you add to the mirror preserve the WWN of the primary component.
- There must be at least two virtual disks in a storage pool to enable legacy mirroring.
- The pool to which you migrate or mirror must to be in the same storage domain as the primary component (original mirrored volume); all mirror components must reside in the same storage domain.
- When you create a legacy volume, the system lists the available legacy pools to which you can add it. A legacy pool is defined by the Legacy profile, which ensures that you can manage the legacy volume and not allocate it as raw storage for new volumes. The wizard does not allow you to place a legacy volume in a pool other than a legacy pool. The system provides a default legacy pool. Alternatively, you can create your own legacy pool (and place it in any storage domain), which must be defined by the preconfigured Legacy profile; you cannot create a custom Legacy profile.

Using mirroring as part of the data migration process is shown in FIGURE 4-2.

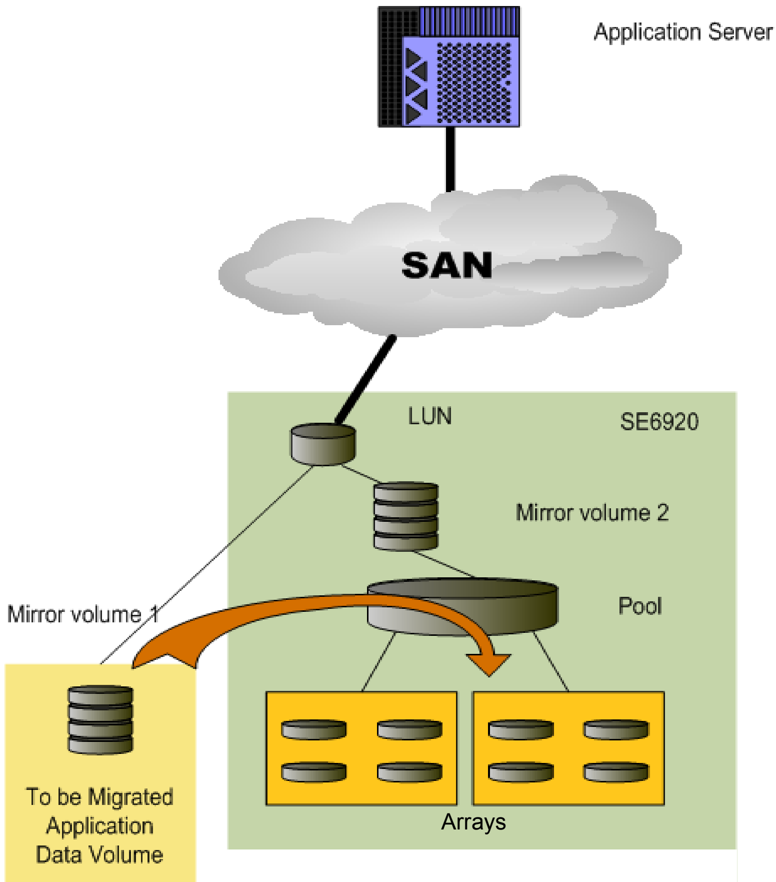

FIGURE 4-2 Mirroring as Part of Data Migration

Procedure

When you have determined how the volume will be accessed from the location to which it is migrated, you are ready to perform the migration.

To migrate your data:

1. Select the volume that includes the data you want to migrate, click Mirror, and follow the steps in the wizard to make it a mirrored volume.
2. For the destination of the migration, create a volume with capacity equal to or greater than the capacity of the original volume.

3. Add the new volume as a mirror component of the original (now mirrored) volume.

   The system takes care of migrating the data for you, using an internal copy mechanism. This is done without host I/O intervention or interference with access to your data from the host. In other words, it is online from the host's perspective. The migration runs as a background task whose progress you can monitor on the Job Details page for that job.

4. When the job has finished, perform a split or break operation to separate the mirror component (original mirrored volume/primary component) from the mirror.

For a related example, see "Example 7-1. Migrating Externally Stored Data to the Sun StorEdge 6920 System" on page 159.

Result

The host from which you migrated your data accesses data from the newly created volume. You can use the original volume for other purposes, or you can simply delete the volume and return the storage to the pool for other uses.

Example 4-4. Using Mirroring to Migrate Data From One Quality of Service to Another

This example describes a best practice method for altering the quality of service of a user volume while preserving access to the data. A data mirror is used to move the volume to a new location having the desired characteristics. You specify the mirror destination or source, with the option of using a legacy volume. User data remains online and accessible to the host throughout the procedure.

Prerequisites

When altering a volume's quality of service, consider the following questions:

- What profile characteristics will the volume have?
- Does a pool exist that has those characteristics and also sufficient space to hold a copy of the user data?
- Will the operation of copying the user data create a performance conflict with other storage volumes?

Note

The pools used for mirroring must reside in the same storage domain.

Procedure

When you have prepared a pool having the characteristics you want, you are ready to move the volume into that pool.

To move the volume:

1. Select the volume whose characteristics you want to alter, click Mirror, and follow the steps in the wizard to make it a mirrored volume.

2. In the wizard, specify a single copy with a mirror isolation property of “optional.” Select a pool with the characteristics that you want the volume to acquire.

3. Complete the mirror creation.

   The system takes care of moving the data for you, using a background task that can be tracked on the Job Details page. A large volume might take a long time to move.

4. When the job has finished, perform a break operation to separate the original user data from the mirror.

   Once this volume is separated, you might delete it; its space will be recovered by its storage pool. The host remains mapped to the mirror and now accesses its data through the new component.

5. Break the copy of the data from the mirror. Because this is the last component, the mirror is turned into a simple volume, which consumes fewer system resources than a mirror. This volume has the quality of service defined by the new pool now containing the user data.

Result

The host accesses data from the newly moved volume using the quality of service defined by the pool’s profile.

Example 4-5. Using Reverse Rejoin

This example describes a best practice method for rolling back user data from an independent copy created at some prior time. It enables you to recover data when an application jeopardizes a data set in a volume. The reverse rejoin operation enables you to use a historical copy of the mirror (a split component) to move the data set back to the time of the split.

This process involves two procedures. The first procedure prepares a known, good copy of the user data using the Sun StorEdge Data Mirror software. The second procedure moves this good data back into the original volume.

Prerequisites

Consider the following questions:

- How often do you want to checkpoint the original data volume?
- Is there sufficient space available to maintain a copy of the data?
- How will you decide which data set contains correct, consistent data?

Note

A split component can be on completely separate storage so that the read operations performed on the split component have no impact on I/O to the mirror, yielding an independent copy for analysis.

Preparation Procedure

To prepare the data you will roll back:

1. Select the volume whose data you want to preserve, click Mirror, and follow the steps in the wizard to make it a mirrored volume. In the wizard, specify the same pool as the original volume and a mirror isolation property of “optional.”

   The system copies the data for you, using a job that runs in the background while maintaining host access to the data throughout the operation.

2. When the job has finished, quiesce the user application and perform a split operation to separate the copy of the data from the original (mirrored) volume.

3. Restart the user application.

Result

The split component represents a consistent set of user data at the time of the split. The host continues to access and update the original user volume. You can now map and access the split volume for analysis, backup, or any other purpose.

Recovery Procedure

If you detect that the application has corrupted data in the user volume, use this procedure to recover good data from the split component.

1. Map the split component to a host and verify that it contains a good, consistent set of user data.
2. Unmap the split component.
3. Stop I/O activity and flush/synchronize the data.
4. Perform a reverse rejoin operation to move data from the split component back into the mirror and thereby back into the original volume. The copy operation occurs as a background job that can be tracked by the Job Details page for that job.
5. Once the reverse rejoin operation has started, you can remap the host to the mirror and restart the application. When the host application begins accessing the data, it gets the data from the split component.
6. When the copy operation is complete, you can again quiesce the user application and then split the mirror component to preserve it for a future recovery procedure.

Result

The split component represents a consistent set of user data at the time of the split. The reverse rejoin has made that data available to the application.

Troubleshooting Mirrors

Following are issues that might occur with mirrors and recommended actions toward resolution.

Mirror Component Does Not Break

When an error message indicates that a break operation is unsuccessful, you must remap the volume and repeat the operation.

Mirror Component Does Not Split

When an error message indicates that the split is unsuccessful, do the following:

1. Wait for the component to complete resilvering.
2. Re-map the mirror.

Mirror Component Does Not Rejoin

When an error message indicates that a rejoin operation is not working, wait for the component to completely resilver (synchronize) and repeat the operation.

CHAPTER 5

Working With Snapshots

A snapshot is an instantaneous copy of volume data at a specific point in time. Snapshots reside in the same storage domain as the volume for which they are taken, called the parent volume.

This chapter provides instructions on working with snapshots. It contains the following sections:

Understanding Snapshots

Sun StorEdge Data Snapshot software enables you to create a dependent copy (snapshot) of a parent volume. Available to servers on the storage area network (SAN) as a unique volume, a snapshot always retains an association with the parent volume.

You can work with a snapshot as you would with any other volume, with the exceptions that you cannot take a snapshot of a snapshot, mirror a snapshot, or replicate a snapshot. A maximum of eight snapshots can be associated with one parent volume at any time. Each snapshot can be accessed independently by other applications.

You can use snapshots for the following purposes:

- To reduce downtime for backup operations

  Rather than using online data, back up your snapshots to enable critical transactions to continue during the backup process. Mount the snapshot to the backup server, and then back up the snapshot's data to tape.

- To perform data analysis and test applications with actual, current data. When you do this, you are still running your test application on the same physical resources for all read I/Os directed at data that hasn't changed since the snapshot was taken. This should work if the volumes were created with performance characteristics that are adequate to sustain both workloads simultaneously. Use the most recent snapshot for data analysis or testing, instead of working with operating data that might interfere with critical transactions.
- To restart applications from the snapshot volume

  If an application problem causes questionable data to be written to the parent volume, restart the application with the last known good snapshot until it is convenient to perform a full recovery or rollback.

How They Work

When a write operation occurs for a data block of the parent volume in which the data has not changed since the snapshot's creation, the system does the following:

- Copies the (original) data about to be changed from the parent volume to the snapshot reserve space. Snapshots can accept write operations, and this data is also dependent on the snapshot reserve space.
- Adds a record that indicates the new location for the old data and an association with each snapshot that will use this data.
- Writes the new data to the parent volume.

Subsequent rewrites of that same data block on the parent volume do not recopy the data.

FIGURE 5-1 illustrates how snapshots work. The Snapshot Virtual Volume Manager software performs the operations necessary to virtualize physical volumes into snapshots that you can use for backups or in place of their parent volumes when necessary.

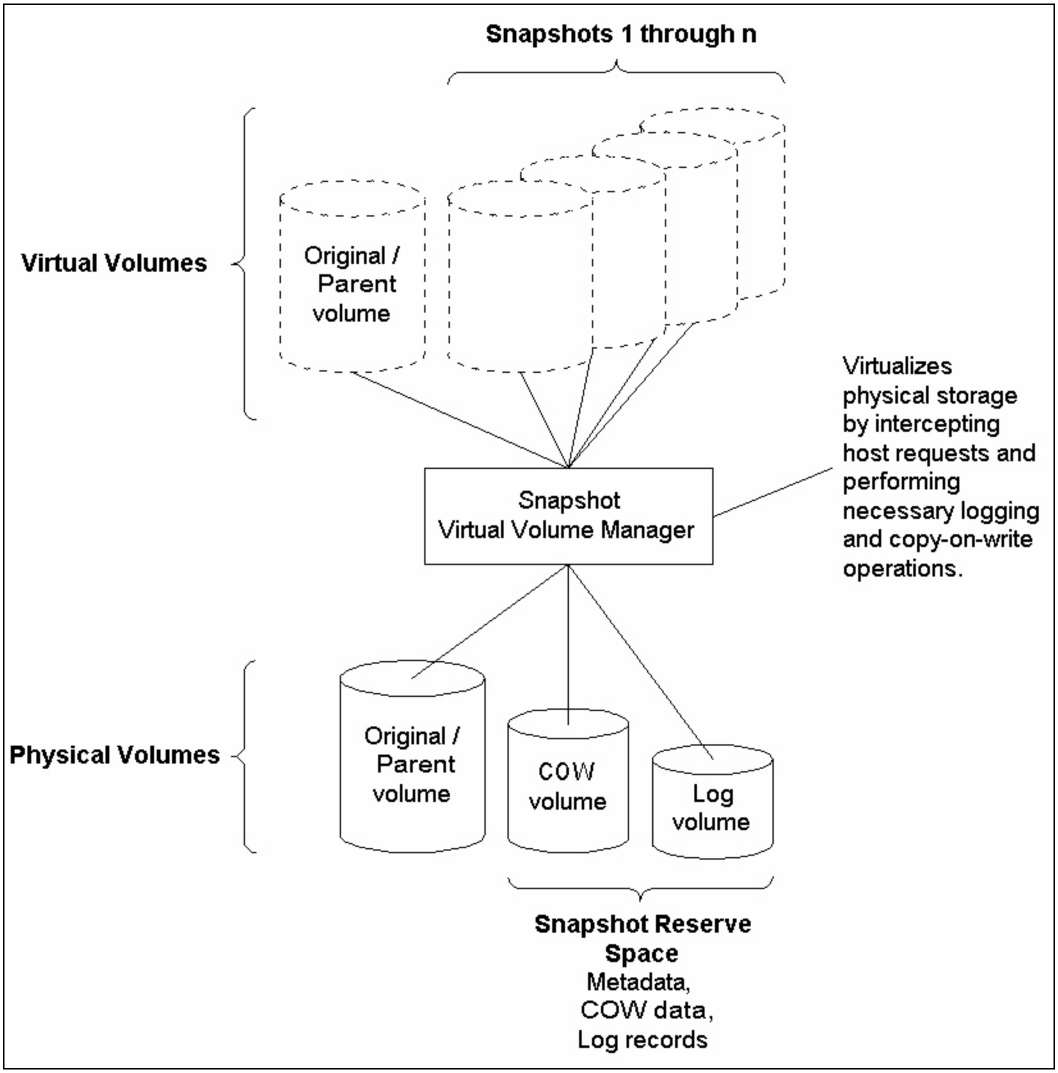

FIGURE 5-1 How Snapshots Work

When a read operation occurs for the (parent) volume for which one or more snapshots have been taken, there is no change to the normal process; it is serviced the same as a volume with which no snapshots are associated.

When a read operation occurs for a snapshot, the data is retrieved from the parent volume, if that data has not changed since the snapshot was created. If the data has changed since the snapshot was created, it is retrieved from the snapshot (COW) reserve space. Read I/Os to snapshots load the same physical drives as read I/Os to the parent volume.

Understanding Snapshot Reserve Space

The system stores snapshots in an area associated with the parent volume called snapshot reserve space. Snapshot reserve space can be a simple partition or a concatenated group of virtual disks. For each parent volume that will contain snapshots, only one area of snapshot reserve space is needed. When necessary, you can expand snapshot reserve space.

You determine the appropriate size of the snapshot reserve space for your business practice based on two factors: the number of snapshots that will be taken of the parent volume, and the anticipated percentage of changed data on the parent volume for each snapshot.

For example, assume that a 2-terabyte production database generates 100 gigabytes of changes on a daily basis. If the requirement is to keep daily copies online for a week (seven snapshot copies of the data), the set of copies requires approximately 700 gigabytes of snapshot reserve space.

How It Works

Snapshot reserve space is consumed when a block is written to the parent volume that has not been written since the snapshot was created, and when a block is written to the volume for the first time. Estimating the rate at which these activities occur can be difficult. If you are unable to estimate the amount of snapshot reserve space needed, the best practice is to set the level of activity to 100 percent, thereby reducing the risk of running out of space.

As you monitor your storage, use the threshold notifications as a measure of how quickly the snapshot reserve space is being consumed. At several thresholds, the system generates messages to indicate the level of space remaining. The system reports the snapshot full percentage for each volume in 5% increments. When a volume’s filled space reaches 70 percent, the system generates a Warning notification. At the 90, 95 and 100 percent thresholds, the system generates a Critical notification.

The system reports the current snapshot usage percentage when any of the following occurs:

- A snapshot is added to the system.
- A snapshot is deleted from the system.
- There is a system reboot or failure.
- Snapshot reserve space is created.

FIGURE 5-2 shows the Snapshot Virtual Volume Manager providing passage for two images of the original/parent volume to be moved into snapshot reserve space, the parent volume, and two snapshots that have been taken of the parent volume.

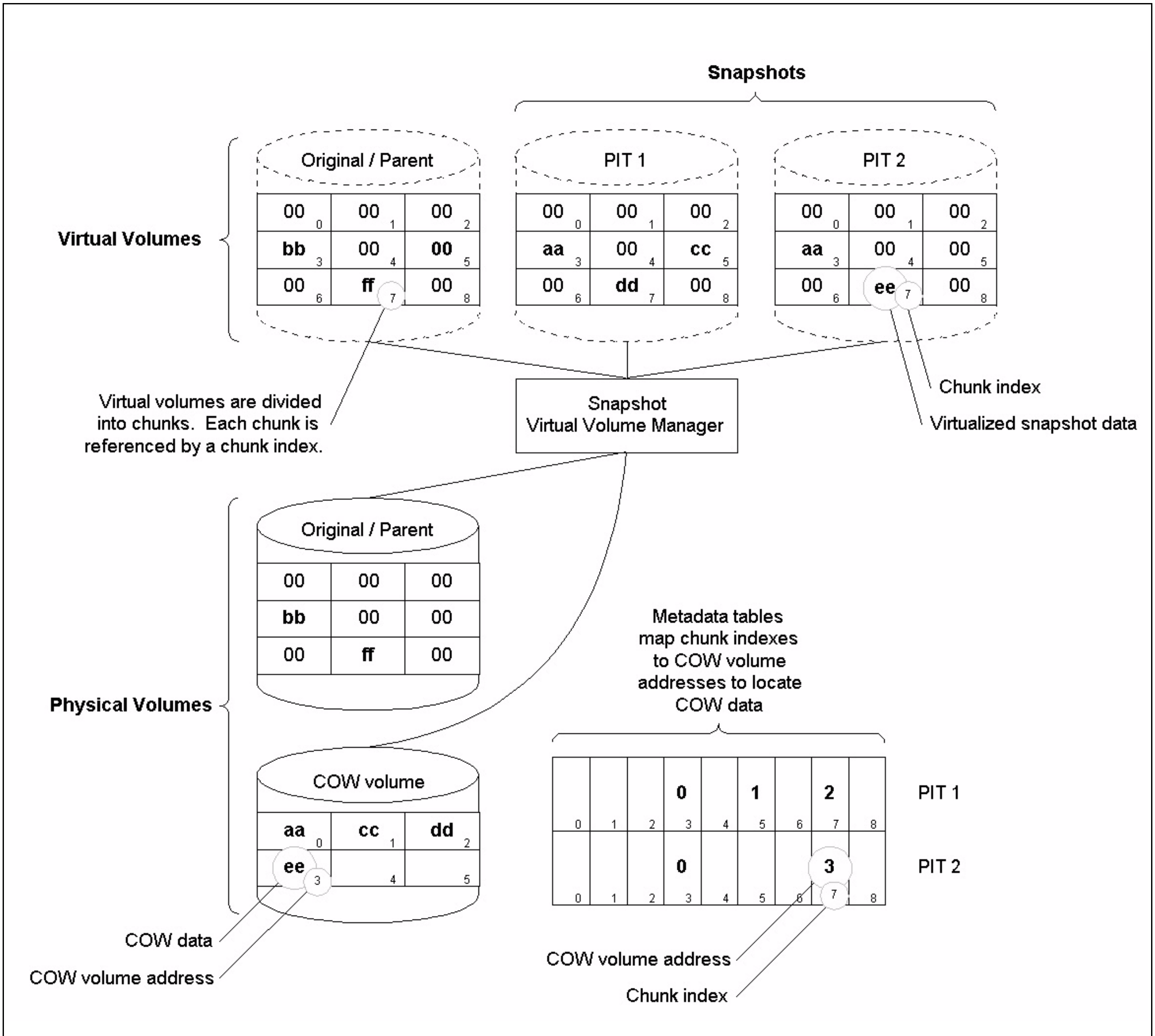

FIGURE 5-2 How Snapshot Reserve Space Works

Planning Snapshots and Snapshot Reserve Space

The first step in planning snapshots is planning the reserve space upon which they depend. You can allocate the snapshot reserve space for a volume when you create the volume, or you can allocate the snapshot reserve space later. A volume with which snapshots are associated is called a parent volume.

You must consider two important factors in determining the size of snapshot reserve space:

- The predicted quantity of snapshots.
- The frequency of write operations expected for the parent volume. Only overwritten data is copied into the snapshot reserve space.

The Snapshot wizard guides you through the process of allocating snapshot reserve space. The system automatically provisions space for the snapshot metadata, which requires a fraction of the space used by the parent volume.

When you allocate snapshot reserve space, you are prompted to specify a snapshot count that can be as high as 32. However, a maximum of eight snapshots can be associated with one parent volume at any time. The ability to allocate up to 32 snapshots makes it possible for you to implement the following scenario.

Suppose you want to take a snapshot of a volume's data on the first day of a 31-day month (March 1, for example) to keep for the entire month (we'll call it SNAPMarch) and then take daily snapshots for a week (SNAP1-7). On the eighth day (March 8), you plan to delete the snapshot that is seven days old (SNAP1) and take a new snapshot for the new day (SNAP8). By the end of the month, you will have taken 31 snapshots (one per day on a rotating weekly schedule or SNAP1-31) plus the one you took on the first day of the month (SNAPMarch). Since the original snapshot was never deleted, snapshot reserve space for 32 snapshots is necessary to accommodate this scenario.

Snapshot reserve space is reclaimed when snapshots are deleted or resnapped in an oldest-to-newest manner. That is, you must delete or resnap the snapshots in the same order in which they were taken (delete #1 before deleting #2; resnap #1 before resnapping #2; and so on).

Because snapshot reserve space is not reclaimed until the oldest snapshot is deleted, and this might not occur for a full month, space for 32 snapshots is provisioned, although there can remain only as many as eight snapshots of a parent volume at any time.

If the parent volume is expected to change substantially between snapshots, making copies using mirroring technologies might be more appropriate, as described in Chapter 4.

Prerequisites for Snapshot Reserve Space

To plan snapshot reserve space, you need to know the following:

- Amount of snapshot activity

  Once a snapshot is taken, subsequent changes to the parent volume data are saved in the snapshot reserve space. Over time, as more blocks of the parent volume are updated, more data is copied to the snapshot reserve space. Therefore, the snapshot reserve space begins to fill as the parent volume changes. When there are multiple snapshots, only one copy of each old data block is maintained, which results in space-efficient operations.

  As the snapshot reserve space is consumed, messages are generated at several thresholds to indicate the amount of space remaining. You can reclaim snapshot reserve space by deleting or resnapping snapshots in the order of oldest to newest.

- Number of snapshots to be taken

  All snapshots of a parent volume share space from the same snapshot reserve space. The larger the number of snapshots taken, the more snapshot reserve space is likely to be required.

Note – There can be at most eight snapshots per volume. Remaining snapshot reserve space is currently not in use or is waiting to be recovered.

- Degree of write activity anticipated on the parent volume

  Write activity is the number of storage blocks that might change on the parent volume after each snapshot is taken. It is a measure of the net change in the parent volume over time. Changes to the parent volume and associated snapshots are stored in the snapshot reserve space.

- Age of the snapshots

  The older a snapshot, the more likely the data it represents has changed, although the rate of change has probably slowed over time. A snapshot never consumes more space than the size of the parent volume (when 100 percent of the volume's data has changed since the snapshot was first taken).

Note – After you have allocated the size of the snapshot reserve space, you can later increase it if necessary. However, you cannot decrease it without deleting and then re-creating the snapshot reserve space.

FIGURE 5-3 shows how snapshot reserve space is reclaimed as data chunks are deleted or resnapped in the order in which they were taken.

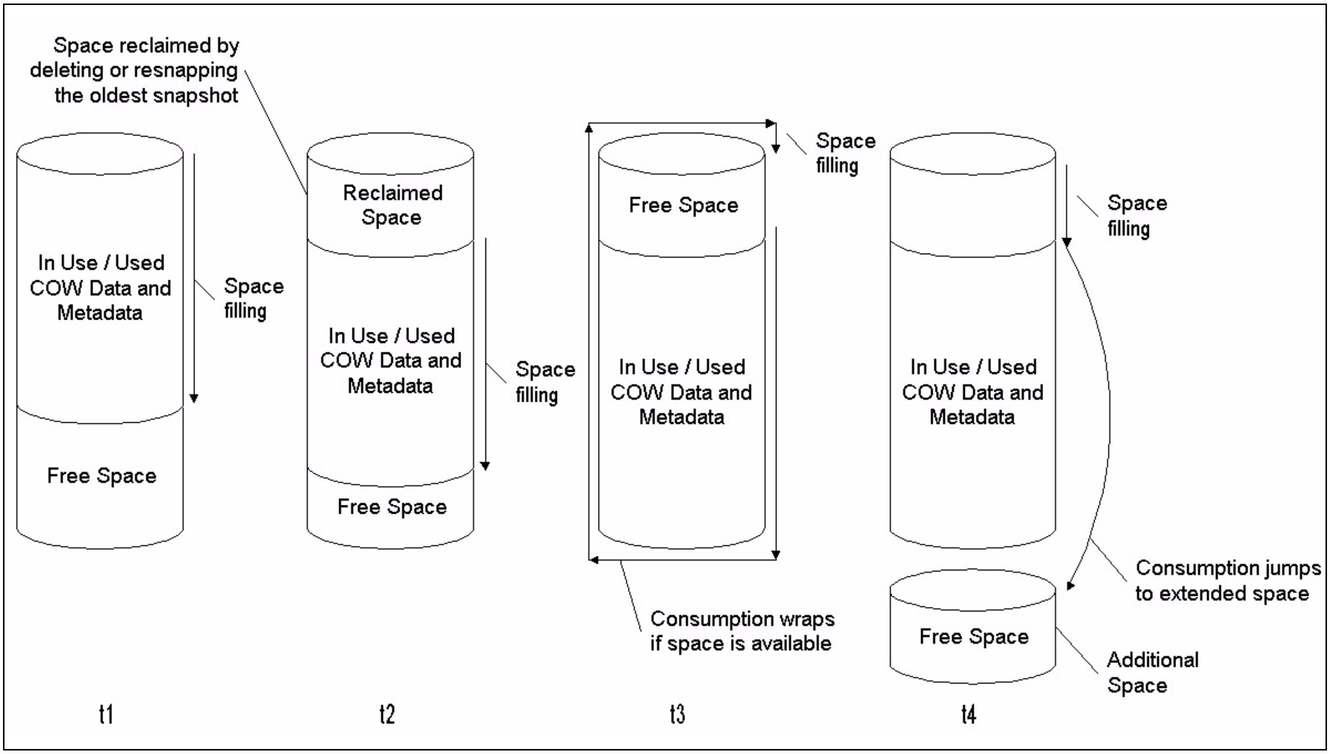

FIGURE 5-3 Snapshot Reserve Space Is Consumed and Reclaimed

Snapshot Virtualization Strategy

When creating a volume for which you plan to take snapshots, make the volume striped rather than concatenated. Striping takes best advantage of the distributed nature of the snapshot function.

The snapshot workload can be divided among the resources of the Data Services Platform (DSP). In particular, when a volume is created out of storage from several arrays and each of the arrays has a dedicated central processing unit (CPU) resource assigned to it, then each of these CPUs performs the snapshot processing for its own volume component.

Typically, storage tends to grow, and the most recent blocks are the ones that exhibit the most write activity. In a concatenated volume, these active blocks likely reside in a single concatenated element. This concentrates the snapshot workload and creates an unnecessary problem area or hotspot. It is therefore better to configure the volume using striping, which distributes the snapshot workload more evenly among the available resources.

Snapshots Compared With Split Mirror Components

If your organization has licensed both the Sun StorEdge Snapshot software and the StorEdge 6920 Data Mirror software, there several reasons to consider whether to use split mirror components or snapshots as images of your data at a specific time.

Snapshots

The snapshot feature enables you to track data through precise, point-in-time images that reside in snapshot reserve space. A snapshot is dependent on the existence of the parent volume and the snapshot reserve space allocated within it. You can take up to eight snapshots for a parent volume. You can modify the snapshot reserve space of a mirrored volume as you would any other volume.

Because snapshots don't need to resilver to update, the resnap operation is quicker than splitting a mirror component, rejoining the mirror component, waiting for resilvering to complete, and then splitting the component once again.

One popular use for snapshots is reducing downtime for backup operations. Rather than using online data, you can back up your snapshots to enable critical transactions to continue during the backup process. You can also create frequent snapshots and, in the event of an error, execute a rollback operation to restore your data.

A snapshot uses less space than a split mirror component. The amount of space required by a snapshot is proportionate to the amount of data that changes after the snapshot is taken. The older a snapshot, the more likely the data it represents has changed; thus, the more space it requires. If the parent volume is expected to change substantially between snapshots, using split mirror components might be more appropriate because of the overhead of copy-on-writes.

Split Mirror Components

When you want an image of the data set represented in a mirror, you can split a mirror component from the mirror and save it as an independent copy or reference point of your data at that time. After you have analyzed the data in the split mirror component, you can rejoin it to the mirror, in which case it is synchronized to match the mirror's data set. Alternatively, if you want the other mirror components to be synchronized to match the data of the split component, you can reverse rejoin it to the mirror. Reverse rejoining a split mirror component is similar to rolling back a snapshot, because it causes the data set of the mirrored volume to match the data set of the point-in-time (PIT) image represented by the split component. One difference between using a split mirror component and using a snapshot rollback as an image of your data is that the mirroring process does not require snapshot reserve space.

You can rejoin the split component to the mirror, to synchronize it back to the mirror's data set, or you can perform a reverse rejoin action on the previously split component if you want the other mirror components to synchronize to its data. In effect, reverse rejoining a mirror component is similar to rolling back a snapshot. One difference in using a split mirror component as an image of your data is that it does not require snapshot reserve space. If the application will need to write to the split/snapshot volume, use a split component to avoid consuming snapshot reserve.

You can choose to use lower-cost legacy storage for the split mirror component when the mirror is created. Whereas a snapshot will need to read from the parent volume, a split component stands alone. If an application needs to write to the image, you can use a split component to avoid consuming snapshot reserve space.

One popular use for split mirror components is part of a rotating strategy used to track changes in your data set over a period of time. Suppose you want to add a fourth mirror component to your three-component mirror, as part of a strategy in which you rotate mirror components. Each day you add a mirror component, wait for it to resilver so that each mirror component is identical to the other, and then perform a split operation to save an image of the data at the time of the split.

Summary of Differences

The major differences in using split mirror components or snapshots as images of your data at a specific time are summarized in TABLE 5-1.

TABLE 5-1 Split Mirror Components Compared to Snapshots

Characteristic	Split Mirror Component	Snapshot
Point-in-time (PIT) image availability	A complete, independent copy of the data.	A copy of the data that is dependent on both the original volume component and the snapshot reserve space.
Capacity used	Each image requires storage for a complete copy of the original volume.	Storage required depends on how many changes have occurred on the parent volume since the snapshot was created.
Quantity of PIT images	Up to 3.	Up to 8.

TABLE 5-1 Split Mirror Components Compared to Snapshots *(Continued)*

Characteristic	Split Mirror Component	Snapshot
Time until PIT access	The resilvering process must run before the mirror can be split. The first time a mirror component is added, the entire volume must be copied and this can take a long time. The length of time the resilvering process needs after a rejoin depends on how many changes have occurred on the mirror since the split.	Ready to access as soon as the snapshot is created. This takes only a few seconds.
Capacity reclaim	Components of a mirror can be broken off and deleted to reclaim space.	Snapshot reserve space is reclaimed after the oldest snapshot is deleted. For more information, see FIGURE 5-3 on page 78.

Snapshot Policy

Based on your data usage, you can set the snapshot policy to favor the parent volume's data set or to favor snapshots you take of the data set.

- Favor Parent Volume Over Snapshots (default)

  This is the default policy and is the best practice for snapshots used for backup purposes. Select this policy for applications like data mining, to avoid problems with the snapshot reserve space that might interfere with business operations on the original data set. Running out of snapshot reserve space impedes the data mining application; however, normal operations continue. If the system runs out of snapshot reserve space, it continues to accept write commands to the parent volume, but all existing snapshots become invalid, and system activity pertaining to these snapshot images fails.

- Favor Snapshot Validity Over Volume

  Select this policy when it is imperative that you preserve the snapshot data. Since problems updating the snapshot reserve space cause all attempted activity performed on the parent volume and snapshots to cease until you can address the problem, this policy preserves the integrity of the snapshots.

  This policy prevents the snapshots from getting out of sync by failing any writes to the original data that cannot be tracked by snapshots. This ensures that the snapshots stay current, but might affect the availability of the original data.

Best Practices: Using Snapshots for Backups

This section describes the best practices (recommended processes) for using snapshots to facilitate tape backups of data. Different performance requirements, the availability of snapshot reserve space, and different backup clients might require different procedures.

The backup process for user data often requires the data to be offline so that a consistent copy can be made. With large storage capacities, the time required to back up data can be significant. This can lead to prolonged periods of unavailability.

Using snapshots to make backups can significantly decrease the time during which the data is unavailable compared with traditional backup methods. Through creation of an instantaneous point-in-time (PIT) image of the data, backup data can be copied from a snapshot. This process maintains a consistent copy of the data at the time the backup is started, even as the original data stays online and is modified.

A snapshot can be created quickly. The original data is unavailable only for as long as it takes to create the snapshot.

The following examples describe best practices for using snapshots for backups:

- "Example 5-1. Creating a New Snapshot for Each Backup" on page 83
- "Example 5-2. Reusing an Existing Snapshot for Backups" on page 84

Prerequisites

Ensure that data is synchronized when you are working with virtualization features. All client data must be flushed to disk so that the snapshot is fully synchronized with the expected data. The steps required to do this are application-specific but might require unmounting and remounting a volume, explicitly requesting a file system synchronization, or closing or suspending an application.

Note – If a snapshot's contents change due to a resnap operation, the backup client must be notified of the change.

Notes

When the snapshot feature is used, there are several issues to consider. The following issues pertain to using snapshots for backups:

- Snapshot reserve space

  The system uses snapshot reserve space to keep images consistent. This enables efficient snapshot creation, because data is copied only as needed after the original snapshot operation. However, for the snapshot operations to be successful, sufficient snapshot reserve space must be allocated.

- Snapshot failure policy

  For snapshot images to be maintained, snapshot reserve space operations and metadata updates must be performed as data is changed. If errors are encountered (such as running out of snapshot reserve space), either the snapshots must be abandoned or the original data cannot be updated, depending on the snapshot policy you have selected. Therefore, when you plan to use snapshots to back up your data, the best practice is to select the Favor Parent Volume Over Snapshots policy. This policy reduces downtime for the original data. If errors occur while the system attempts to maintain the snapshots (for example, if there is insufficient snapshot reserve space), the snapshots fail, but access to the original volume is preserved.

- If a snapshot fails, it must be deleted and re-created. It is no longer updated and might not contain the correct data.

  For more information, see "Snapshot Policy" on page 81.

Example 5-1. Creating a New Snapshot for Each Backup

This procedure requires relatively little snapshot reserve space. However, this approach might require additional configuration of the backup client, because the World Wide Name (WWN) of the snapshot changes for each backup.

Procedure

To do a backup with a new snapshot:

1. Synchronize the original data.
2. Create the snapshot. This creates a new volume with a unique WWN.

   For details, see "Creating a Snapshot Using the Browser Interface" on page 97.

3. Remount the original data if it was unmounted during synchronization.

   At this point, the original data is made available and can be used in a normal fashion.

4. Map the snapshot to the backup client.

5. Mount or gain access to the snapshot from the backup system. Backup applications differ, but it might be necessary to set up the backup application with the snapshot's WWN.

6. Start the backup procedure by copying data from the snapshot.

When the backup procedure is complete, the snapshot can be unmounted, unmapped, and deleted. As a best practice, your backup procedure might include a copy and verify phase.

Example 5-2. Reusing an Existing Snapshot for Backups

It might be simpler to preserve the WWN of the snapshot used for the backup so the backup application doesn't have to be reconfigured with a new WWN before each backup. Instead of creating and deleting a new snapshot for each backup, you can resnap the existing snapshot, enabling it to be reset to the current data without changing its WWN.

Note – It is best practice to resnap snapshots in the order in which they were taken, from oldest to newest.

There are some disadvantages to this procedure. When a snapshot is present, there is overhead incurred while the original data is accessed. If the snapshot is not deleted when the backup is completed, the overhead associated with maintaining it continues. In addition, you might consume more snapshot reserve space.

Prerequisites

Before starting this procedure, make sure that an existing snapshot has been created for backups and is mapped to the backup client.

Procedure

To do a backup with an existing snapshot:

1. Ensure the backup client recognizes data has changed in the snapshot when it is resnapped. This might require unmounting the snapshot from the backup client. Optionally, the snapshot can be un-mounted as described in Step 7 of this procedure.
2. Synchronize the original data, as described in "Prerequisites" on page 82.
3. Resnap the existing snapshot. This changes the snapshot's data to reflect the data currently in the parent volume while preserving the WWN.

Note – The resnap operation is quicker than using mirroring to update the data, which would involve rejoining a split component to the mirror and waiting for resilvering to take place before splitting the component again.

---

4. At this point, the original data can be used in a normal fashion. Remount the original data if it was unmounted during synchronization.
5. Complete the notification to the backup client with the information that the snapshot data has changed.
6. If the snapshot was unmounted earlier, remount it on the backup client. The snapshot should still be mapped to the backup client.
7. Start the backup procedure by copying data from the snapshot.
8. When the backup procedure is complete, unmount the snapshot in anticipation of repeating this procedure later.

Interoperability of Sun StorEdge Data Snapshot and Backup Software Products

Although many applications can continue to function while copy and backup operations take place, the amount of data and the time it takes to copy the information in some cases might exceed what an application can cache during its freeze state. To solve this problem, Sun Microsystems has worked with Computer Associates International and VERITAS Software Corporation to enable the Sun StorEdge 6920 system to work in conjunction with these companies' software backup products. As a result, the Sun StorEdge 6920 system can efficiently perform backup operations with high reliability.

The Sun StorEdge Data Snapshot software now works with the integration scripts of both Computer Associates BrightStor ARCserve Backup Agent for Oracle and VERITAS NetBackup software products to enable the workflow associated with copying and backing up Oracle data set volumes. Although the Oracle 9i Database software provides the fundamental capabilities needed to enable and simplify the process, it must operate with Sun StorEdge Data Snapshot software, in conjunction with either the VERITAS or Computer Associates backup products, to provide complete data integrity.

Many backup applications support both full and incremental backups. The incremental mode saves time and storage resources (such as tapes) by archiving only the changes since the last full backup. Although this might be a resource savings up front, recovery might be more complicated since restoring both the full and several incremental archives might be necessary. In addition, a means of determining what has changed since the last backup is needed. Often, backup applications track the changes between full and incremental backups by modifying the source data. For example, a file system backup application might set an i-node archival bit or update timestamps (ctime, mtime) to track incremental changes. When you perform backups from newly created or reset snapshots, such changes to the source data are not present at the start of the next backup.

Therefore, a backup solution using a tracking mechanism that is not dependent on modifying the source data is the best practice. Alternative approaches include cataloging backup information and using that to drive the criteria for incremental changes. One reason for performing incremental backups is to reduce the amount of downtime involved when the source data is taken offline for the duration of the backup procedure. By backing up data from snapshot devices, you greatly reduce the downtime and might negate the need for incremental backups. In other words, performing full nightly backups from snapshot devices are likely to be less intrusive than traditional offline incremental backups.

Performing effective and efficient backups is a non-trivial task and a discussion of all the issues involved is beyond the scope of this document. The key point is that using snapshots for backups requires a change to the approach used to archive data.

Best Practices: Working With Snapshot Reserve Space

For snapshot operations to be successful, you must allocate sufficient snapshot reserve space. The Sun StorEdge 6920 system also supports the expansion of snapshot reserve space, even with snapshots present. In addition, the snapshot reserve space can be allocated on separate storage devices. The following examples describe best practices for allocating and expanding snapshot reserve space:

- "Example 5-3. Allocating Snapshot Reserve Space" on page 87
- "Example 5-4. Expanding Snapshot Reserve Space" on page 90

Notes

- You determine the size of the snapshot reserve space appropriate for your business practice based on two factors: the number of snapshots that will be taken of the parent volume, and the anticipated percentage of changed data on the parent volume for each snapshot.
- To minimize the need for (or frequency of) expansion of the snapshot reserve space, slightly overestimate how much space you will require.
- There might be times when snapshots are no longer needed. In such cases, you can reclaim existing snapshot reserve space by deleting or resnapping snapshots in order from oldest to newest. When snapshot reserve space is reclaimed, it can be reused for other snapshots.
- The snapshot process involves the management and tracking of differences between the original data and the data in the snapshots. The Sun StorEdge 6920 system tracks these differences in groups of blocks or chunks to enable efficient snapshot creation and optimal usage of snapshot reserve space. Data is copied only as needed after the original snapshot operation.

Example 5-3. Allocating Snapshot Reserve Space

The most important thing to understand when managing snapshot reserve space is how much will be needed. The rate of change, locality of reference, and number of snapshots all affect the amount of snapshot reserve space you will need for your business practices. Because such usage characteristics can be difficult to estimate and might change as your business needs change, the Sun StorEdge 6920 system also enables you to expand snapshot reserve space later, after the initial allocation.

Prerequisites

Before starting the allocation process, plan your snapshot reserve space carefully. For more information, see "Planning Snapshots and Snapshot Reserve Space" on page 76.

Notes

- Estimate desired size
  - You need to estimate the amount of snapshot reserve space needed, how it will be consumed, the expected rate of change, the locality of reference, and the number of snapshots that will reside within it.
  - There are limits on the number of times snapshot reserve space can be extended. Any concatenated device in the system can have at most 16 partitions. Keep this in mind when estimating needed space and performing extensions.

- It might be better to slightly overestimate capacity needs to avoid running out of options. However, having huge amounts of snapshot reserve space might slow performance during snapshot reserve space operations.
- If you expect 100 percent of data to be copied for each snapshot, using mirroring with split or broken components might be a more suitable approach. For more information, see "Working With Mirrors" on page 51.

- Choose appropriate storage
  - The location and class of storage used for snapshot reserve space are important considerations. Because logging and metadata updates might be required with snapshots present, performance can be impacted by this additional overhead. Choosing storage that is efficient can be crucial to reducing undesired latency.
  - The specific configuration and connectivity have an impact. It is best to provision snapshot reserve space so it is accessed from the same Storage Service Processor as the original data. Not doing so creates additional overhead because operations are passed between processors. You can use the same virtual disk as the original data for snapshot reserve space. Alternatively, you might designate a separate target that could improve performance, so long as it is accessed by the same processor.

Procedure

To allocate snapshot reserve space:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Logical Storage > Volumes.

   The Volume Summary page is displayed.

2. Select the volume for which you want to allocate snapshot reserve space.

   The Volume Details page is displayed.

3. Click Snapshot Reserve.

4. Specify the settings you want for the following fields:
   - Storage Pool
   - Space Allocated For
   - Expected Write Activity

The following table might be helpful in estimating your snapshot reserve requirements.

TABLE 5-2 Snapshot Reserve Space Requirements for Expected Write Activity

Parent Volume Size	Write Activity	Percentage of Space Allocated	Number of Snapshots	Snapshot Reserve Space Size
10 gigabytes	Low	10%	2	2 gigabytes
10 gigabytes	Very Little	25%	2	5 gigabytes
10 gigabytes	Little	40%	2	8 gigabytes
10 gigabytes	Average	50%	2	10 gigabytes
10 gigabytes	High	75%	2	15 gigabytes
10 gigabytes	Full	100%	2	20 gigabytes

5. In the Snapshot Reserve Space section of the page, select the snapshot policy you want:
   - Favor Parent Volume Over Snapshots (default)

     This is the default policy and is the best practice for snapshots used for backup purposes. If the system runs out of snapshot reserve space, it continues to accept write commands to the parent volume, but all existing snapshots become invalid, and all system activity pertaining to these snapshot images fails.
   - Favor Snapshot Validity Over Volume

     Select this policy when it is imperative that you preserve the snapshot data. This ensures that the snapshots stay current, but might affect the availability of the original data.

     For more information, see "Snapshot Policy" on page 81.
6. Click Save.

Example 5-4. Expanding Snapshot Reserve Space

You can expand the existing snapshot reserve space while the original data and snapshots are online and under I/O load. There is no need to halt I/O, unmount volumes, or perform host synchronization (as is typically necessary with other virtualization operations).

Prerequisites

In order to expand snapshot reserve space, you must have additional capacity available on the array volumes from which the parent volume is constructed.

Notes

- The most conservative approach is to allocate 100 percent for each desired snapshot. Although this consumes disk space, it helps ensure that even if every block in the parent is altered, all snapshots are preserved.
- You can expand the snapshot reserve space of a mirrored volume as you would that of any other volume.

Procedure

To expand snapshot reserve space:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Logical Storage > Snapshots.

   The Snapshot Summary page is displayed.

2. Select the parent volume for which you want to expand snapshot reserve space.

   The Volume Details page is displayed.

3. Click Snapshot Reserve.
4. Increase the values of the settings, as appropriate, for the following fields:
   - Space Allocated For
   - Expected Write Activity
5. Click Save.

Rolling Back Snapshots

When you roll back snapshot data, you copy data from a snapshot to the parent volume, resetting the volume's data to become identical to that of the snapshot. This feature is particularly helpful if there are issues with a volume's data.

You start the rollback process by selecting the snapshot that you want to roll back; this need not be the newest snapshot.

The snapshot reserve space data can be reclaimed as it becomes unused by other snapshot images when the rollback completes. During the rollback process, additional snapshot reserve space might be consumed to keep the older images up to date.

The snapshot rollback feature allows for successful completion of scheduled jobs (such as cron jobs) even when a rollback is in progress. You can also create snapshots during a rollback operation.

During a rollback, you cannot perform read/write activity on the parent volume or snapshots associated with it. In addition, a rollback operation might take a significant amount of time to complete, depending on the amount of data that is being copied from the snapshot back to the parent volume. When the rollback operation is complete, the create time for the rollback snapshot volume is updated to reflect the time when the rollback operation started.

It is a good idea to make a backup copy of data before issuing a rollback. During a rollback operation, the parent volume data is in flux; it is a composite of the parent volume and the rollback snapshot's data. The parent volume is dependent upon the snapshot reserve space, the metadata, and the snapshot log. Failure of any of these can lead to loss of the original data for the parent volume and loss of the snapshot being rolled back.

Note – During the rollback process, the parent volume is offline and unavailable to the host.

When performing a rollback, the best practice is to select the default policy, Favor Parent Volume Over Snapshots. With this policy, the system attempts to preserve the volume's data if it encounters an error during a rollback operation. If the system runs out of snapshot reserve space, it continues to accept write commands to the parent volume, but all existing snapshots become invalid, and all system activity pertaining to these snapshot images fails.

The rollback can continue copying data from the snapshot reserve space to the parent volume.

Note – If you select the other policy, Favor Snapshots Over Parent Volume, and the system then encounters an error, the rollback process is impeded.

Example 5-5. Rolling Back a Snapshot

When you roll back snapshot data, the data for the original (parent) volume is reset to be identical to the snapshot.

Notes

- Rolling back a snapshot for a large parent volume might take a significant amount of time.
- You cannot roll back a snapshot of a mirrored or replicated volume, with these exceptions: If you break a mirror down to a single volume, you can rollback its snapshots. On a primary volume during replication, you can roll back a snapshot if the replication set is deleted. In both these cases, you can re-create the mirror or replication set after the rollback is complete.

  For more information about using snapshots with data replication, also see "Using Snapshots With Data Replication" on page 122.
- Unless you are planning to delete the parent volume and snapshots associated with it, you should not cancel a rollback operation that is in progress. A canceled rollback operation causes the data in the volume and snapshot reserve space to be inconsistent and the condition of the canceled rollback snapshot to be "Non Recoverable Error."

Browser Interface Procedure

To initiate the rollback process:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Logical Storage > Snapshots.

   The Snapshot Summary page is displayed.
2. Select the snapshot that you want to roll back.

   The Snapshot Details page is displayed.
3. Click Rollback.
4. When prompted, click Yes to confirm that you want to roll back the data.

   A rollback job is created, and the Snapshot Summary page is updated to show that the snapshot's condition has changed to Snapshot Rollback.

CLI Procedure

Assuming that `archive` is a volume that exists with mapped initiators, and the snapshot volume you want to roll back is `archive_snap1`, you could run the following command on the administration to execute a snapshot rollback operation from the `archive_snap1` snapshot volume to the master `archive` volume.

```
sscs snapshot -b(--rollback) [-S(--sdomain) <storage_domain>]
volume <snapshot>
[admin_host]# sscs snapshot -b -S DatanDomain volume archive_snap1
Job 2005.01.20_14:31:19.571_1 started: PITRollback
[admin_host2]#
```

To monitor the rollback job:

```
[admin_host]# sscs list jobs
Job ID: 2005.01.20_14:31:19.571_1
Job ID: 2005.01.17_16:38:15.126_0
# sscs list jobs 2005.01.17_16:38:15.126_0
Job ID: 2005.01.17_16:38:15.126_0
Description: PITRollback
% complete: 100
Start: Wed Aug 11 16:38:15 MST 2005
```

Managing Snapshots

This section describes the following procedures for managing snapshots:

- "Displaying Snapshot Information" on page 94
- "Identifying Snapshots Using CLI Commands" on page 95
- "Using Snapshots With Oracle Databases" on page 98

For other procedures, see the online help.

Displaying Snapshot Information

To display snapshot information:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Logical Storage > Snapshots.

   The Snapshot Summary page displays information about all snapshots in your system.

2. Select a snapshot to see its detailed information.

   The Snapshot Details page displays information about the selected snapshot.

Identifying Snapshots Using CLI Commands

To identify the snapshots, list the volumes in the `Data`*n*`Domain` storage domain by volume name. In this example, use `redo_01 redo_01_snap1`. The output lists the volume name, the WWN (GUID), and the State as `Mapped`. For the snapshot volume, the Type is `Snap`.

```
[admin_host]# sscs list volume redo_01
Volume: redo_01
Description: redo_01
Creation Date: Thu Aug 12 00:00:00 MST 2004
WWN: 600015D00020E1000000000000006946
Storage Domain: Data2Domain
Pool: r5pool
Profile: r5profile
Size: 80.0000 MB
State: Mapped
Condition: OK
Type: Stripe
Snapshot Policy: Favor original volume over snapshots
Snapshot Pool: r5pool
Space Allocated For: 2 Snapshots
Expected Write Activity: Average(50%)
Snapshot Reserve Size: 100
Snapshot Reserve Percent Full: 5
Snapshots:
Snapshot: redo_01_snap1 Creation Time: Wed Aug 11 14:58:31 MST2005
Associations:
Initiator: host2_hba1 LUN: 21 Permission: Write/Read Map State:
Online
Initiator: host2_hba2 LUN: 21 Permission: Write/Read Map State:
Online
[admin_host]# sscs list volume redo_01_snap1
Volume: redo_01_snap1
Description:
Creation Date: Wed Aug 11 14:58:31 MST 2005
WWN: 600015D00020E1000000000100006946
Storage Domain: Data2Domain
Pool: null
Profile: null
Size: 80.0000 MB
State: Free
Condition: OK
Type: Snap
Associations:
```

The WWN of a snapshot differs by only one numeral from the WWN of the original volume. For example, the WWN of `redo_01` is as follows:

`600015D00020E10000000000000006946`

The WWN of `redo_01_snap1` differs by just one numeral, the snapshot identifier in the ninth position from the right:

`600015D00020E10000000000100006946`

In this example, `redo_01_snap1` is the first snapshot. If an additional snapshot is taken, the snapshot identifier will increase by one, as follows:

`600015D00020E10000000000200006946`

Note – The snapshot identifier number never gets smaller. Though it will increase as additional snapshots are taken, it will not decrease when snapshots are deleted.

Note – It is possible to assign different LUN numbers to a single volume when you map it to different initiators, but using this process is not a best practice.

Creating a Snapshot Using the Browser Interface

To create a snapshot:

1. In the system's browser interface, click Sun StorEdge 6920 Configuration Service > Logical Storage > Volumes.

   The Volume Summary page is displayed.

2. Select the check box to the left of the volume for which you want to create a snapshot.

3. Click Snapshot.

   The New Snapshot wizard is displayed.

4. Follow the directions in the wizard.

   Click Help for more information.

   Upon completion of the new snapshot, a confirmation message is displayed.

Note – When you create a snapshot from the Volume Summary page, the system automatically names it (such as snap050301164439). To change its name, start by selecting the volume for which you created the snapshot. On the Volume Details page for that volume, go to Additional Information, and click Snapshots. Then select the snapshot and change the name.

Creating a Snapshot Using the CLI

Following is an example of how to create a snapshot volume using the `sscs` command line. Assume that `redo_01` is a volume that exists with mapped initiators, and that you are running the following command on the administration host where the SSCS CLI is installed.

Example

```
[admin_host]# sscs snapshot -Pr5pool -vredo_01-C2-L average -S
DatanDomain \volume redo_01_snap1
```

The `-v` option specifies the volume to be snapped, the `-C` option specifies the number of snapshots to support, and the `-L` specifies the read/write activity expected. The last portion of the command specifies the name of the snapshot volume. When the command completes, you can map the snapshot to initiators, and then mount the snapshot as you would any other volume.

Using Snapshots With Oracle Databases

The snapshot application works at a volume level, independent of higher-level applications. Therefore, to maintain data integrity for other operations, ensure that the Oracle database is in a known state before creating the snapshot. To do this, place the Oracle database in cold backup or hot backup mode, so that if an Oracle instance is started on the snapshot, the database is in a consistent state.

The Oracle database needs to be quiesced only long enough for a snapshot to be created. To take the snapshot, the system briefly suspends I/O to the parent volume, establishes the necessary pointers for the snapshot, and then resumes I/O to the parent volume. For more information on using snapshots with Oracle databases, see the *Sun StorEdge Data Snapshot Software With Oracle Databases Usage Guide.*

Troubleshooting Non-Recoverable Snapshots

The Storage Automated Diagnostic Environment generates alert messages at 5 percent intervals from the time when the reserve space reaches 70 percent full until the space is 100 percent full.

When the snapshot reserve space allocated for the snapshot reaches the 100 percent full state, all snapshots for that primary volume become non-recoverable. The snapshot volumes and the data within the snapshots are no longer available.

If the reserve space becomes 100 percent full, all snapshots for that primary volume must be unmapped and deleted in order for the snapshot reserve space to be recovered. For more information, see the online help.

CHAPTER 6

Working With Data Replication

The Sun StorEdge Data Replicator software offers a volume-level and volume set (consistency group) replication mechanism to protect your data. You can use this software to replicate volumes between physically separate primary and secondary Sun StorEdge 6920 systems in real time. The software is active while your applications access the data volumes, and it continuously replicates the data to the remote site.

This chapter provides instructions on working with data replication. It contains the following sections:

Understanding Data Replication

Data replication has traditionally been employed in disaster recovery and business continuance strategies to provide redundant storage of critical information across physically separate sites. In addition to information protection, the investment in multiple data copies can now be leveraged into new information-sharing opportunities across distributed enterprises. These opportunities can lead to operational efficiency for physically disparate computing environments and yield a strategic advantage due to increased information access.

Data replication offers the following benefits:

- Data protection and disaster recovery

  As part of a disaster recovery and business continuance plan, the software enables you to keep up-to-date copies of critical application data at remote sites. You can also rehearse a data recovery strategy to fail applications over to remote sites. Later, you can update the remote site with any changes that occurred on the production data set during the rehearsal, as well as restore any data that was changed on the remote site as part of the rehearsal.

- Near-real-time online backup

  It is standard practice to perform regularly scheduled data backups. However, the time periods that are available for performing these backups has decreased to support increased operating hours, while the applications have become larger. By using replicated data on the secondary server for backup processing, you can keep the business-critical resource at the production location available during the required backup processing.

- Reduced downtime during data migration

  Data replication features can significantly reduce downtime during data migration, such as when you upgrade or move your data center. In such cases, it is common to install and ready a new server while the existing server continues to provide service. However, even with parallel systems, the downtime involved in copying, transporting, and restoring data from an old server to a new server can be lengthy. You can significantly reduce this downtime by using Sun StorEdge Data Replicator software to continuously replicate changes to the new server.

How It Works

The Sun StorEdge 6920 system replicates data from a primary site volume to a secondary site volume. The association between the primary and secondary volumes constitutes the replication set. After the volume in a replication set is initially synchronized, the software ensures that the primary and secondary volumes contain the same data on an ongoing basis.

Note – It is best practice to maintain the same RAID type for the primary and secondary replication sites.

When replicating data, the software preserves write order consistency. That is, the software ensures that write operations to the secondary volume occur in the same order as the write operations to the primary volume.

If you need to ensure write order consistency across multiple volumes, such as for an application that builds its database on multiple volumes, you can place multiple replication sets into a consistency group. A consistency group enables you to manage several replication sets as one, because it is a collection of replication sets with the same primary and secondary roles and modes for the primary and secondary sites. By using a consistency group, the software maintains write ordering for volumes in a group to ensure that the data on all secondary volumes is a consistent copy of the corresponding primary volumes.

The software transports data between the two Sun StorEdge 6920 systems, using either a Fibre Channel (FC) connection or Gigabit Ethernet network link (replication link).

FIGURE 6-1 shows data replication over a network link.

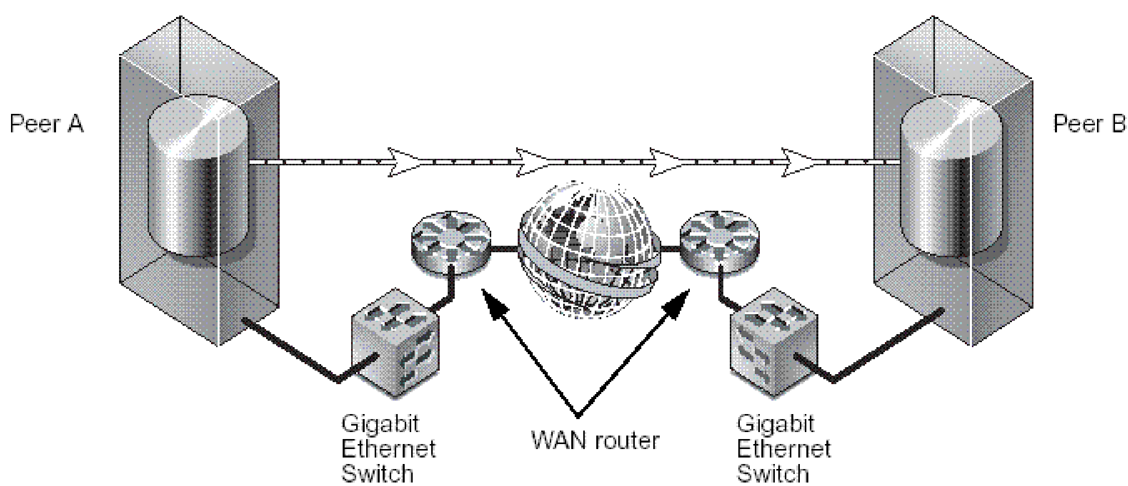

FIGURE 6-1 Data Replication Over a Network Link

The means by which data is transported between the two Sun StorEdge 6920 systems is either synchronous replication or asynchronous replication. During synchronous replication, data will be committed to storage at both the primary site and the secondary site before a write to the primary volume is acknowledged. During asynchronous replication, write operations are written to the primary site and to the asynchronous queue on the primary site. The asynchronous queue then forwards queued writes to the secondary site as network capabilities permit.

Note – The Sun StorEdge 6920 system does not provide built-in authentication or encryption for data traveling outside your data center over a long-distance replication link. It is assumed that customers implementing remote replication strategies with multiple Sun StorEdge 6920 systems replicate the data over secure leased lines or use edge devices to provide encryption and authentication. For help setting up appropriate security, contact Sun Professional Services.

If there is a break in the network or the secondary peer is unavailable, the software automatically switches to suspended mode, in which it ceases replication and tracks changes to the primary peer in the replication bitmap. When communication is re-established, the software uses the replication bitmap to resynchronize the volumes and then returns to replicating the data.

As a failover technique, you can initiate a role reversal of the primary and secondary peers so that the software restores data from a secondary volume to a primary volume. The application software accesses the secondary volume directly until you can correct the failure at the primary peer. The old, failed, or primary volume can function as a new secondary volume until you coordinate another role reversal to get back to the original configuration.

Note – Ensuring the primary and secondary volumes are the same size enables role reversal.

Replication Sets

The system administrator at each site must create and configure a replication set on each system. The replication set definition for the secondary peer must be equivalent to the replication set definition for the primary peer. The role of the secondary peer is different in the two corresponding replication sets, as are the remote volume and peer node descriptors.

A replication set includes the following:

- A volume residing on a Sun StorEdge 6920 system and a reference to a volume residing on another, physically separate, Sun StorEdge 6920 system. One system is the primary peer that replicates the data. The other system is the secondary peer that is the recipient of the data.

- The role that the peer plays within the replication set, either as a primary or as a secondary peer.
- A replication bitmap volume on each system to track write operations and differences between the volumes. The primary volume's bitmap records write actions issued at the primary peer. The secondary peer's replication set also includes a replication bitmap in case you initiate a role reversal and the secondary peer becomes the primary peer. The replication bitmap defines the differences between the primary and secondary peers. This enables the software to resynchronize only the blocks that have changed since the last synchronization.
- If you choose asynchronous mode replication, an asynchronous queue associated with each set, or one queue associated with all sets in a consistency group. The asynchronous queue associated with the primary peer is used to buffer remote write operations destined for the secondary peer. The queue associated with the secondary peer is not used until a role reverse results in the replication set taking on the role of primary peer.

You can update the secondary volumes 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. However, you might choose to copy the primary to the secondary with a tape Fast Start or an original full synchronization. As application I/O operations are written to the primary volume, the Sun StorEdge Data Replicator software copies the changes from the primary volume to the secondary volume, keeping the two images consistent.

FIGURE 6-2 shows the relationship between the two peers and their corresponding replication sets.

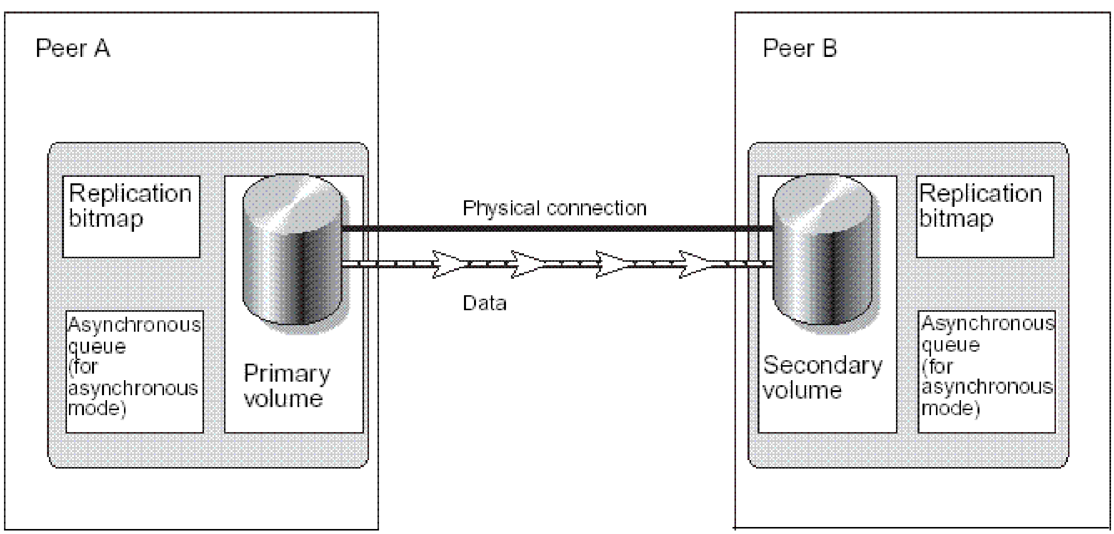

FIGURE 6-2 Peer and Replication Set Relationship

After you create a replication set, you can perform volume operations, such as the following:

- Extending replicated volumes by adding storage first to the secondary volume and then to the primary volume. The system increases the size of the replication bitmaps accordingly. The size of the replication set is dictated by the primary volume.

Note – If the primary or secondary volume in a replication set is a legacy volume, you cannot extend the volume.

- Creating a snapshot of the volume within the replication set.

  You can do this at any time. You can also enable the Snap With Primary or Snap Before Synchronization option so that the software creates snapshots on both systems before synchronizing the primary and secondary volumes, or so that the software creates a snapshot of the secondary volume whenever it creates a snapshot of the primary volume. For more information, see "Using Snapshots With Data Replication" on page 122.

Note – Any link outage puts the primary set into suspended mode, causing the ensuing update synchronization not to preserve write ordering at the secondary site. A best practice is to create a snapshot on the secondary volume before initiating the update synchronization. This ensures that there is a consistent data set if a primary site failure occurs while the update is in progress.

- Replicating user data to and from a legacy volume.
- Deleting a replication set. When you delete a replication set, you remove the association between the primary and secondary volumes, and the volumes revert to independent volumes.

If the primary volume becomes unavailable, the secondary volume can assume the role of primary volume. This role reversal enables applications to continue their operations by using the newly designated primary volume. When the former primary volume is again available, you must synchronize it with the more recent data on the other volume to restore the functions of the replication set pair.

Consistency Groups

To ensure write order consistency across multiple volumes, you can group multiple replication sets into a consistency group. A consistency group is a collection of replication sets that have the same group name, primary and secondary roles, and replication mode. A consistency group can be associated with only one primary path and one alternative path.

When you perform an operation on a consistency group, the operation applies to all the replication sets, and consequently their volumes, in the consistency group. If you make a change to a consistency group, the change occurs on every replication set in the consistency group; if an operation fails on a single replication set in the consistency group, it fails on every replication set in the consistency group.

When you configure a consistency group, the system preserves write ordering among the volumes in the replication set's consistency group. Because you control the replication set's consistency group as a single unit, data replication operations are executed on every member of the consistency group. Write operations to the secondary volume's consistency group occur in the same order as the write operations to the primary volume's group. By using a consistency 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.

Note – The system automatically detects mixed groups, in which replication modes are asynchronous for one replication set and synchronous for another replication set, and changes the mode of the added set to match that of the existing members in the consistency group.

Reversing Primary and Secondary Roles

If the primary and secondary volumes are the same size, you can reverse roles in a disaster or link failure to provide access to your critical data. The software can run on what was originally the secondary peer. When the original primary peer failure has been rectified, another role reversal returns both peers to their original roles.

It is a best practice to flush the file system's cached data on the primary site before a role reversal. Otherwise, the file system will need to undergo a file system check before it can be mounted on the secondary site.

If your organization has licensed snapshot as well as data replication features, a best practice is to create snapshots of the secondary volume regularly in case its data becomes inconsistent due to an in-progress or terminated synchronization operation. If you have a snapshot available, you can perform a rollback operation from the snapshot to restore the data after you have removed the replication set.

Caution – If I/O operations to the primary and secondary peers occur during a role reversal, data is lost when you initiate a synchronization operation. That is, the current secondary volume is brought into synchronization with the current primary volume, and any write operations that have been written to the secondary volume are lost.

Data Replication Architecture

FIGURE 6-3 illustrates the data replication architecture of the Sun StorEdge 6920 system.

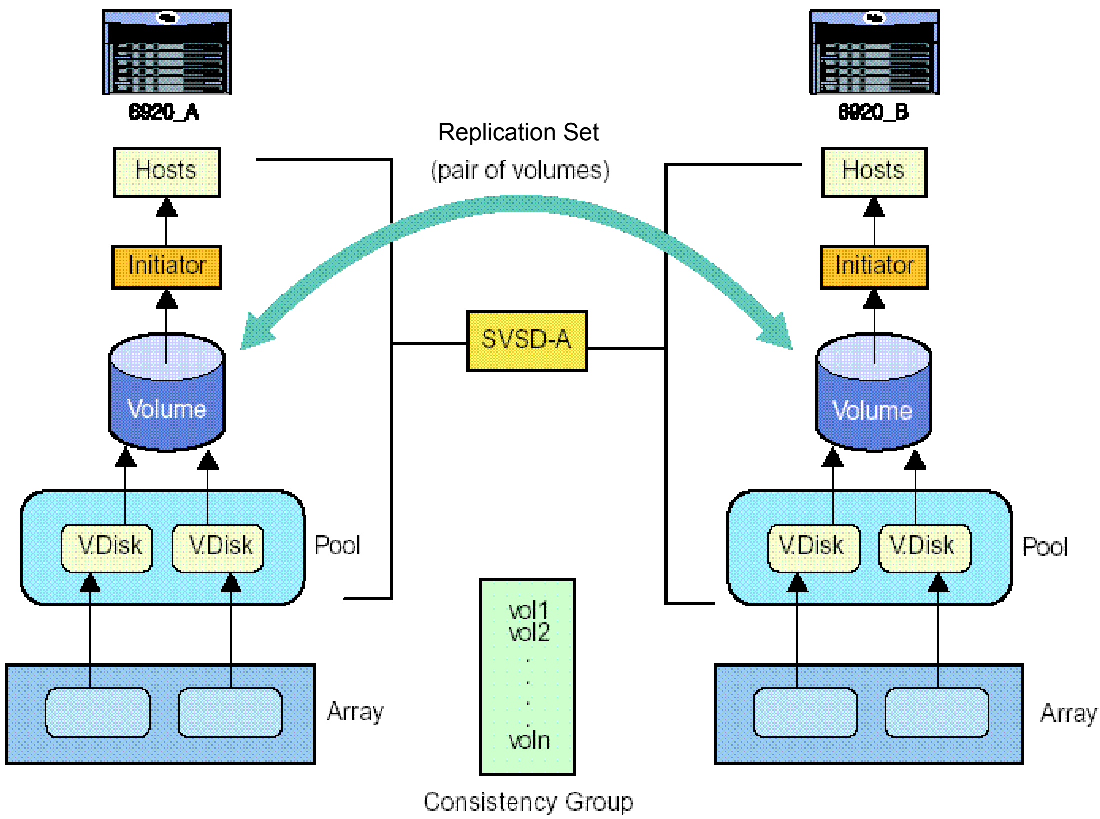

FIGURE 6-3 Sun StorEdge 6920 System Data Replication Architecture

FIGURE 6-4 shows the requirements for Sun StorEdge 6920 Data Replicator software.

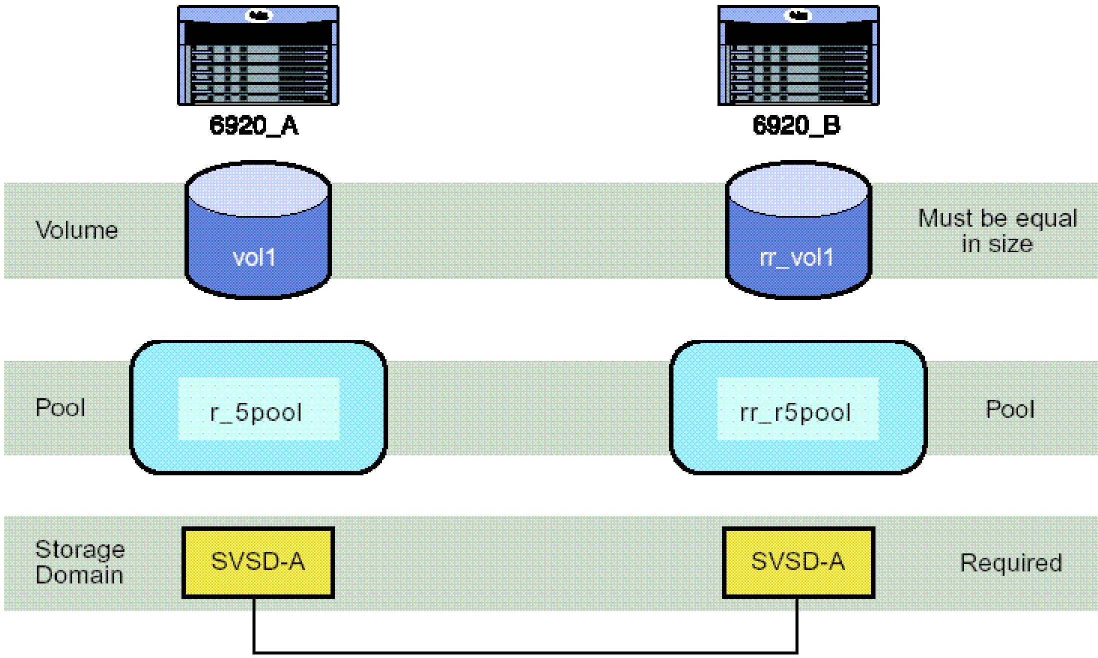

FIGURE 6-4 Data Object Requirements for Replication

A replication link is a logical and physical connection between two Sun StorEdge 6920 systems that enables data replication. A replication link transports data between the primary and secondary peers. This link transfers data as well as replication control commands.

You can use either Fibre Channel (FC) or Gigabit Ethernet ports for data replication. You must enable the same types of ports on both systems and verify connectivity to establish the replication link.

Note – You can configure only two replication links at a time. You cannot mix port protocol types.

If you use FC ports for data replication, you must configure any FC switches that you use to make the connection to the remote site for long-distance operations, and you must apply zoning practices, as shown in FIGURE 6-5.

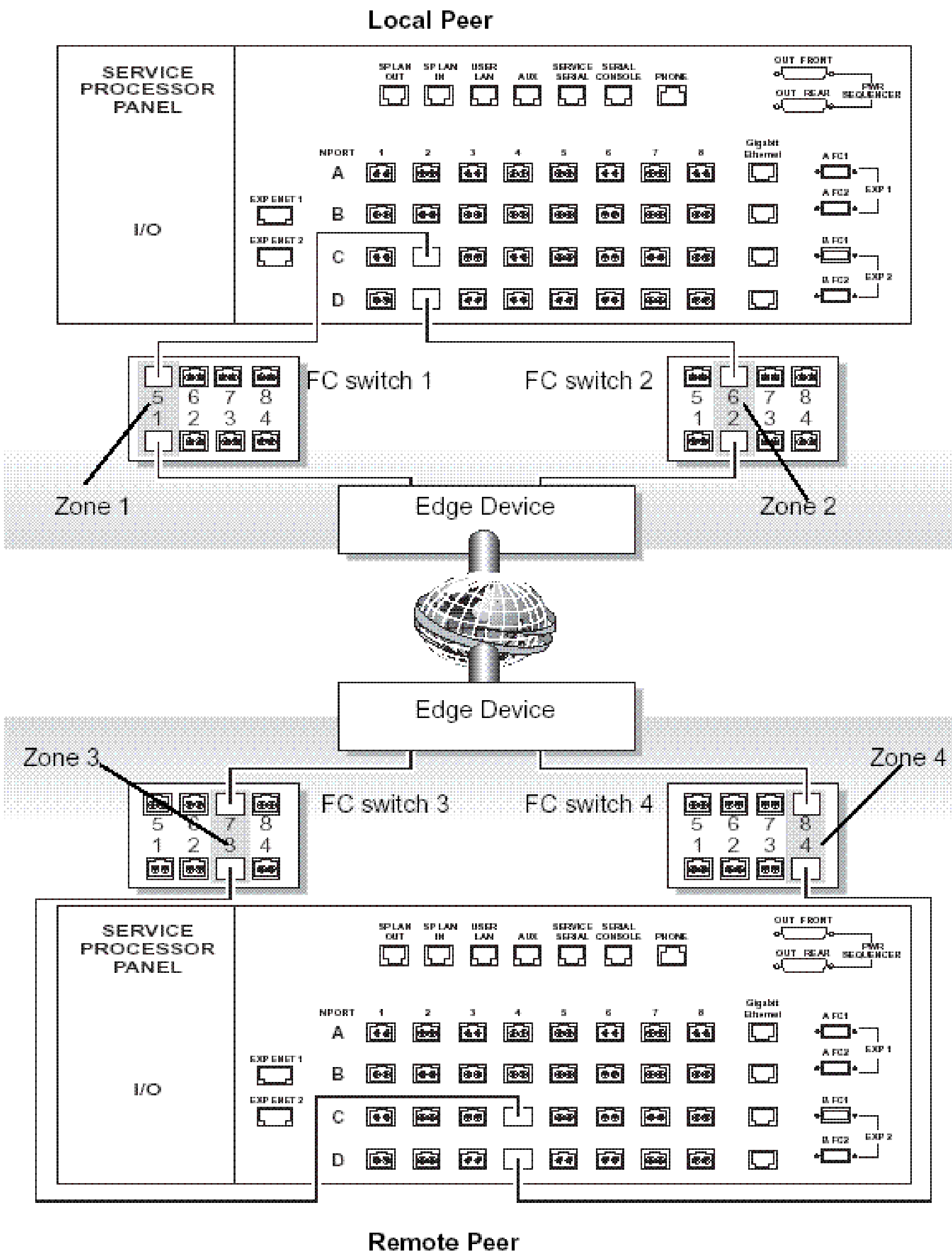

FIGURE 6-5 Zone/Switch Configuration on Local and Remote Peers

Planning for Data Replication

Replicating data and modifying replication options and operations make a significant change to your system's configuration. You must plan accordingly before you perform data replication operations.

Consider the following:

- Business needs – When you decide to replicate your business data, consider the recovery point objective (RPO), or how far behind the data on the secondary peer can become. Additionally, it is important to determine the recovery time objective (RTO), procedures to be implemented, in order to ensure that business objectives are met at the time when data integrity becomes jeopardized.
- Application write load – Understanding the average and peak write loads is critical to determining the type of network connection required between the primary and secondary peers. To make decisions about the configuration, collect the following information:
  - Average rate and size of data write operations

    The average rate is the number of data write operations that occur while the application is under typical load. Application read operations are not important to the provisioning and planning of your remote replication.
  - Peak rate and size of data write operations

    The peak rate is the largest amount of data written by the application over a measured duration.
  - Duration and frequency of the peak write rate

    The duration is how long the peak write rate lasts, and the frequency is how often this condition occurs.
- Network characteristics – The most important network properties to consider are the network bandwidth and the network latency between the primary and secondary peers.

Note – The system supports asynchronous and asynchronous replication of up to 128 volumes.

Data Replication Tasks Checklist

TABLE 6-1 provides a checklist of data replication tasks.

TABLE 6-1 Data Replication Tasks Checklist

✔	To Do
	Consider network latency, bandwidth, and security when you choose the path through the storage area network (SAN) or wide-area network (WAN) that you will use for data replication. **Note:** As the software synchronizes the volumes within a replication set, there might be a large number of additional I/O operations, which might impact performance.
	Make sure that the remote Sun StorEdge 6920 system is available to the your system by means of an existing physical link.
	Log on to the system using the `storage` user role.
	Configure both the primary and secondary volumes as you would any other volume, and make sure that they have identical configurations. The capacities of the volumes must be identical.
	From the Volume and Port Summary pages, record the following: • WWN of both the local and remote Sun StorEdge 6920 systems nodes (located on the table title bar) • WWN of both the local and remote volumes that will make up the replication set • IP addresses of both the local and remote ports, if you are using a Gigabit Ethernet port **Note:** If you enter an incorrect WWN in the Create Replication Set wizard, you must delete the replication set and create a new one.
	For configurations that use FC ports, configure the FC switches that provide the connection between peers for long distance operations. See the FC switch vendor's documentation for information about operating over long distances.
	On both the local and remote Sun StorEdge 6920 systems, enable the FC or Gigabit Ethernet ports that you plan to use for replication. If the port is a Gigabit Ethernet port, complete the Enable Gigabit Ethernet Replication Port wizard at both sites to create a replication link.
	Before creating a replication set, consider the following: • Replication roles, modes, and priorities. • Snapshot requirements. • The replication peer port must be initialized before the replication set can be initialized.
	Use the Create Replication Set wizard to do one of the following: • Create a replication set • Create a replication set and add it to an existing consistency group • Create a consistency group at the same time as you create a replication set
	Make sure that the replication sets and consistency groups for both peers are configured identically.

TABLE 6-1 Data Replication Tasks Checklist *(Continued)*

✔	To Do
	Consider the different operations involved with data replication: • Normal synchronization operation • Full synchronization operation • Normal suspend operation • Fast Start operation • Role reversal
	Consider the following functional parameters: • There can be a maximum of 128 replication sets per system. • There can be a maximum of 8 consistency groups per system. • Replication volumes support inline snapshots. • The system can have replication volumes that are either synchronous or asynchronous. • Consistency groups can be either synchronous or asynchronous. They cannot have a mix of both types of volumes. • Two-way replication is supported.
	Consider the following topology limitations: • A maximum number of 2 DSP ports can be configured for replication (1 primary path and 1 alternate path). • A maximum of 2 Sun StorEdge 6920 systems can be in a data replication mesh. • The system does not support the use of both Ethernet replication and FC replication. You must implement one or the other. • During replication, a port on the same storage processor can only be used for host I/O but not storage.
	Consider the following operational constraints: • Mirrors cannot be replicated. • Snapshots cannot be replicated. • You must put all replication into suspended mode before initiating a rolling upgrade.

Planning to Create Replication Sets and Consistency Groups

To replicate data to a secondary peer, you must first run the Create Replication Set wizard to create replication sets and consistency groups. The Create Replication Set wizard enables you to create a consistency group at the same time that you create a replication set.

Before you run the wizard, you must do the following:

- Make sure that you have configured volumes and enabled ports for data replication on both systems.
- Coordinate with the system administrator at the remote site to ensure that you both define volumes, replication sets, and consistency groups with properties that complement each other appropriately.
- Determine the World Wide Name (WWN) for the remote peer and the secondary volume. You can find these WWNs on the Volume and Port Summary pages for the remote peer.
- Determine whether you need to preserve write order consistency across volumes or want to manage multiple volumes as a group.

Note – It is best practice to add a volume to a group only when it is essential that write ordering is preserved for the group of volumes or when the I/O performance of the volumes is not top priority.

- Define the replication set or consistency group:
  - Assign the primary role to the replication set or consistency group, and ensure that the system administrator at the remote site assigns the corresponding replication set or consistency group with the secondary role.
  - Configure the consistency group as synchronous or asynchronous replication mode, in line with the business requirements already determined.

    For more information, see "Replication Modes" on page 117.
  - If you choose asynchronous mode, determine the asynchronous queue parameters and consider how they might affect performance.
  - Determine whether you want the system to assign a higher or lower priority to the rate of synchronization relative to I/O activity. A high synchronization rate might lead to degraded I/O performance, so you can choose a faster synchronization rate if your environment can sustain slower I/O activity. Conversely, you can choose a slower synchronization rate to ensure quicker I/O activity.
  - Determine whether you want the system to synchronize the volumes on both peers whenever there is a working link, or whether you want to initiate synchronization operations manually.

- Determine whether you want to create snapshots whenever you synchronize the volumes on both peers. If you do, select the Snap Before Sync option in the wizard. You must also ensure that you have allocated sufficient snapshot reserve space.
- Determine whether you want to take simultaneous snapshots of the primary and secondary volumes. If you do, select the Snap With Primary option in the wizard. For more information, see "Using Snapshots With Data Replication" on page 122.
- Determine the pool from which you want to create the replication bitmap. For asynchronous replication, also determine the pool for the asynchronous log.

Preparing to Create a Replication Set

Before you create a replication set, consider the factors described in TABLE 6-2.

TABLE 6-2 Replication Set Prerequisites

Topic	Issues
Ports	You must enable the ports on both the local and remote Sun StorEdge 6920 systems for data replication before you run the Create Replication Set wizard. If the port is a Gigabit Ethernet port, you must also create a replication link.
Storage domains	All elements of a replication set (the local volume, the replication bitmap, and the optional asynchronous queue) must be from the same storage domain. **Note:** A replication set is automatically associated with the storage domain of the local volume.
Volumes	• You can associate replication sets with volumes. • Each replication set contains one volume. • Volumes must be in storage domains that are named the same on each system. **Note:** Though the storage domains on both systems have the same name, they are not the same storage domain. • You can use simple or legacy volumes to create a replication set. **Note:** You cannot use a mirrored volume to create a replication set. • You cannot use a snapshot to create a replication set, but you can use a volume that includes snapshots. • A volume can be the primary or secondary volume for only one replication set. • The size of the secondary volume must be equal to or greater than the size of the corresponding primary volume. If you initiate resynchronization on a replication set in which the secondary volume is smaller than the primary volume, the operation fails.

TABLE 6-2 Replication Set Prerequisites *(Continued)*

Topic	Issues
Replication sets	• A replication set can be configured to communicate with only one peer system. • A replication set can be a member of only one consistency group at a time. • A replication set must be configured with the same role as the consistency group to which you want to add it. • The replication set specifications at the primary and secondary peers must match and complement each other. • A system that includes two storage resource card (SRC) sets supports up to 128 replication sets. **Note:** Multi-hop and one-to-many data replication are not supported.

Preparing to Create a Consistency Group

Before you create a consistency group or place a replication set in a consistency group, consider the factors described in TABLE 6-3.

TABLE 6-3 Consistency Group Prerequisites

Topic	Issues
Write order	If you need to ensure write order consistency for a set of volumes, you must place replication sets in a consistency group. You can create the consistency group when you create the first replication set, or you can add an existing replication set to the consistency group from the replication set's Details page.
Volumes	• All primary volumes in a consistency group must reside on the same primary peer. You cannot preserve write ordering when you have primary volumes originating on different systems. • All secondary volumes in a consistency group must reside on the same secondary peer. You cannot place secondary volumes on different systems and expect to preserve write ordering among them.
Replication sets	• Replication sets in a consistency group must be part of the same storage domain. • You must define identical consistency groups for both the primary and secondary peers in order to preserve write ordering on data arriving at the secondary peer. • Replication sets in a consistency group must have the same replication mode. • Replication sets in a consistency group must have the same role. If you attempt to add a replication set to a consistency group with a different role, the operation will fail. • If you add a replication set that is configured for asynchronous replication to a consistency group, the local asynchronous queue is deleted. The consistency group has its own asynchronous queue. • Each consistency group supports up to 128 replication sets.

Replication Set Properties

When you define a replication set or a consistency group, you set replication properties. To set these properties, do the following:

- Set the role of the replication set on one system to primary, and set the role of the complementary replication set on the other system to secondary.
- Set the replication mode to synchronous or asynchronous.
- Set the priority of the synchronization rate relative to I/O activity to high, medium, or low.
- If appropriate for your environment, enable the autosynchronization option.

  Autosynchronization is an alternative to manual synchronization. The autosynchronize option supports both replication sets and consistency groups. If you enable the autosynchronize option on the primary peer, the software synchronizes the volumes on both peers and resumes replication as soon as possible.

  If autosynchronization is enabled, the software attempts to synchronize the secondary volume with the primary volumes if the replication set or consistency group is placed in suspended mode because of a link failure or system shutdown. The software does not perform an autosynchronization operation if the replication set or consistency group was placed in suspended mode through either of the following methods:

  - You manually set the replication set or consistency group to suspended mode.
  - The asynchronous queue exceeded its limits while the replication link was active.

Note – When a replication set or consistency group resumes data replication, the autosynchronize option is not disabled.

- If you choose the Fast Start option, disable the autosynchronize option so that you can manually start the synchronization operation when you have restored the secondary volume and it is ready for replication.
- If appropriate for your environment, enable the option to create snapshots automatically.

Combining Replication Sets and Consistency Groups

There are several possible methods of putting replication sets in consistency groups. This section describes these methods.

Note – It is a best practice to plan for consistency groups ahead of time and create them at the same time that you create replication sets.

Creating a New Replication Set With a New Consistency Group

To create a new replication set and group it with a new consistency group, use the Create Replication Set wizard to create the new replication set and a new consistency group.

Creating a New Replication Set With an Existing Consistency Group

To create a new replication set with an existing consistency group:

1. Place the existing consistency group into suspended mode.
2. Use the Create Replication Set wizard to create the new replication set and to select the existing consistency group.
3. From the Consistency Group page, resume the consistency group.

Moving a Replication Set Into a Consistency Group

To move a replication set into a consistency group:

1. Place the existing consistency group into suspended mode.
2. From the existing replication set's Details page, move the replication set into the newly created consistency group.

Using Data Replication

This section describes replication modes and operations, as well as how to use snapshots with data replication.

For best practice examples, see "Best Practices: Using Data Replication" on page 123.

Replication Modes

The replication mode is a user-selectable property that defines the communication mode for a replication set. The software supports two modes of data replication: synchronous and asynchronous.

Synchronous Mode

In synchronous mode replication, a write operation to the primary volume is not acknowledged until the write has been completed on both the local and remote sites. The data written to the primary storage is committed to back-end storage on both the primary and secondary systems, before the write is acknowledged to the user, so that both copies are synchronized.

Asynchronous Mode

In asynchronous mode replication, data is written to the primary volume and to a local asynchronous queue. A write operation is confirmed as complete before the remote volume has been updated. Later, write operations that have accumulated in the asynchronous queue are forwarded in sequence to the remote peer.

The asynchronous queue exists to absorb bursts of application writes. You can select how the asynchronous queue operates when it becomes full and causes application writes to wait for room in the queue:

- Blocking mode – If the asynchronous queue fills, all writes to the primary volume and replication writes to the secondary volume are delayed until the queue drains enough to enable a write to occur. Blocking mode, which is the default option, ensures write ordering of the data to the secondary peer. If the asynchronous queue fills with the blocking option set, response time to the application might be affected. Write operations to the secondary volume must be acknowledged before being removed from the queue on the primary peer, so they can block further write operations to the queue until space is available.

- Suspended mode (also called scoreboarding) – The system keeps a change bitmap of data that has been updated but not replicated, perhaps because the link between the primary and secondary peers is temporarily unavailable, or because the data has not yet been committed to the secondary peer. Each bit in the change bitmap represents 64 kilobytes of storage. Writes issued at the primary peer are noted in the primary bitmap, and writes issued at the secondary peer are noted in the secondary bitmap. Merging the primary and secondary bitmaps results in a bitmap that defines the data that differs between the primary and secondary peers. The software uses this replication bitmap to reestablish data replication through optimized update synchronization rather than through a complete volume-to-volume copy.

  Before replication can resume after being in suspended mode, the secondary volume must be synchronized with the primary volume.

Consider the following issues when you choose asynchronous mode replication:

- All replication sets in a consistency group share a single asynchronous queue.
- If you choose the Suspend option, the software switches from replicating to suspended mode when the queue is full. Write ordering is preserved and write operations to the primary data volume are completed and tracked by the primary bitmap. To continue replication mode once the queue has drained, perform a synchronization operation.
- The minimum size of an asynchronous queue for a consistency group is 16 megabytes.
- You must choose the proper queue size for your environment. If the asynchronous queue fills, subsequent write operations must wait to be placed in the queue. As a result, the application response time increases. To improve the response time for the application, increase the asynchronous queue size based on its usage.

  If you need to extend the size of the asynchronous queue, do the following:

  a. Place the replication set or consistency group into suspended mode.

  b. On the replication set or consistency group Details page, change the asynchronous queue size.

  c. Initiate a synchronize operation for the replication set or consistency group to synchronize both peers and resume replication.
- You can specify the threshold above which the queue is considered full according to the number of queued disk blocks or the length of time an entry is in the queue. To set asynchronous queue parameters, use the Create Replication Set wizard or make changes on the replication set or consistency group Details page.
- Asynchronous mode accommodates bursts of write activity in which the write rate exceeds the bandwidth of the replication link. The asynchronous queue must be large enough to handle bursts of write traffic associated with the application's peak write periods. A large queue can handle prolonged bursts of write activity, but this activity causes the secondary peer to become further out of synchronization with the primary peer. The disk storage area previously allocated for queue usage is returned to the associated storage pool.

- If you add a replication set that is configured for asynchronous replication to a consistency group, that replication set's own queue is deleted and the group queue is used instead.
- With an asynchronous queue, the secondary volume is always somewhat out of date with the primary volume. The extent to which it is out of date depends on how much data there is in the asynchronous queue, as well as on the latency of the link.
- If you choose blocking mode and the queue becomes full, writes are blocked until the queue drains. The software maintains write ordering; however, application write operations are impacted.

You can change the replication mode at any time during the life of a replication set. However, you must first place the replication set in suspended mode. If the replication set is a member of a consistency group, you must place the consistency group in suspended mode.

Synchronous Compared With Asynchronous Replication Modes

When you choose a replication mode, consider the differences between synchronous and asynchronous replication described in TABLE 6-4.

TABLE 6-4 Synchronous Compared With Asynchronous Replication Modes

Characteristic	Synchronous Replication	Asynchronous Replication
Write operations	A write operation to the primary volume is not complete until the user data is written to the secondary peer.	The write operation is complete as soon as the user data is written to both the primary and secondary data volume.
Network response time	The response time depends on the network. The network latency must be low enough that the application response time is not affected dramatically by the time of the network round trip of each write operation. Also, the bandwidth of the network must be sufficient to handle the amount of write traffic generated during the application's peak write period. If the network cannot handle the write traffic at any time, the application response time is affected.	The response time does not depend on the network, as long as the queue is not filled. The bandwidth of the network link must be able to handle the write traffic generated during the application's average write period. During the application's peak write phase, the excess write operations are written to the local asynchronous queue. They are then written to the secondary peer at a later time when the network traffic allows. The application response time can be minimized during bursts of write traffic above the network limit as long as the asynchronous queue is properly sized. **Note:** Response time to the application might be affected if the asynchronous queue fills with the blocking option set.

TABLE 6-4 Synchronous Compared With Asynchronous Replication Modes *(Continued)*

Characteristic	Synchronous Replication	Asynchronous Replication
Availability	Synchronous replication provides high data availability at the expense of longer latency across the WAN.	Asynchronous replication offers higher performance and availability in exchange for delayed synchronization of all copies of data. The secondary peer lags behind the primary peer, depending on the latency, queue size, and number of write operations.
Write response time	Synchronous replication might exhibit an increased 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 usually has a quicker response time than synchronous replication.
Write order	Because every data change is replicated to the secondary peer in real time, the secondary peer is write order consistent with the primary peer. Write ordering across volumes is preserved at the secondary peer.	The secondary peer is write order consistent with the primary peer, except when it is receiving a synchronization operation from the primary peer.
Data loss during a disaster	Synchronous replication is best for data volumes that have zero tolerance of data loss at the secondary peer.	Asynchronous replication is best for data volumes that have some level of tolerance for a delay at the secondary peer.

Replication Operations

You can choose one of the following data replication operations from either the Replication Set Details page or the Consistency Group Details page:

- **Resume**, which provides the following two options:
  - Normal synchronization – A copy operation of differences between the primary and secondary volumes. Normal synchronization 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. After normal synchronization is completed, the volumes are write order consistent with each other.
  - Full synchronization – 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. A full synchronization is typically performed only when the remote volume is damaged due to a disaster or has questionable data integrity.
- **Suspend**, which provides the following two options:

- Normal – An operation to stop replication or synchronization. While the system is in suspended mode, data is written only to the local volume and the bitmap keeps track of changes. No replication occurs. At a later time, such as when an operation is completed or a link is reestablished, the volumes are synchronized, either through the autosynchronize option or manually. The replication bitmaps at the primary and secondary peers are compared, and the changed blocks for the primary peer volume are copied to the secondary volume. After the volumes are synchronized, replication starts again.
- Fast Start – An operation that signifies that you are performing the initial synchronization manually by means of a backup tape. You initiate the Fast Start option at the primary peer before you create the backup tape.

Use the Suspend operation as follows:

- Use the Fast Start option of the Suspend operation to minimize data replication I/O traffic when you set up a copy of the data at a new site. You can initialize the remote peer by shipping a backup tape that contains a copy of the user data and loading the backup tape directly onto the remote system. This option is useful if you have an extensive amount of user data, are concerned about the cost of the peer link, or are concerned with the speed of the peer link.
- Use a Suspend operation to save on telecommunications or connection costs. However, you risk the cost of data loss. If you lose the primary peer, you will have an RPO of the time the suspension took effect, thus losing any updates that occurred on the primary.

The replication sets can go into suspended mode for either planned or unplanned reasons, such as when you make a change to a consistency group or a network link failure occurs. If a network link fails and causes the software to cease replication, synchronization occurs when the link is reestablished.

While the software is in suspended mode, there might be a physical connection. However, there is no communication and no data replication between the two peers. When the software resumes data replication from suspended mode, the following occurs:

1. When synchronization is initiated, there is communication between the two peers, but write order consistency is not restored.
2. When the peers are synchronized, there is communication between the two peers, and write order consistency is restored. If the replication mode is asynchronous, there might be a time delay, but the data is write order consistent.

Note – If a replication set is a member of a consistency group, you cannot change the replication mode of the replication set. The replication set’s attributes must match those of the consistency group.

Rolling Upgrade

When replicating data between two DSP nodes, it is a best practice to upgrade one DSP node at a time. You can suspend the replication set and keep the host I/O running, using the following procedure:

1. Set up a replication set with autosynchronization on for a full data sync between DSP nodes A and B.

2. Suspend replication on the primary peer.

3. Upgrade the secondary DSP to new code, while monitoring the primary node in suspended mode as it awaits the corresponding node to resume the data replication.

4. Monitor the secondary DSP node as it comes online.

   The primary node will resume its data replication to node B, populated with new DSP firmware and Storage Service Processor patches.

5. If autosync resumes data replication, run data verification on node B.

6. Complete a role reversal test, in which you turn node B into a data replication master and run a full data replication synchronization.

Using Snapshots With Data Replication

If your organization has licensed snapshot as well as data replication features for your Sun StorEdge 6920 system, you can use snapshot functionality to take point-in-time copies of your data, as a safeguard in case issues occur during synchronization. For example, snapshots guarantee a known, write-ordered copy of your user data in the case of a network link failure during a synchronization operation.

Although data replication does not affect the integrity of the primary data, the secondary data is vulnerable during synchronization because write order is not preserved. Although you cannot access the secondary volume for read and write operations unless it is in suspended mode, you can create a snapshot of the secondary volume to access your data while the volume is being replicated.

To ensure a high level of data integrity on both peers during normal operations or data recovery, it is a best practice to create snapshots before synchronization so that a consistent copy of data is available. The snapshot provides a point-in-time data copy from which the data can be restored or rolled back. For more information, see "Example 5-5. Rolling Back a Snapshot" on page 92.

The system provides two options for creating snapshots automatically:

- Snap Before Synchronization – A snapshot of the secondary peer is automatically created before synchronization. When the synchronization option is completed successfully, the snapshot is deleted. If the synchronization operation is interrupted, you can use the snapshot to roll back the secondary volume to a consistent state.
- Snap with Primary – Two snapshots, one of the primary volume and one of the secondary volume, are created. The resulting snapshots at the primary and secondary peers represent the same point in time in your data.

Note – Currently, the web browser does not perform the snapshot operation automatically. You must manually create the snapshot copy on the primary peer before launching remote replication. When you take a snapshot of the primary volume, a snapshot is automatically taken of the secondary volume.

If you are using an asynchronous queue, a marker is inserted into the queue, and when the marker reaches the other side, the snapshot is initiated. Thus, the snapshot on the secondary volume is initiated after all the data that was in the queue when the snapshot was taken has been written to the secondary volume.

When using snapshots with data replication, follow these guidelines:

- Ensure you have allocated sufficient snapshot reserve space for snapshot operations to succeed.
- If you plan to create several snapshots at the primary peer and do not want to consume resources at the secondary peer, disable the Snap With Primary option.
- In order to ensure that the snapshot includes consistent data, quiesce the application, take a snapshot of the primary volume, and then back up the snapshot on the secondary volume.

  For more information, see Chapter 5.

Best Practices: Using Data Replication

The following examples describe best practices for using the Sun StorEdge Data Replicator software:

- "Example 6-1. Using Data Replication for Disaster Recovery" on page 124
- "Example 6-2. Using Data Replication for Remote Data Availability" on page 125
- "Example 6-3. Using Data Replication for Synchronous Operation" on page 127
- "Example 6-4. Using Data Replication for Asynchronous Operation" on page 129
- "Example 6-5. Sizing Network and Queues for Asynchronous Replication" on page 134

Example 6-1. Using Data Replication for Disaster Recovery

Remote replication can help your business recover from some disasters. You can use either synchronous or asynchronous replication for disaster recovery.

Prerequisites

To accurately prepare for disaster recovery, perform the following tasks:

- Undertake a risk assessment (RA) and business impact analysis (BIA) for your organization to ensure that all dependencies are fully explored.
- Understand all aspects of the business process to be protected.

Note

- An effective replication strategy is usually part of a larger business continuance plan (BCP) and requires management support.

Procedure

To use data replication for disaster recovery:

1. Engage a business continuance professional to carry out an RA and BIA.
2. Get support from stakeholders and upper management to ensure success.
3. Evaluate process dependencies to ensure that the proposed replication strategy satisfies requirements.
4. Ensure that processes are documented and tested regularly.
5. Follow processes and procedures for the replication methodology (synchronous or asynchronous) outlined in this guide and the online help.

Result

When you have completed this process, your organization will have a plan to handle all contingencies. A well-documented, tested plan establishes viable business practices to continue operations following a disastrous situation.

Example 6-2. Using Data Replication for Remote Data Availability

Data replication not only copies a database but also synchronizes a set of replicas so that changes made to one replica are reflected in the other. This functionality enables many users to work with a local copy while the database is updated, just as if all users were working on a single, centralized database.

Note – For database applications whose users are widely distributed geographically, replication is often the most efficient method of database access.

The data replication process of the Sun StorEdge 6920 system can run in synchronous or asynchronous mode and can be used in conjunction with snapshot features. You can establish a process in which the replicated data is used for reporting or backup at the primary or secondary site. This best practice example focuses on the use of snapshot functionality at the secondary location, but snapshot functions can be used at both ends of a replicated process.

Prerequisites

Before you can use data replication at a secondary site, the following prerequisites must be met:

- Replication of a process must be established between primary and secondary sites.
- Snapshots must be established on the secondary volumes.
- You must be able to quiesce the process (application) for snapshots to be taken.
- You must test all scripts before their introduction into production.

Notes

- Take snapshots during nonpeak operations. Occasionally shut down the system temporarily so it can quiesce.
- The length of time for which an application needs to be quiesced depends on how far out of sync the replication is. Replication must be synchronized up to the point at which the snapshot is taken. Synchronous replication is almost instantaneous.

Procedure

To use replication for remote data availability:

1. When creating a secondary volume, use the wizard to enable snapshots. The secondary volume should be the same size as the primary volume, with the snapshot sized to handle the changes during its utilization.

2. Establish synchronous or asynchronous replication between the sites.

3. Run a test to verify that you can quiesce the application under replication.

4. Create a script to enable the application to quiesce under replication.

5. Restart the application and create a snapshot using the command line (as a part of the script) or through the wizard.

Result

Execution of a snapshot on the secondary location creates a mountable image of the replicated data. This point-in-time copy can be used for reporting or backup purposes. It also provides a rollback path in the event of later corruption in an application.

For more information about snapshots, see Chapter 5.

Example 6-3. Using Data Replication for Synchronous Operation

Data replication can run in synchronous mode, during which each I/O is written in parallel to both the local and remote sites before the write is acknowledged to the user. Data is committed to back-end storage on both local and remote sites, so that the data is fully synchronized. Synchronous replication performance is affected by network latency, because write operations on both peers must finish before control is returned to the application.

Prerequisites

Before replicating a data volume in synchronous mode, perform the following tasks:

- Make sure that peer ports (Fibre Channel and Gigabit Ethernet) and peer links (Gigabit Ethernet only) are configured on the local and remote sites.
- Verify with your network provider that the Fibre Channel or Gigabit Ethernet replication link characteristics (throughput and latency) meet the needs of your application.
- Verify that there is enough disk space at the secondary site to accommodate the replicated volume and any potential snapshots that might be generated.

Notes

- There is a limit of two peer ports per system.
- If possible, create replication sets while host I/O is quiesced or at nonpeak loads.
- The initial synchronization of the data volume might take a long time over relatively slow replication links. Consider using the tape Fast Start option to reduce the initial synchronization time.
- You can disable replication of data from the primary to the secondary site by issuing the Suspend command from either site using the browser interface.
- Coordinate any snapshots of the replicated volume with application usage at the primary site. Restarting the application or flushing operations before taking snapshots provides a more consistent data set for potential future use.
- You can use the replication set's Suspend command at any time to disable replication between the two volumes. This is particularly useful for planned network outages or remote site maintenance.

Procedure

To use replication for synchronous operation:

1. If no replication ports are configured, use the Physical Storage / Ports / Enable Replication button to select at least one FC or Gigabit Ethernet port on the primary and secondary sites for replication. If you select a Gigabit Ethernet replication port, you must create a peer link as well.
2. Use the New volume wizard to create a volume on the secondary site. This volume should be the same size as the primary volume.
3. Create the replication set and assign the remote peer and WWNs for the primary and secondary volumes.
4. Designate the remaining replication set parameters on the primary and secondary sites. Both sites must be set for synchronous mode and have a replication bitmap allocated from a storage pool.
5. Select the appropriate snapshot option:
   - Snap Before Sync enables snapshots of the replicated volume at the secondary site to be taken before a sync is initiated from the primary site.
   - Snap With Primary coordinates consistent snapshots at both the primary and secondary sites.

Note – If autosynchronization was configured on the replication set, there is no need for a manual synchronization.

---

6. If autosynchronization is not enabled for this replication set, complete a manual synchronization at the primary site.

   A synchronization must also be run after the replication link state changes (in a "link down-link up" transition) if autosynchronization is not enabled.

Result

The host continues to access the newly replicated data volume. The data for the volume is continuously replicated between the primary and secondary sites. In the event of a disaster, the host's application data is available at the secondary site in the same state as at the failed primary site.

Example 6-4. Using Data Replication for Asynchronous Operation

This example provides recommendations for configuring storage resources for asynchronous replication.

To obtain a sense of how best to lay out storage for asynchronous replication, consider the data flow that results from a host write, as shown in FIGURE 6-6.

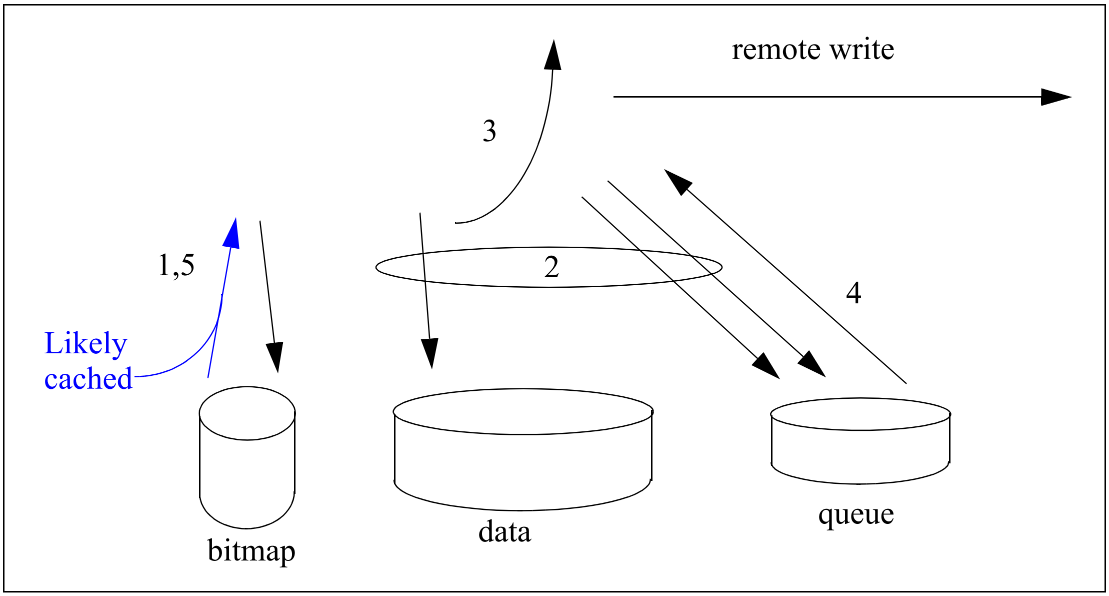

FIGURE 6-6 Data Flow During Asynchronous Replication

The data flow in FIGURE 6-6 is as follows:

1. The bit is set in the bitmap (read/modify/write).
2. Data is written to the data disk and queue disk (two writes).
3. The write is acknowledged.
4. The flusher reads from the queue and performs a remote write.
5. Upon acknowledgment of the remote write, the bit is cleared from the bitmap (read/modify/write).

Notes

- A host write will cause 5 to 7 I/Os on the Data Services Platform (DSP) for every host write. Assuming read cache hits for the bitmap reads, the number is 6.
- The read from the queue potentially contains several queued writes that are broken into the queued writes to increase the efficiency of the queue read and reduce contention for this volume. In this case, a more accurate approximation might be that data replication creates 5 I/Os to the Sun StorEdge 6920 subsystem for every 1 host I/O.
- Reads to a replicated volume involve no extra I/O and very little overhead.

Provisioning for Predictable Performance

Virtualization abstracts the low-level configuration from the administrator. Although virtualization can provide ease of use, there are situations in which specificity in the configuration is important. Since asynchronous replication turns a logical write from the host into several physical writes on the Sun StorEdge 6920 system, you must ensure that bottlenecks are not inadvertently created by contention for the same resources. Also, you must be sure that your configuration provides the expected performance level.

Avoid creating one storage pool for a given application, because it might create a configuration in which contention for disk resources produce unexpected performance regressions. Consider a case, for example, in which all data replication constituents–data, queue and bitmap volumes–are allocated from the same storage pool. In this configuration, the single queue reader per replication set or consistency group competes with potentially hundreds of writers to the same storage resources, in addition to data reads and writes, and bitmap reads (mostly cached) and writes. You can allocate the disk queue to disks that are independent of data and bitmap storage to remove a portion of this contention.

Consider the following example, in which all data replication constituents are allocated from the same storage pool, compared with a scenario in which separate pools are created with separate storage resources for data and queue volumes. This example shows the advantage of partitioning storage resources in order to prevent disk contention.

The following system setup was used to explore how different configurations affect local as well as network I/O:

- Two Sun StorEdge 6920 systems, each with two 2x2 arrays configured with two RAID-1 LUNs
- One Gigabit Ethernet link.

FIGURE 6-7 depicts the test configurations.

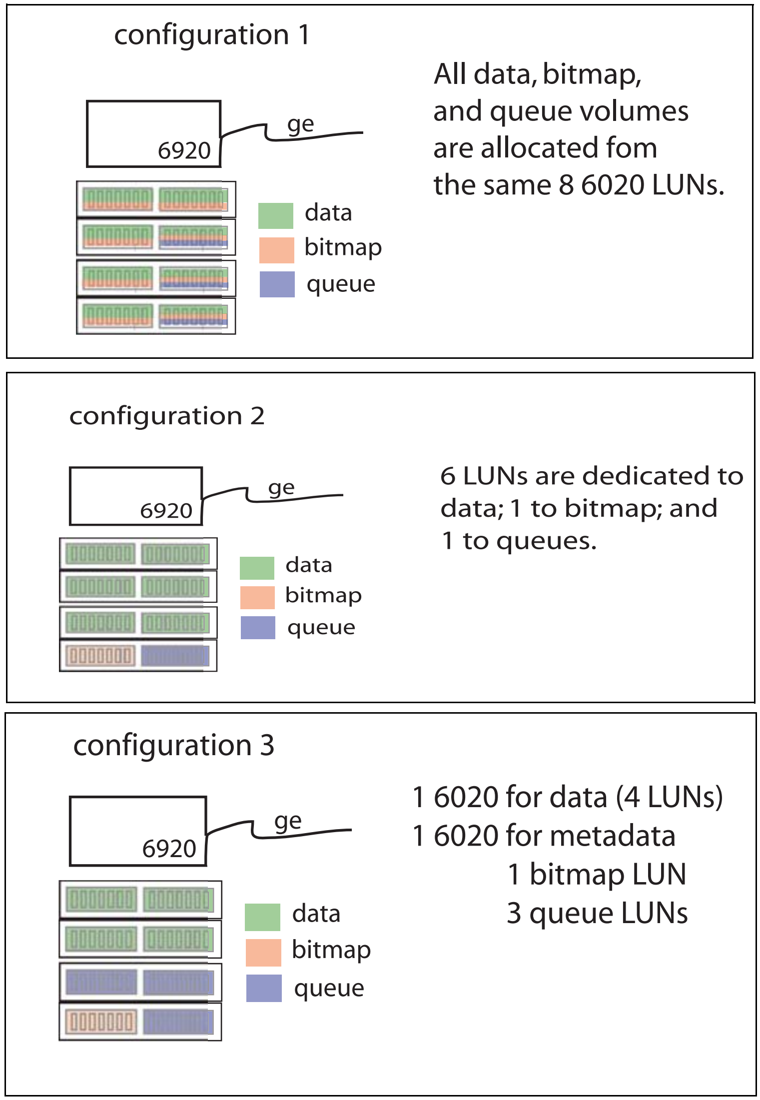

FIGURE 6-7 Provisioning for Predictable Performance – Configuration Examples

---

Note – Testing of these configurations consisted of the use of vdbench to test 8-kilobyte and 128-kilobyte random and sequential write workloads. Filebench was also used to simulate an OLTP workload.

Each of the configurations depicted in FIGURE 6-7 was tested with 16 replication sets in one consistency group and with 16 replication sets without a consistency group. TABLE 6-5 shows the test results.

TABLE 6-5 Configuration Test Results

	16 Replication Sets, 1 Consistency Group			**16 Replication Sets, No Consistency Group**		
	I/O per Sec	**ASVC Time**	**Net Writes/Sec**	**I/O per Sec**	**ASVC time**	**Net Writes/Sec**
Config 1	**Data volumes, queue and bitmap volumes distributed over all LUNs**					
8k sequential	170	730	30	640	200	85
8k random	160	810	23	630	200	350
128k random	86	1500	3	160	770	25
128k sequential	77	1700	2	140	920	40
filebench	230	2600	50	600	1000	240
Config 2	**Data, bitmap and queue volumes on separate LUNs**					
8k sequential	190	680	68	250	500	45
8k random	180	700	46	350	520	200
128k random	100	1200	8	48	2700	14
128k sequential	100	1200	4	45	2600	14
filebench	280	2100	75	400	1500	200
Config 3	**Data, bitmap and queue volumes on separate LUNs Additional LUNs dedicated to queue volumes**					
8k sequential	400	320	44	520	250	65
8k random	400	320	40	630	200	320
128k random	220	560	36	130	960	27
128k sequential	220	590	38	140	900	60
filebench	600	1020	33	800	765	300

To improve network throughput for replication sets in a consistency group, you can isolate queue and bitmap volumes from data volumes. Isolating these storage resources eliminates drive contention and subsequent starvation of the queue reader when there are many incoming write threads competing with a single queue reader thread.

You also can significantly improve local write throughput by doing both of the following:

- Isolating storage resources by function
- Adding queue devices to the consistency group's disk queue

There is some degree of parallelism allowed in writing to the disk queue, and striping across multiple LUNs provides greater concurrency.

Therefore, a best practice for configuring storage for data replication is to isolate data, bitmap and queue resources to minimize contention. This also makes capacity planning more straightforward, as illustrated in FIGURE 6-8.

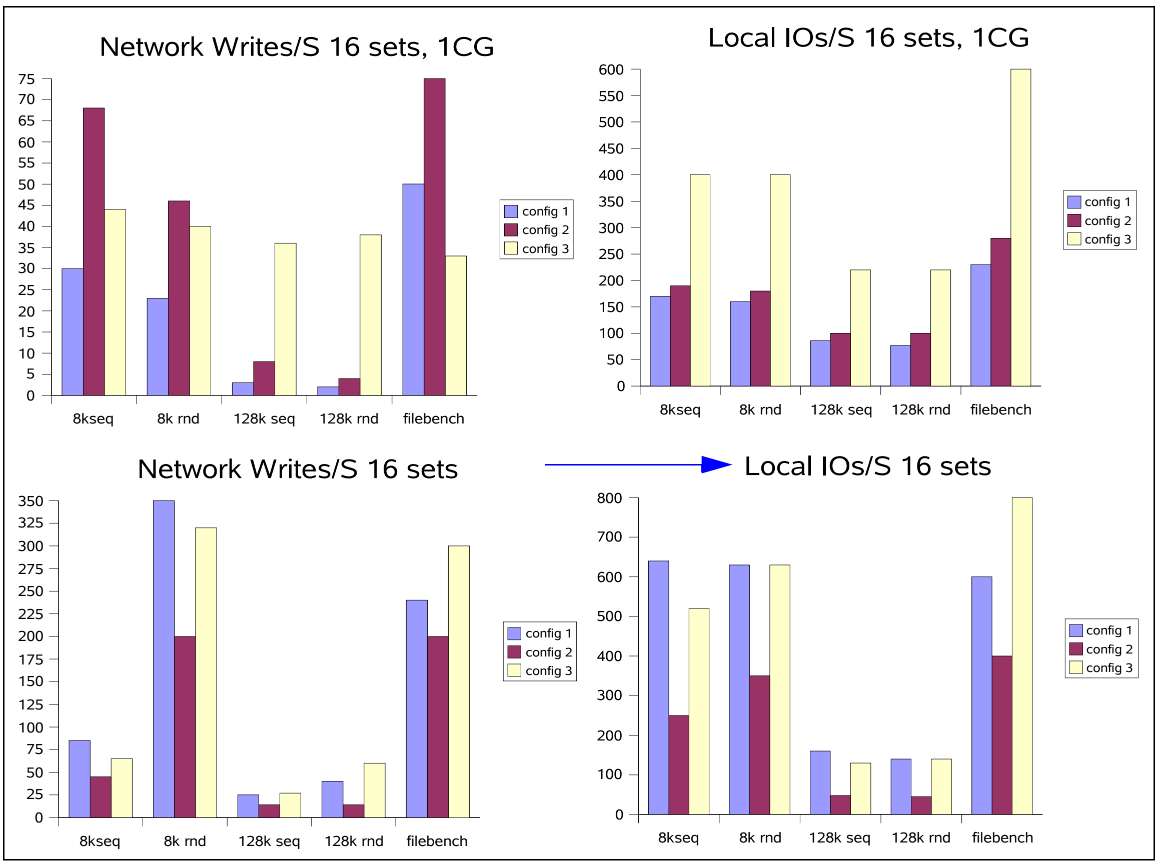

FIGURE 6-8 Configuring Storage for Data Replication

Example 6-5. Sizing Network and Queues for Asynchronous Replication

This example provides a method for sizing queues and provisioning network bandwidth for replication.

Prerequisites

Before you can size network and queues for asynchronous replication, you must do the following:

- In order to properly size asynchronous queues and provision the appropriate capacity for the WAN, it is essential that you characterize the expected workload. You can do this on a Solaris host with `iostat` and on a Windows system with `perfmon`.
- After determining the host workload for a typical day and choosing the network bandwidth (or knowing the capacity of a preexisting connection), size the asynchronous queues appropriately. The queues' purpose is to absorb bursts of write activity that exceed the replication rate, thus decoupling remote writes from the application write service times.
- Use the host workload data in sizing the network and determine whether data replication is a practical option for the application and network being considered.
- Choose the configuration that determines what data replication will do if the queue fills. There are two choices: Suspend or Block.
  - The Suspend option will suspend the replication set (or sets, if they are in a consistency group) when the queue is full. Network replication ceases, the queue is discarded, and local writes are recorded in the bitmap. At this point, restarting data replication through an update synchronization is a manual process (the autosync option does not apply here).
  - If the replication set is configured with the Blocking option and the queue becomes full, application writes to the set (or sets) will be blocked, waiting for room on the queue.

Notes

- Assuming proper queue sizing, a queue-full condition should be uncommon. When a queue full condition occurs, the unplanned write load that caused the condition may continue for an indefinite amount of time. The Suspend option is recommended over the Block option, since the Suspend option does not affect application writes.
- Consistency groups guarantee write ordering across all data volumes in the group. This is implemented through a single queue for all sets in the group and through the tagging of network writes with a sequence number to ensure write ordering at the secondary site. A consequence of this implementation is that the queue can become the bottleneck both for

writing and for network replication. It is a best practice to use consistency groups only when write-ordering across volumes is necessary, rather than merely as an administrative convenience.

Procedure

To size networks and queues for asynchronous replication:

1. Collect performance data for the volumes in question.

   For example, run `iostat` for 24 hours, generating statistics at 30-minute intervals (`iostat -Cnxz 30`).

2. Based on the average write bandwidth to the volumes to be replicated, select a network bandwidth.

3. Calculate the maximum queue size, based on the I/O load and replication bandwidth.

4. Select a queue size, based on the parameters you have chosen.

   The average bandwidth calculated from the example shown in TABLE 6-6 is 2220 kilobytes per second. Assuming a 45-megabit network link and conservatively choosing 3000 kilobytes per second as a replication rate, one can estimate that the queue will grow to a maximum of 23 gigabytes. Therefore, it makes sense to select a 30-gigabyte queue size to accommodate variances from the day's monitored workload.

   Further sizing can also be done with the information gathered. Note that the highest write rate (I/Os) during the 24-hour capture is 1000 writes per second, with the average write size being 8 kilobytes. When allocating storage for the queue, you must ensure that the storage is capable of sustaining 1200 I/Os per second (1000 8-kilobyte writes, 200 128-kilobyte queue reads) so that the queue does not become a bottleneck.

TABLE 6-6 24-Hour iostat Statistics

	Time	Kilobytes Per Second	I/Os	Net Kilobytes	Queue Growth (Megabytes)	Queue Size (Megabytes)
1	A	B	C	D	E	F
2						0
3					=((B3-D3)*1800)/1024	MAX(0;F2+E3)
4	0:600	900	113	3000	-3691	0
5	06:30	200	25	3000	-4922	0
6	07:00	1000	125	3000	-3516	0
7	07:30	1200	150	3000	1758	1758
8	08:00	4000	500	3000	-1582	176
9	08:30	2100	263	3000	-1406	0
10	09:00	2200	275	3000	0	0
11	09:30	3000	375	3000	0	0
12	10:00	8000	1000	3000	8789	8789
13	10:30	5000	625	3000	3516	12305
14	11:00	5300	663	3000	4043	16348
15	11:30	2000	250	3000	-1758	14590
16	12:00	500	63	3000	-4395	10195
17	12:30	200	25	3000	-4922	5273
18	13:00	1000	125	3000	-3516	1758
19	13:30	1200	150	3000	-3164	0
20	14:00	1600	200	3000	-2461	0
21	14:30	3000	375	3000	0	0
22	15:00	3300	413	3000	575	575
23	15:30	800	100	3000	-3867	0
24	16:00	1200	150	3000	-3164	0
25	16:30	2000	250	3000	-1758	0
26	17:00	950	119	3000	-3604	0
27	17:30	1000	125	3000	-3516	0
28	18:00	1200	150	3000	-3164	0
29	18:30	800	100	3000	-3867	0
30	19:00	1000	31	3000	-3516	0
31	19:30	1000	31	3000	-3516	0
32	20:00	200	6	3000	-4922	0

TABLE 6-6 24-Hour iostat Statistics *(Continued)*

	Time	Kilobytes Per Second	I/Os	Net Kilobytes	Queue Growth (Megabytes)	Queue Size (Megabytes)
33	20:30	300	9	3000	-4746	0
34	21:00	5000	39	3000	3516	3516
35	21:30	8000	63	3000	8789	12305
36	22:00	8000	63	3000	8789	21094
37	22:30	4000	31	3000	1758	22852
38	23:00	1000	8	3000	-3516	19336
39	23:30	1000	8	3000	-3516	15820
40	00:00	200	2	3000	-4922	10898
41	00:30	50	0	3000	-5186	5713
42	01:00	0	0	3000	-5273	439
43	01:30	3000	23	3000	0	439
44	02:00	3000	23	3000	0	439
45	02:30	4000	31	3000	1758	2197
46	03:00	8000	63	3000	8789	10986
47	03:30	2000	31	3000	-1758	9229
48	04:00	0	0	3000	-5273	3955
49	04:30	0	0	3000	-5273	0
50	05:00	100	1	3000	-5098	0
51	05:30	200	2	3000	-4922	0
AVG		**2220**	**8**			

Troubleshooting Data Replication

Following are issues that might occur with data replication and recommended actions toward resolution.

Before following instructions in this section, verify the following:

- Data paths to local data volume, bitmaps, and logs are functioning properly.
- DSP chassis hardware is functioning properly.
- There are no hardware failures or FRU problems present, and failover is not in progress.
- Configuration of the replication sets is correct, as described in the online help.

Replication Does Not Occur

When data replication does not occur, verify the following:

- The ports at both sites are enabled for data replication.
- There is an active replication link or connection.
- Only two replication links are configured, and the links are both on the same type of port. For example, both ports must be either Fibre Channel (FC) ports or Gigabit Ethernet ports, not a mix of the two.
- The correct remote volume and peer node World Wide Names (WWNs) were entered in the Create Replication wizard. If they were not, you must delete the replication set and create a new one with the correct WWNs.
- Both peers have complementary replication set and consistency group definitions.
- The secondary volume includes snapshot reserve space if the secondary replication set has the Snap Before Synchronization option enabled.

For information about FC and Gigabit Ethernet port failures, log in to the Java Web Console and click Sun Storage Automated Diagnostic Environment.

Port Integrity Issues Exist

If the PWR LED is on and green, try the following actions:

- Inspect the port LEDs to check the integrity of the physical ports.
- Reset the cable connector.
- Jiggle the cable.
- Inspect the equipment at the far end of cable connection.
- Ensure that the power cord is securely fastened.

Connected Ports Are Not Enabled

Ports with attached devices should have LEDs that are green and on or blinking. If they are not, check the device and the connector. Ensure that the expansion switch is depressed to enable ports 13 through 24.

If the ports are still disabled, use external test equipment to diagnose the problem.

Logical Network Connection Issues Exist

- If replication is not enabled on the port, enable it.
- If local IP, netmask, and gateway addresses are incorrect, configure them correctly.
- If remote host definitions are not correct on each SCSI link, configure the correct remote host definitions.
- If the default gateway is not in the Address Resolution Protocol (ARP) table, the default gateway is needed when the remote host is in a different IP subnet from the local system.

  Try the following:

  - Verify the netmask. An incorrect mask might cause system to seek the remote host on the local subnet.
  - Use an external sniffer to diagnose problem.

Default Gateway ARP Entry Is Incorrect

An incorrect ARP entry might mean the wrong system is responding to the configured IP address. In this case, you should purge the incorrect ARP entry and use an external sniffer to diagnose problem.

Remote Host Is Not in ARP Table

The default gateway is not needed if the remote host is in same IP subnet as the local system. Try the following:

- Verify the netmask. An incorrect mask might cause the system to seek the remote host through the default gateway.
- The remote host might not be responding to ARP. Use an external sniffer to diagnose the problem.

Remote Host ARP Entry Is Incorrect

An incorrect ARP entry might mean that the wrong system is responding to the configured IP address. In this case, purge the incorrect ARP entry. Use an external sniffer to diagnose the problem.

Default Gateway Does Not Respond

Ping the default gateway, and use an external sniffer to diagnose failures.

Remote Host Is Not Reachable or Does Not Responding

Ping the remote host, and use an external sniffer to diagnose failures.

Login Status of the Local System to Remote Peer Shows "Connection Failure"

The local system must complete the SCSI login to the remote system. The system automatically attempts login every 10 seconds until it succeeds. Try the following:

- Ensure that the remote system is configuring a peer port for replication.
- Open port 3260 on the remote site firewall.

Login Status of the Local System to Remote Peer Shows "Login Failure"

Ensure that the remote system is configuring a peer link to the local system.

Remote System Is Not Logged In to Local System

The local system must accept the SCSI login attempt of remote system to permit access to replicating volumes. The local site firewall might block passage of port 3260 traffic to the local system. In this case, open port 3260 at the local site firewall.

Local System Rejects Login Attempts of Remote Systems

Repeated failed attempts to log in to the local system might indicate a misconfigured remote system. The timestamp of login activity identifies recent attempts. A remote system might be configured to use the incorrect slot or port on the local system, or the wrong remote system might be trying to connect to the local system. Verify the remote system configuration.

There Is a Replication Set Error

The Sun StorEdge Data Replicator software replicates data from a primary volume to a secondary volume. The association between the primary and secondary volumes and a corresponding replication bitmap at each site make up a replication set. After the volumes in a replication set are initially synchronized, the software ensures that the primary and secondary volumes contain the same data on an ongoing basis.

If you enable the Snap Before Synchronization option for a secondary replication set, and the secondary volume does not include snapshot reserve space, the software returns the following error message to the primary peer when you try to resume data replication:

```
PIT failed for sync startup
```

Add snapshot reserve space to the secondary volume to resume data replication successfully. If you disable the Snap Before Synchronization option, the secondary volume is only partially replicated on the secondary replication set.

CHAPTER **7**

Working With External Storage

Located outside the base and expansion cabinets, an external storage device is a physical disk or storage array connected to the Sun StorEdge 6920 system. Also known as non-captive storage, external storage devices are not directly managed at the logical unit number (LUN) level. They are presented to the Sun StorEdge 6920 system as virtual disks assigned to storage pools for dissemination to hosts in the form of volumes.

The Sun StorEdge 6920 system allows a number of third-party storage arrays to be attached externally. This enables owners of such storage to retain the use of these legacy devices while taking advantage of the data services available in the Sun StorEdge 6920 system.

This chapter provides instructions on working with external storage. It contains the following sections:

Understanding External Storage

An external storage device is a storage array that is physically connected to the Sun StorEdge 6920 system service panel and located outside the system cabinet. An external storage device is located outside the management path and presented as a logical unit number (LUN). Because the storage is outside the management path, you must use the vendor's management software to configure the LUN.

Note – A LUN from an external storage array can be used in the Sun StorEdge 6920 system as a legacy volume or as a raw device in a storage pool.

You can use external storage in any of the following ways:

- Preserve the data on an external device by creating a legacy volume, and use snapshot or mirroring on the legacy volume, as if it were any other volume in the Sun StorEdge 6920 system.
- Increase storage capacity of the Sun StorEdge 6920 system by using external storage as raw storage. The external device is added to the system as a virtual disk.
- Migrate data from an external storage device to the Sun StorEdge 6920 system's internal storage by using data mirroring.

The External Storage Summary page in the browser interface lists an external storage device's LUN as an uninitialized disk. You can manage this storage by using the Manage External Storage wizard. When the wizard has finished running, the system displays the external storage as a virtual disk.

Note – The system management software does not configure external storage devices. You must use the native configuration tools for the external storage device to make changes to external LUNs and lower-level components. For more information, see the documentation that ships with the external storage device.

FIGURE 7-1 illustrates how a LUN on an external storage device can be used by the Sun StorEdge 6920 system, as a legacy volume or part of a storage pool.

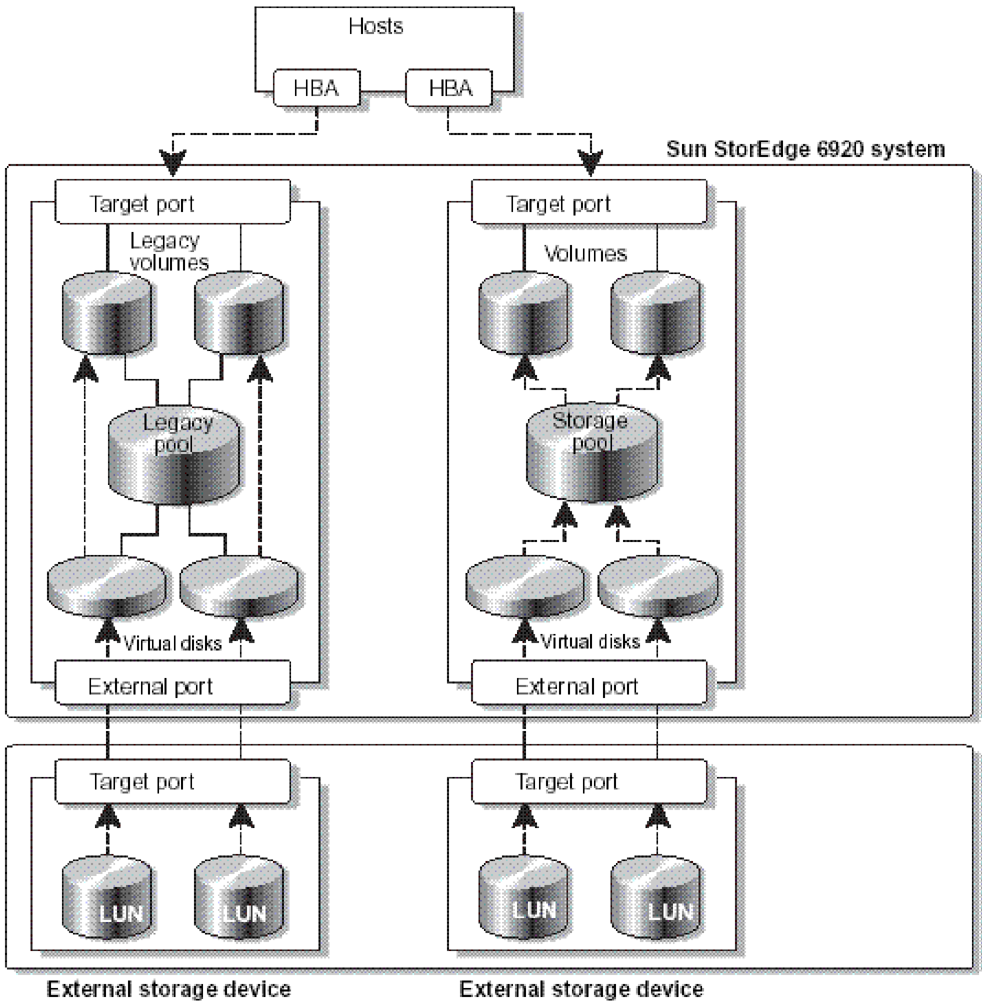

FIGURE 7-1 External Storage Utilization

Supported External Storage Devices

Currently supported external storage devices include the following:

- Sun StorEdge T3 array with 1-gigabyte cache
- Sun StorEdge 3510 array
- Sun StorEdge 3511 array
- Sun StorEdge 6120 array
- Sun StorEdge 6130 array
- Sun StorEdge 6320 system
- Sun StorEdge 9980 system
- Sun StorEdge 9970 system
- Sun StorEdge 9960 system
- Sun StorEdge 9910 system
- Sun StorEdge 3910/3960 systems (with Sun StorEdge T3 arrays with 1-gigabyte cache)
- Sun StorEdge 6910/6960 systems
- EMC CLARiiON CX 400 array
- EMC CLARiiON CX 700 array
- Hewlett-Packard EVA 3000 virtual array

For an up-to-date list of supported external storage devices, see the *Sun StorEdge 6120 Array Release Notes*.

For more information about how to cable an external storage device to the base cabinet, see the *Sun StorEdge 6920 System Getting Started Guide.*

Connecting External Storage Devices

You can connect external storage devices directly to the system's service panel or by means of a pair of redundant FC switches, as shown in the sample configuration depicted in FIGURE 7-2.

Note – You cannot connect more than one set of redundant FC switches between the Sun StorEdge 6920 system and an external storage array.

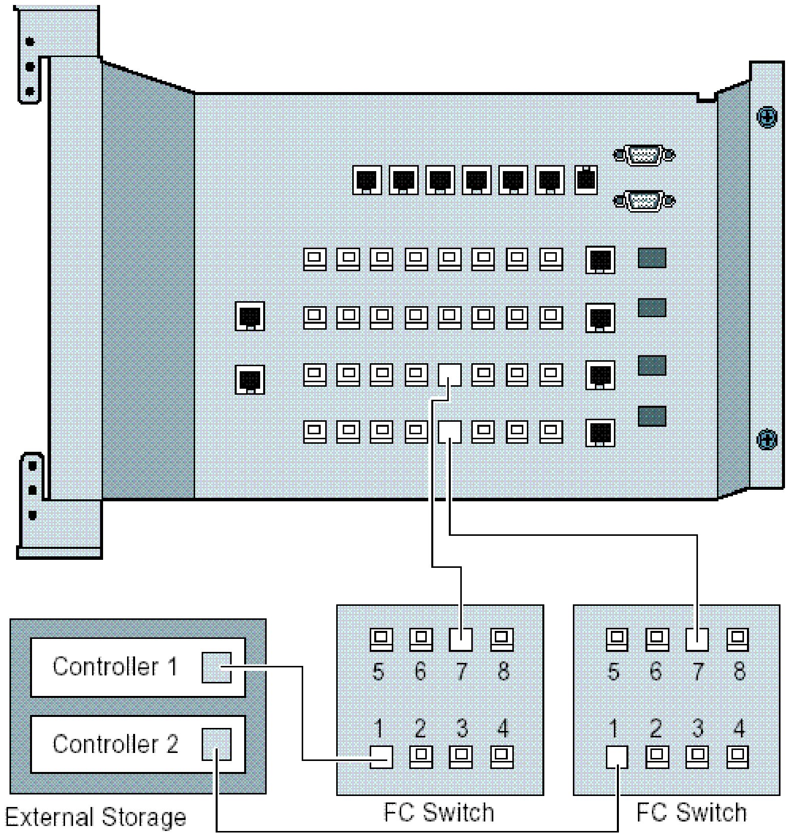

FIGURE 7-2 Connecting External Storage

Note – The primary and secondary SIO cards must be configured identically.

When you connect an external storage device to the system cabinet, the system automatically discovers the storage device. Using the host SSCS CLI software command `sscs list externalstorage`, list the summary of all external storage disks or list the details for each of the external storage disks.

About Legacy Volumes

A legacy volume (also called virtual legacy volume or VLV) is a Sun volume that supports virtualization features but whose user data resides on a legacy device and whose metadata resides elsewhere in the Sun StorEdge 6920 system. You use a legacy volume when you want to add virtualization features to a legacy device without compromising its user data.

A legacy volume is composed of an entire logical unit number (LUN) on an external storage array. You can apply the system's data services to a legacy volume; however, you cannot extend a legacy volume.

You can use a legacy volume for the following purposes:

- To preserve the user data and manage the legacy volume as if it were any other volume on the system
- To preserve and migrate the user data from the external storage device onto the internal storage in the system
- To replicate the legacy volume

After you create a legacy volume, you can perform the following volume operations:

- Map the legacy volume to an initiator
- Use the legacy volume to create a mirror component
- Add the legacy volume to an existing mirror for resilvering, which destroys any existing data that resides on the legacy volume
- Create a snapshot of a legacy volume and use the internal storage to hold the snapshot
- Roll back snapshot data to restore a legacy volume

You can delete a legacy volume from the Volume Summary or Volume Details page. When you delete a legacy volume, it reverts to being an uninitialized LUN on the external storage device. The external storage device retains any data written to the legacy volume; if no data was written to the legacy volume, the LUN retains its original data completely.

Keep in mind the following legacy volume information:

- Heterogeneous arrays are externally managed by way of native tools.
- The system supports 128 legacy volumes.
- Multipath failover support.
- Legacy volumes enable replication of volumes that are not native to the Sun StorEdge 6920 system.
- With legacy volumes, a mirrored volume enables data migration to Sun StorEdge 6920 system storage.
- Arrays can be used as raw storage for Sun StorEdge 6920 system volumes.

Planning to Use External Storage

Before you can manage external storage, you must do the following:

- Use the external storage device's native configuration tools to ensure that the external storage device includes at least one LUN:
  - A LUN from a single port on an external array cannot be mapped to more than one port on the Sun StorEdge 6920 system.
  - A LUN cannot appear on more than two DSP controller ports.
  - To support multipath failover, the LUN must appear on two of the external storage device's controller ports.

Note – The Sun StorEdge 6920 system designates the entire LUN on the external storage device as a legacy volume. You cannot specify more than one LUN per legacy volume.

- For information about how to connect an external storage device to the system, see the *Sun StorEdge 6920 System Getting Started Guide*.
- If the external storage device is connected directly to the system, make sure that any LUNs mapped to the system are not accessible by other hosts.
- Initialize at least two virtual disks from the system's internal storage. If you create legacy volumes, the system requires two initialized virtual disks for administrative purposes. When you create a legacy volume, you cannot delete or uninitialize the last two remaining virtual disks on the system unless you delete the legacy volume first.

FIGURE 7-3 shows a sample configuration in which the external storage device is connected to the system's service panel by means of a pair of FC switches. Each FC switch includes a zone with an exclusive path from a controller on the external storage device to a port on the system's service panel.

Note – Refer to the FC switch vendor's documentation for information about zoning.

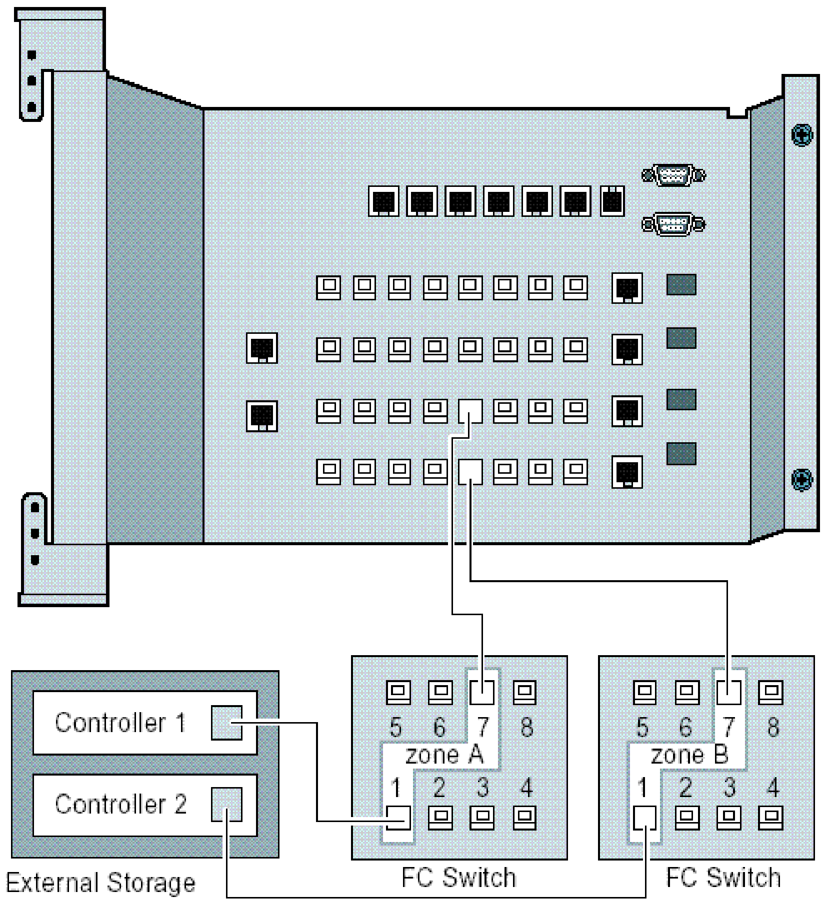

FIGURE 7-3 Zoning External Storage

There are a number of decisions you must make before you run the Import External Storage wizard:

- Determine your user data preservation strategy:
  - If you want to preserve existing user data on a LUN on the external storage device, create a legacy volume. This operation ensures that the system does not overwrite existing user data by using it as raw storage.
  - If you do not need to preserve existing user data on a LUN on the external storage device, designate the storage as raw storage for use by the system. This operation overwrites the existing user data.
- If you are creating a legacy volume, choose a legacy volume name with a meaningful description for your environment.
- Determine the storage pool to which you want to add the legacy volume or raw storage:
  - If you create a legacy volume, the system lists the available legacy pools to which you can add the legacy volume. A legacy pool is defined by the Legacy profile, which ensures that you can manage the legacy volume and not allocate it as raw storage for new volumes. The wizard does not allow you to place a legacy volume in a pool other than a legacy pool.

    The system provides a default legacy pool. Alternatively, you can create your own legacy pool (and place it in any storage domain), which must be defined by the preconfigured Legacy profile; you cannot create a custom Legacy profile.

    Snapshot reserve space for a legacy volume must reside in a storage pool other than a legacy pool. You cannot allocate snapshot reserve space for a legacy volume from the legacy pool in which it resides.
  - If you designate the external storage as raw storage, the system lists the available pools to which you can add the raw storage. The list includes all the storage pools on the system (other than legacy pools). Because the system cannot determine the attributes of a LUN on the external storage device, make sure that the LUN's attributes are consistent and compatible with the profile of the storage pool to which you plan to add the raw storage. For example, you do not want to add RAID-1 external storage to a pool defined by the RAID-5 profile.
- After you create a legacy volume, you have full access to the user data. You must map the legacy volume's WWN, which the system provides, to access the data.
- The system has the following limits on connections to an external storage device:
  - The system supports 1024 objects, including legacy volumes.
  - Each port can support 128 legacy volumes, as long as there are no initialized virtual disks active on the port.
- Unlike data paths to internal storage, the data paths to external storage devices might not be redundant. Therefore, if the data path between the system and an external storage device becomes unavailable, you risk losing access to your data. A best practice is to provide redundant paths between the system and the external storage device to minimize this risk.

FIGURE 7-4 shows the relationship between LUNs on an external storage device and the legacy volumes on the system.

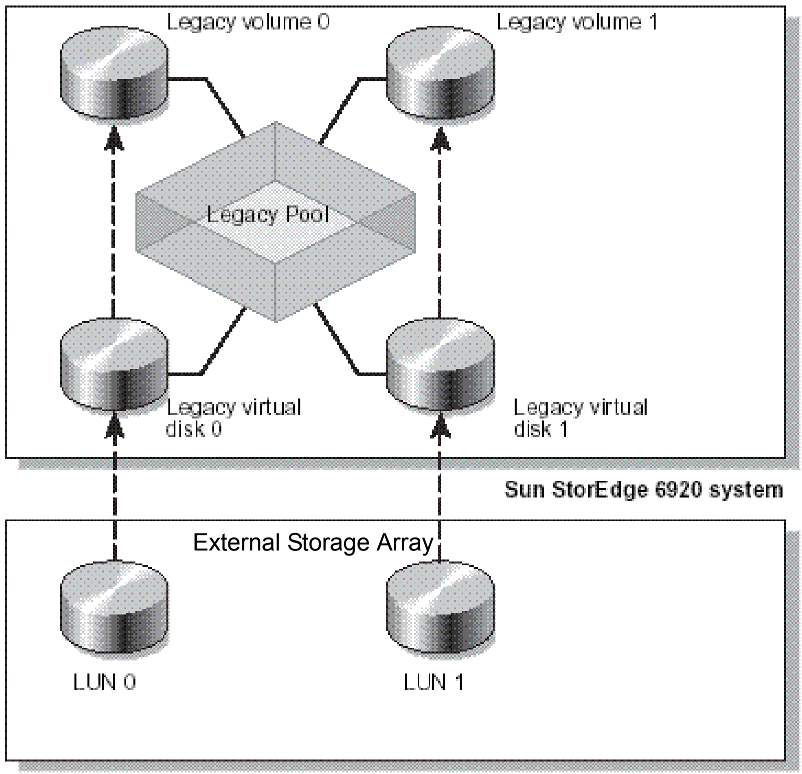

FIGURE 7-4 Mapping of External LUNs to Legacy Volumes

Migrating Data

The Sun StorEdge 6920 system can simplify the data migration process and reduce the risk of disruption to your usual business practices. After the initial system configuration, the data migration process can be performed without operational downtime, because the volume continues to be available to the host during the data transfer phase. After migration, the Sun StorEdge 6920 system enables you to continue using the legacy storage as part of a new data management architecture or to retire the legacy storage completely.

There are a number of reasons your organization might implement a data migration process, including the following:

- Consolidating islands of information present in your organization
- Consolidating business applications or processes
- Moving to a more cost-effective data management environment as a result of a total cost of ownership (TCO) study
- Moving data to a more cost-effective environment as part of a data lifecycle management (DLM) strategy

Data migration activities must be properly planned. Key elements of success are a well-defined project plan, assessment of your environment, and documented requirements translating this information into a risk mitigation strategy.

Assessing Requirements

Before defining a data migration plan, you must conduct a detailed assessment of your environment that includes both the IT environment and the business processes depending on the IT environment.

You can gather key information in a variety of ways, such as customer workshops, interviews with relevant personnel, and assessment of existing computer infrastructure.

It is a best practice to begin the assessment with a customer workshop. This helps set the scene and provide a forum for gathering initial information. It also helps to clarify any additional information you might still need and to identify relevant sources.

A range of factors influence the requirements of a data migration project, including:

- Criticality of availability of the data and related business processes
- Criticality of the data to the business
- Business processes and interdependencies
- Host (operating system) platforms involved
- Storage subsystems involved
- Size of data volumes to be migrated
- Growth expectations of volumes
- Time-critical factors such as downtime and nonavailability of maintenance windows
- Data security and availability requirements

It is important to identify which data is critical to the business and which data can be classified as legacy data that can be excluded from data migration activities. Consider business requirements compared with the cost of migrating the data when making decisions as to what data to archive and what data to migrate.

It is not enough to identify the location, size, and number of volumes to be migrated. For example, reviewing business processes can help identify interdependencies, up-time requirements, and performance requirements. This type of information might determine the sequence by which the volumes are migrated from the old to the new environment.

Ensure that the infrastructure to which the data will move meets the data availability and performance requirements. This removes the risk that, after the migration, application response times are longer.

Be sure to identify all the components in the IT environment that are affected by the data migration project. Version dependencies of components (hardware from different vendors, host operating system versions and patch levels, vendor, type and firmware levels of host bus adapters) must be verified against the current support matrix. You might decide to add elements in the plan to upgrade hosts, install new patches, or plan for firmware upgrades. In a production environment, strict change control procedures might be in place. This requires very careful planning.

You must also verify backup and restore procedures. If adequate processes are not in place, creating a backup and restore plan is time well spent. Also, you must define a strategy for how to handle the backup process during the data migration project.

Note – If you use replication on the legacy storage arrays, migrating to a new environment requires a separate project plan.

Using the Sun StorEdge 6920 System for Data Migration

Consider the following scenario: For several years, an application server has been storing application data on a number of volumes configured on a storage array. Technology has now improved, offering higher performing and higher capacity storage that is easier to manage. A study has concluded that investment in a new storage subsystem and its associated maintenance costs would be less of an expense than the operating costs of the current storage subsystem. The new platform of choice is the Sun StorEdge 6920 system, with its ability to easily migrate data from a legacy environment.

FIGURE 7-5 shows a volume of data ready to be migrated to the application server through the storage area network (SAN).

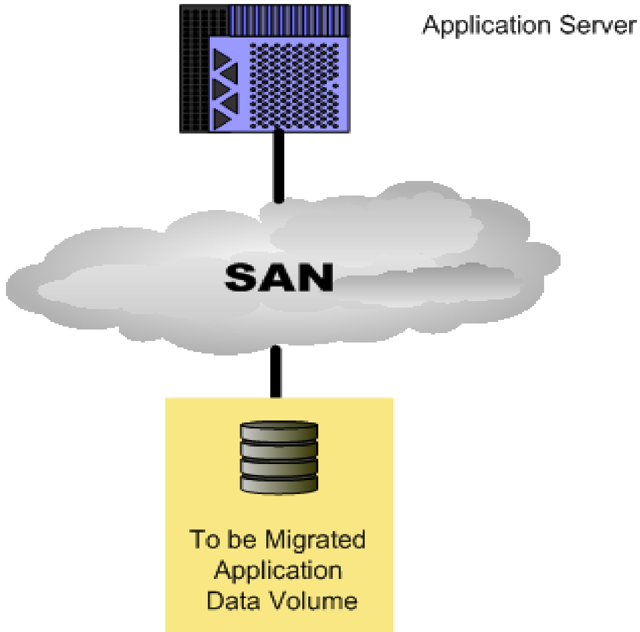

FIGURE 7-5 Legacy Storage Subsystem

Adding the Sun StorEdge 6920 System

The first step is to gather information about the components currently used in the I/O stack of the current infrastructure: hardware platform, operating system version, type of SAN switches used, type of host bus adapter (HBA), driver stack on hosts, and firmware revision levels installed on the various components.

FIGURE 7-6 shows examples of the types of information you can gather before migration.

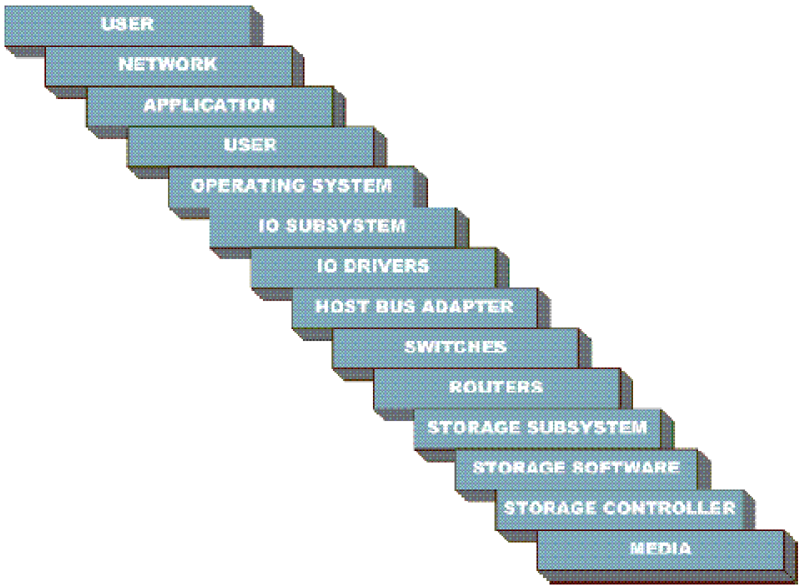

FIGURE 7-6 Information to Gather Before Migration

You might also need to plan for a firmware upgrade or change of components as part of the migration process. Document these activities carefully, estimating any scheduled downtime and impact on business processes, resources, and materials. In addition, you must properly scope the requirements for volume size, growth, and application use with the Sun StorEdge 6920 system.

When the requirements are determined, configure the Sun StorEdge 6920 system accordingly and connect to the SAN. Verify access from the hosts to the volumes. Finally, according to your plan for downtime on the hosts used by the legacy storage subsystems, disconnect the legacy storage array from the SAN and connect it to the Sun StorEdge 6920 system.

Importing Legacy Volumes

When the legacy storage subsystem is connected to the Sun StorEdge 6920 system, its data volumes are configured on the system to become legacy volumes. The Sun StorEdge 6920 system determines the World Wide Name (WWN) that is displayed for a legacy volume. The system hides the original WWN of the legacy storage array.

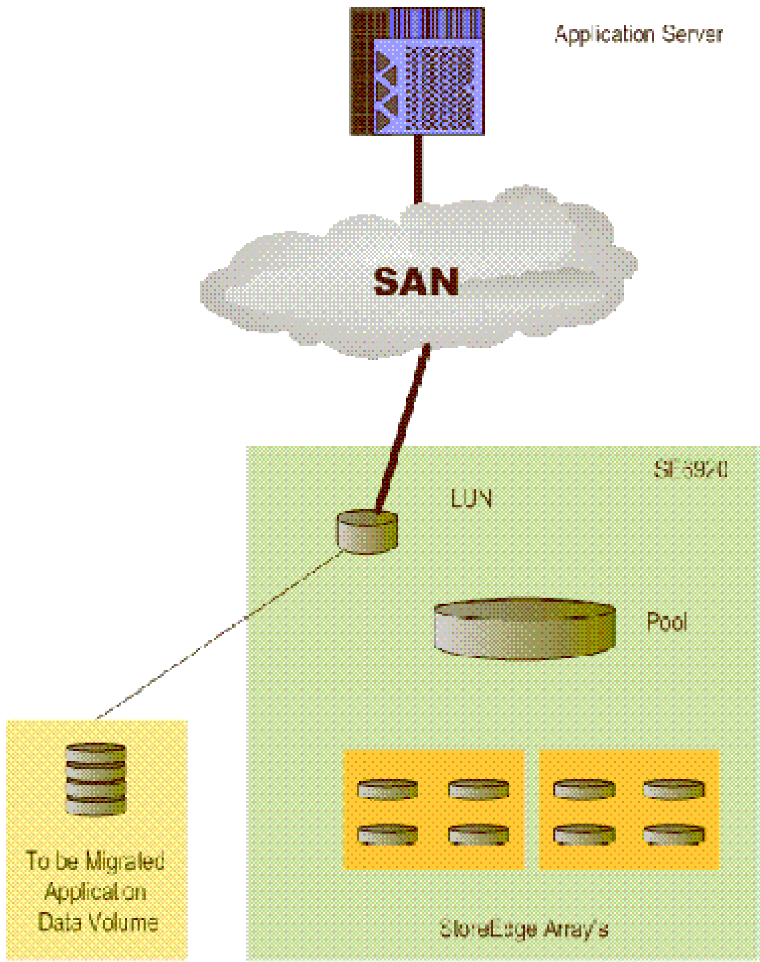

FIGURE 7-7 Legacy Volumes

When the legacy volumes are set up, the hosts have access to the data as if it were presented by the legacy storage array. You can now restart the application process.

Creating Mirrored Volumes

The next step in your data migration plan is to add mirrored volumes to the legacy volumes. A legacy volume acts as one surface of a mirror while the other side resides on a volume as part of a pool within the Sun StorEdge 6920 system, as shown in FIGURE 7-8.

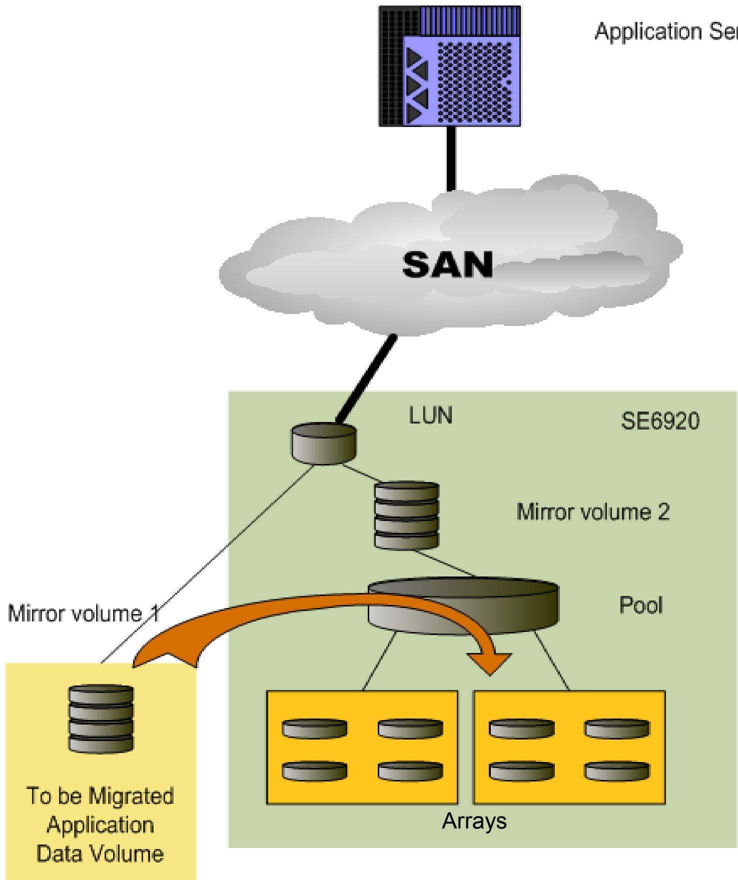

FIGURE 7-8 Mirrored Volumes as Part of Data Migration

You can mirror the data on a legacy volume to a normal volume, during which time the system can continue to access the legacy volume. When mirroring is complete, the mirror components are resilvered so that the legacy data also exists as an independent copy on the new mirrored volume.

Phasing Out Legacy Storage

When the mirrored volumes of the Sun StorEdge 6920 system are in sync with the data volumes on the legacy storage array, you might break the mirror component that includes legacy volume data from the mirror and delete it. Data is now migrated to Sun StorEdge 6920 system's volumes with performance I/O characteristics that meet the application requirements and access to all the data management functions of the Sun StorEdge 6920 system.

Note – When you attach a Sun StorEdge 6910 system to the same zone as a Sun StorEdge 6920 system, the 6920 system sees the 6910 system as an initiator instead of as a target. There is no workaround for this, because the configuration is not valid.

Best Practices: Using External Storage

The following examples describe best practices for using external storage:

- "Example 7-1. Migrating Externally Stored Data to the Sun StorEdge 6920 System" on page 159
- "Example 7-2. Using External Storage to Add to the System's Capacity" on page 162

Example 7-1. Migrating Externally Stored Data to the Sun StorEdge 6920 System

Migrating data that is stored on external devices to internal storage on the Sun StorEdge 6920 system is a benefit of the architecture. Through the use of mirroring features, the administrator can accomplish this migration with little interruption to the existing host. With the exception of the time taken to attach the legacy storage to the system as an external array, the migration process can occur online with the application running.

Prerequisites

- The Sun StorEdge 6920 system must be implemented and space must be available on the array to mirror the data from the external storage.
- The host that the data resides on must be at a level supported by the Sun StorEdge 6920 system. This includes OS and patches, HBA, and SAN connectivity.
- For each migration that you perform, complete a full assessment of the variables to ensure that all risks are fully documented and understood.
- As a risk mitigation strategy, document all existing configurations so that you can back out of the migration process should difficulties arise in the process.

- Ensure a fully documented project plan exists, as well as a strategy to ensure that the application is available at a given time.
- When planning a migration using the system's data mirroring features, ensure that the maximum number of objects within a given Sun StorEdge 6920 system does not exceed the 1024 limit.
- You must have at least two virtual disks in a storage pool to enable legacy mirroring.

Notes

- No more than 128 LUNs can be migrated at a time.
- You can have switches between external devices and the Sun StorEdge 6920 system.
- It is possible to migrate data from one external storage device to another under control of the Data Services Platform (DSP). Legacy volumes are used only when data is to be preserved on legacy storage systems.

Procedure

To migrate externally stored data to the Sun StorEdge 6920 system:

1. Establish the details of your documentation and migration strategies, as follows:

   a. Document the process by which you will migrate the data to the legacy system. This should be a well-planned project that takes into account all steps involved with the migration.

   b. Determine whether existing HBAs on the host can be connected to the Sun StorEdge 6920 system.

   c. If the HBAs and their firmware are not supported, determine whether the HBAs require a firmware upgrade or whether new HBAs are needed to accommodate the attachment of the Sun StorEdge 6920 system. This is usually done at the time the array is implemented.

   The best practice is to ensure that the legacy system is also supported on the firmware and HBA selected. If the two storage arrays need different levels, separate HBAs might be needed to handle the migration and support the two arrays. This is particularly important if the host must communicate with the legacy system during the migration.

   d. Ensure that any multipathing software that you are running is compatible with both the Sun StorEdge 6920 and the legacy system. If you need two different multipathing software solutions, ensure that they can coexist on the same host. If this is not possible, do not install multipathing software for the Sun StorEdge 6920 system until the migration is complete and the legacy system is no longer directly in the data path.

e. Determine a time period during which the system can have maintenance done. Allow sufficient time to move the legacy storage to the Sun StorEdge 6920 system, generate any WWN changes on the host, and remap the newly presented legacy volumes to the host. This might require a number of reboots to ensure that all volumes are presented and available to the application. Ensure that any queue depth issues are examined and documented for change (if needed) prior to any movement of the hardware.

f. Document all changes that the newly created legacy volumes will make within the application. Ensure that you know all affected mount points. Ensure that you document any hard-coded references to the legacy volumes.

2. Bring down the host that you are migrating.

3. Move all data paths from the legacy storage that will be migrated. Ensure that all data paths to the legacy storage are attached to the Sun StorEdge 6920 system and that their LUNs will be migrated. Any other access points that enable data to travel from the host to the legacy storage can result in data corruption during and following migration. Use LUN masking (where available on the legacy system) to ensure that only the LUNs being migrated are attached to the system. Use switching where needed to allow other LUNs associated with the migrated data path to operate as before.

4. When the legacy volumes are attached to the Sun StorEdge 6920 system, bring each legacy volume into the array to encapsulate these volumes as legacy volume devices. (See the online help for procedures.)

5. When the legacy volumes are created, map them to the initiators identified for the host. These could be new or old HBAs, depending on the requirements of the system. Use proper switching and zoning to ensure the LUNs are seen as specified. In some instances, an HBA might need to detect both the Sun StorEdge 6920 system and the legacy system, depending on the configuration. Document steps to ensure expected results.

6. When the legacy volumes are in place on the host and all mappings to them are correct, verify that the application starts. If not, verify that all volumes were imported as expected. Errors at this point must be carefully tracked. Follow your mitigation strategy to ensure that the maintenance window is not compromised. Follow your backout plan, documenting issues as you see them.

7. When the application is back up and running, start the mirroring processes for each of the legacy volumes. Depending on the number, size, and application requirements, mirroring processes for different mirrors can run simultaneously.

   For more information about mirroring, see Chapter 4.

   When the mirroring processes are complete, the legacy volume is like any other mirror component in the Sun StorEdge 6920 system. The legacy volume component can be removed, and the legacy system can be disconnected.

8. If you need to migrate more than 128 volumes from the legacy system, start again with Step 5, repeating as needed.

Example 7-2. Using External Storage to Add to the System's Capacity

The Sun StorEdge 6920 system enables a number of third-party storage arrays to be attached externally. Owners of such storage retain the use of these legacy devices while having access to the data services available in the system. External storage devices are not directly managed at the LUN level but are presented to the Sun StorEdge 6920 system as virtual disks assigned to storage pools for dissemination to hosts in the form of volumes. External storage devices can be carved up in the same fashion as internal arrays and their LUNs within the Storage Pool Manager.

Prerequisites

- Ensure proper licensing of the internal and external storage devices:
  - Sun StorEdge Storage Pool Manager for Sun storage
  - Sun StorEdge Storage Pool Manager for third-party storage
- Fully implement the Sun StorEdge 6920 system.
- Ensure that the limitation of 1024 objects within the system is not exceeded through the addition of the external storage.
- Follow all SAN rules in conjunction with the attachment of the external storage devices.
- Because the system cannot determine the attributes of a LUN on the external storage device, make sure that the LUN's attributes are consistent and compatible with the profile of the storage pool to which you plan to add the raw storage. For example, you do not want to add RAID-1 external storage to a pool defined by the RAID-5 profile.
- Determine the storage pool to which you want to add the legacy volume or raw storage. If you designate the external storage as raw storage, the system lists the available pools to which you can add it. The list includes all the storage pools on the system other than legacy pools.

Notes

- Implementation of a SAN between the external storage and the Sun StorEdge 6920 system is allowed. Hops are limited to one.
- All data service license requirements are in force, regardless of the internal or external status of the storage.
- All data on the external storage devices is lost as the devices are brought into storage pools within the Sun StorEdge 6920 system. Should data need to be protected, follow the best practices on migrating data from legacy systems described in "Example 7-1. Migrating Externally Stored Data to the Sun StorEdge 6920 System" on page 159.
- If the external storage device is SAN-attached to the system, make sure that any LUNs mapped to the system are not accessible by other hosts.

- The number of arrays that you can connect to the same DSP port through the zoned SAN is a function of how many virtual disks each array contains (such as 4 arrays, each with 4 virtual disks; or one array with 16 virtual disks).
- Many arrays have several configuration parameters (operating modes) to accommodate a wide range of hosts. You must set those parameters correctly for the DSP.

Procedure

To use external storage to add to the system's capacity:

1. Using native tools associated with the storage subsystem, provision all external storage.
2. Determine any LUN masking requirements and establish any necessary zoning requirements between the legacy, external storage and the Sun StorEdge 6920 system.
3. Connect the external storage to the ports of the Sun StorEdge 6920 system.

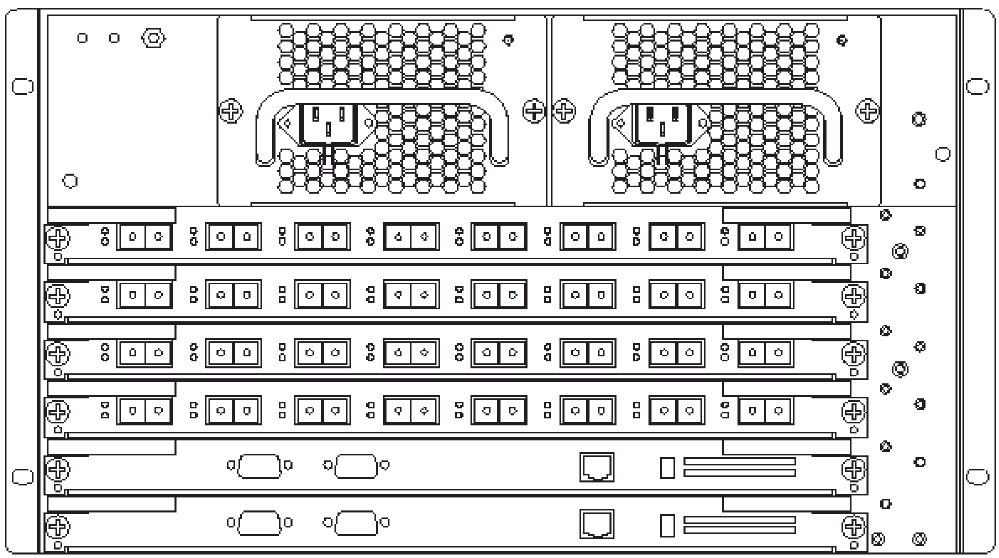

FIGURE 7-9 Sun StorEdge 6920 System Ports–Back View

Note – It is a best practice to ensure the host and disk are on separate port processors.

Troubleshooting External Storage

Following are issues that might occur with external storage and recommended actions toward resolution.

Legacy Volume or External Storage Is Not Available

If a legacy volume or external storage is unavailable, do the following:

- Verify that your SAN is configured properly.
- Verify that you have zoned your fabric appropriately.
- Verify that the external storage device is cabled to the system properly.
- From the External Storage Summary page, click the Rescan Devices button to force a rescan of all available storage in the system.
- Contact Sun Professional Services.

For more information about SANs, FC switches, and zoning, see "Related Documentation" on page xviii.

Diagnostic Support

The Sun StorEdge 6920 system does not provide diagnostic support for problems incurred on external storage devices. Use the diagnostic tools included with the devices, or contact the vendor for more information.

Even if the external storage device is also a Sun product, it is not monitored by the Sun StorEdge 6920 system. A licensed copy of the Storage Automated Diagnostic Environment should be installed on the host connected to the external storage.

CHAPTER 8

Working With the Sun Storage Automated Diagnostic Environment

The Sun StorEdge 6920 system has an integrated health monitoring and fault isolation function that includes the monitoring and diagnostic software engine called Sun Storage Automated Diagnostic Environment software. All subcomponents in the system are integrated with the fault analysis engine that the monitoring and diagnostic software provides. The software monitors functions, devices, and performance of the Sun StorEdge 6920 system and is accessible from the Java Web Console and the command-line interface (CLI).

This chapter provides instructions for the Sun Storage Automated Diagnostic Environment. It contains the following sections:

Understanding the Sun Storage Automated Diagnostic Environment

The Sun Storage Automated Diagnostic Environment software ensures that the responsible system administrators are notified when an event of note (alarm) has occurred. You can configure the notification procedures by using the browser interface.

Monitoring and diagnostic functions apply to the Storage Service Processor, arrays, Data Services Platform (DSP), power relay (when installed), Network Terminal Concentrator (using ping), USB flash disk (through monitoring `/var/adm/messages`), and modem (by monitoring of the UUCP transfer log).

How It Works

The Sun Storage Automated Diagnostic Environment software performs the following functions for the Sun StorEdge 6920 system:

- Collects health, configuration, and other non-customer-related data
- Checks the configuration of the entire system, as well as individual configuration of the Sun StorEdge 6020 arrays within the system
- Shows the user a detailed display of the Sun StorEdge 6920 system configuration
- Evaluates statistical error reports
- Reports statistical data to Sun Solution Centers (when Sun StorEdge Remote Response service is being used)
- Monitors message files for errors to obtain status information about the Sun StorEdge 6920 system
- Makes decisions on actionable service issues
- Notifies designated parties about events when action is required – messages indicate system location, component location, component identifiers, probable cause, and recommended action
- Provides step-by-step instructions to add Sun StorEdge 6020 arrays to an existing Sun StorEdge 6920 system
- Checks device revision of Sun StorEdge 6920 system hardware components
- Checks and installs (updates and patches) software revisions
- Enables manual updates and automatic checking for software revisions of components in the Sun StorEdge 6920 system
- Determines whether the power sequencer is on or off

Editions

There are two versions of the Sun Storage Automated Diagnostic Environment software:

- The System Edition (`SUNWstads`) is pre-installed on the Storage Service Processor to provide diagnostics for the Sun StorEdge 6920 system.
- The Enterprise Edition supports devices outside the Sun StorEdge 6920 system. You must install it separately. The Enterprise Edition includes `SUNWstade` (base package with CLI) and `SUNWstadm` (browser interface).

Security Features

The Sun Storage Automated Diagnostic Environment software tracks version levels of Sun StorEdge 6920 system components to ensure that appropriate firmware is installed and necessary configuration information is restored after a new Sun StorEdge 6920 system component or field-replaceable unit (FRU) is added. The monitoring and diagnostic software discovers the new hardware and invokes the software that tracks version levels of individual components within the Sun StorEdge 6920 system. It then upgrades or downgrades the component firmware as necessary (after confirmation from the user) and then runs an install script to restore the configuration information.

Other main security features of the monitoring and diagnostic software are:

- 128-bit Secure Socket Layer (SSL) encryption support for the user interface between the customer's management local area network (LAN) and the Storage Service Processor
- Password protection in the user interface
- Additional security features, when remote notification by Sun StorEdge Remote Response service is enabled

Browser Interface

The browser interface is designed to mask the complexity that is typically involved in configuring monitoring and diagnostics for a storage system. The Sun Storage Automated Diagnostic Environment software interface includes extensive event management capabilities and enables you to manage the individual system components.

Solution Extract

Solution Extract is a utility that is similar in functionality to the Explorer (`SUNWexplo`) or T3Extractor utility used for Sun StorEdge 6120 arrays. Its purpose is to gather the data used to verify configuration and error conditions.

For more information, see the online help.

Disk Scrubbing

Disk scrubbing is turned on in every array that ships with the Sun StorEdge 6920 system. An interface through the monitoring and diagnostic software is available to turn it off, if required. You turn off disk scrubbing in order to enable the Volume Verify feature.

Converting a 2-Up to a 4-Up DSP

You can convert a Sun StorEdge 6920 system with 16 Fibre Channel (FC) ports to a system with 32 FC ports. This is accomplished by the simple addition of two SRC/SIO cards and the necessary cabling and wiring. No other upgrades (hardware, software, or firmware) are required.

This procedure is documented in the Service Advisor, available from any Storage Automated Diagnostic Environment interface page.

Note – It is a best practice to perform revision checking whenever new hardware is installed, since Sun Storage Automated Diagnostic Environment software does not provide any hardware upgrade procedures.

Preparing for Diagnosis and Monitoring

The following are tasks and operations you must perform before using Sun Storage Automated Diagnostic Environment software to monitor your Sun StorEdge 6920 system:

- Verify site information.
- Ensure that company and contact information is correct.
- Configure remote service monitoring and observation information.
- Establish email notification:
  - Specify email addresses for local email and pager notification recipients.
  - Set up Simple Network Management Protocol (SNMP).
- Enable and configure remote Sun notification recipients, including Network Storage Command Center (NSCC) and Sun StorEdge Remote Response service.
- Manually initiate and run the agent run to verify system health.

For more information about Sun Storage Automated Diagnostic Environment software requirements, see the *Sun StorEdge 6920 System Getting Started Guide*.

Setting Site Information for Diagnostics and Monitoring

Use the Sun Storage Automated Diagnostic Environment application to set diagnostic and monitoring site information for your system. The monitoring and diagnostic software is preconfigured to recognize the hardware configuration, but it does not recognize specific information about your site. Before using the system, you must specify basic site and email notification information.

Note – You can be logged in as a storage or admin user to configure diagnostic and monitoring settings.

Procedure

To set site information for diagnostics and monitoring:

1. From the Java Web Console page, click Sun Storage Automated Diagnostic Environment.

   The Site Information page is displayed.

2. Fill in all mandatory Site Information entries as denoted by red asterisks, and then click Save.

   If you need a description of any of the fields, click the Help button on the Site Information page.

3. To set up email or pager notifications:

   a. Click Administration > Notification > Email.

      The Email Notification page is displayed.

   b. Click New to add a new email or pager notification address.

      The Add Email Notification Page is displayed.

   c. Enter each email or pager notification address you want to add, and click Save.

   d. Fill out the Add Email Notification Form, using the following guidelines:

      Type: Enter email or pager, depending on how you plan on receiving your alerts.

      Email Address: Fill in your email address.

      Categories: Select All Categories.

Priority: Select Critical and above. Any other priority saturates the email address with noncritical events.

Active: Select Yes.

Apply Email Filters: Select Yes.

Skip Components of Aggregated Events: Select No.

4. To test the notification email service:

   a. Click Administration > Notification > Setup.

      The Notification Setup page is displayed.

   b. Enter the name of the SMTP server in the * Use this SMTP server for Email field.

   c. Click Test Email.

      The Send Test Email page is displayed.

   d. Fill in the * To field and type a brief message in the Message field.

   e. Click Send.

      A confirmation message appears if the test email is sent successfully.

   f. Click Close to close the Send Test Email page.

   g. Verify that you received the test email message.

Note – Testing email notification is essential, because the SMTP server is required even if you are not testing email service, and the email path is required if the Sun StorEdge 6920 system is installed behind a firewall.

---

5. To set up SNMP traps, click Administration > Notification > SNMP, and do the following:

   a. Click New, and then enter the IP name or address of the new SNMP recipient. You can provide information for up to five IP addresses.

   b. Select the minimum alarm level for which SNMP notifications are to be sent to the SNMP recipient. The options are Down, Critical, Major, and Notice.

   c. Click OK.

6. To enable remote notification:

   a. Click Administration > Notification > Setup.

      The Notification Setup page is displayed.

b. Select the check box for each remote notification provider you want to enable:

   Network Storage Command Center (NSCC) – Sends notifications by email to the Sun NSCC, which enables Sun to continually improve the product and its support through analysis of this data.

   Sun StorEdge Remote Response service – Sends notifications to the Sun StorEdge Remote Response service.

c. To activate the NSCC provider, enable or disable email encryption by clicking Yes or No.

d. To activate the Sun StorEdge Remote Response provider, enter the frequency (in hours) with which you want to check the communication link to the Sun StorEdge Remote Response provider. Typical values are 6, 12, and 24 hours.

7. Click Save.

   A message indicates that the notification setup was saved.

Monitoring Alarms and Events

Events are generated to signify a health transition in a monitored device or component. For example, a health transition occurs when the state of a device goes from online to offline. The transition from online to offline, not the actual offline value, generates the event. If the state alone were used to generate events, the same events would be generated repeatedly.

Transitions cannot be used for monitoring log files, so log events can be repetitive. To minimize this problem, attach thresholds to entries in the log files. The software includes an event maximums database that keeps track of the number of events generated about the same subject in a single eight-hour time frame. This database prevents the generation of repetitive events. For example, if the port of a switch toggles between offline and online every few minutes, the event maximums database ensures that this toggling is reported only once every eight hours instead of every five minutes.

Events that require action are classified as alarms. For many alarms, information regarding the probable cause and recommended action can be accessed from the alarm view. In most cases, this information enables you to isolate the source of the problem. In cases where the problem is still undetermined, diagnostic tests are necessary. From the browser interface, you can execute tests remotely using the agent. Once the problem is fixed, you can clear the alarms for the device.

There are four alarm types:

- Down – Identifies a device or component as not functioning and in need of immediate service.

- Critical – Identifies a device or component in which a significant error condition is detected that requires immediate service.
- Major – Identifies a device or component in which a major error condition is detected and service may be required.
- Minor – Identifies a device or component in which a minor error condition is detected.

Note – If you use the Solaris operating system or the Netscape browser, you might not see the animation associated with the “barber pole” progress indicator when you view alarms and events. To animate the progress indicator, open the Netscape browser, select Edit > Preferences > Privacy&Security > Images, and select the radio button for looping the image as many times as specified.

Alarms

This section describes general tasks regarding alarms. For more information, click Help in the browser interface.

Viewing an Alarm Summary

To view alarm summary information, click Storage Automated Diagnostic Environment > Alarms.

The Alarms Summary page is displayed.

Viewing Alarm Details

To view detailed information about an alarm:

1. Click Storage Automated Diagnostic Environment > Alarms.

   The Alarms Summary page is displayed.

2. Click Details for the alarm for which you want to display detailed information.

   The Alarm Details page is displayed for the selected alarm.

3. Click the alarm icon on this page to view all the device specific alarms at this level.

Viewing Alarms for a Specific Device

To view detailed alarm information for a specific device:

1. Click Storage Automated Diagnostic Environment > Alarms.

   The Alarms Summary page is displayed.

2. In the Device column of the Alarm Summary table, click the name of the device for which you want to display alarms.

   The Device Details page is displayed.

Acknowledging Alarms

When an alarm is generated, it remains open in the alarms summary until you acknowledge it. Acknowledged alarms remain in the alarms summary until you delete them.

To acknowledge one or more alarms:

1. Click Storage Automated Diagnostic Environment > Alarms.

   The Alarms Summary page is displayed.

2. Select the check box for each alarm you want to acknowledge.
3. Click Acknowledge.

   The Acknowledge Alarms confirmation window is displayed.

4. Enter an identifying name to be associated with this action.
5. Click Acknowledge.

   The Alarms Summary page is re-displayed and the State of the acknowledged alarms is displayed as Acknowledged.

Reopening Acknowledged Alarms

To reopen one or more previously acknowledged alarms:

1. Click Storage Automated Diagnostic Environment > Alarms.

   The Alarms Summary page is displayed.

2. Select the check box for each acknowledged alarm you want to reopen.
3. Click Reopen.

   The Reopen Alarms confirmation window is displayed.

4. Enter an identifying name to be associated with this action.

5. Click Reopen.

   The Alarms Summary page is displayed and the State of the reopened alarms is displayed as Open.

Deleting Alarms

To delete one or more alarms:

1. Click Storage Automated Diagnostic Environment > Alarms.

   The Alarms Summary page is displayed.

2. Select the check box for each alarm you want to delete.

3. Click Delete.

   The Delete Alarms? confirmation window is displayed.

4. Click OK.

   The Alarms Summary page is displayed without the deleted alarms.

Generating Alarm Statistics Reports

To generate updated alarm statistics reports:

1. Click Storage Automated Diagnostic Environment > Alarms > Statistics.

   The Alarms Statistics page is displayed.

2. Click Generate All Reports.

   When the reports are complete, the following confirmation message is displayed:

   "Alarm Statistics Reports have been generated."

Displaying an Alarm Statistics Report

Alarm statistics provide event and alarm trends over days of the week, weeks, and months.

To display an alarm statistics report:

1. Click Storage Automated Diagnostic Environment > Alarms > Statistics.

   The Alarms Statistics page is displayed.

2. In the Actions column, click View for one of the following alarm statistic report types:
   - By device type and month
   - By device type and week
   - By device type and day of week

   The selected alarm statistics report is displayed.

Events

This section describes general tasks regarding events. For more information, click Help in the browser interface.

Displaying the Event Log

To gather additional information about an alarm, you can display the event log to view the underlying events on which the alarm is based.

To display event log information, click Sun Storage Automated Diagnostic Environment > Administration > Event Log.

The Events page displays a summary of all events in the system event log.

Viewing Event Details

To view event details:

1. Click Storage Automated Diagnostic Environment > Administration > Event Log.

   The Events page is displayed summarizing all events in the system event log.

2. Click Details in the row that corresponds to the event for which you want detailed information.

The Event Details page is displayed for the selected event, including information such as the severity of the event, the device/component to which the alarm applies, whether action is required, and an "Info" field that contains non-technical information about why the event occurred.

Viewing Aggregated Events

To view aggregated events for an alarm:

1. Click Storage Automated Diagnostic Environment > Alarms.

   The Alarms Summary page is displayed.

2. Click Details for the alarm for which you want to display detailed information.

   The Alarm Details page is displayed for the selected alarm.

3. Click View Aggregated Events.

   A list of all events associated with the displayed alarm is displayed.

About the Service Advisor

The Service Advisor details maintenance tasks for:

- The DSP
- DSP FRUs
- Array FRUs
- Cabinet FRUs
- Storage Service Processor
- USB flash disk

You can access Service Advisor procedures in two ways:

- If an alarm determines that you might need a Service Advisor procedure to resolve the problem, a hyperlink appears on the Alarm Details page. Clicking the link opens the Service Advisor page for that alarm in the context of the FRU that needs service.
- If you want to perform a service action or implement an option for a selected storage component that has Service Advisor procedures available, click the Service Advisor link in the masthead of the Storage Automated Diagnostic Environment browser interface. A pop-up window opens to the default page. Navigate down the tree in the left frame to find the selected component or procedure.

For more information, see the online help.

Best Practices: Using the Monitoring and Diagnostic Software

The following examples describe monitoring and diagnostic best practices:

- "Example 8-1. Monitoring the Health of the Sun StorEdge 6920 System" on page 177
- "Example 8-2. Upgrading Software and Firmware Using the Monitoring and Diagnostic Software" on page 179

Example 8-1. Monitoring the Health of the Sun StorEdge 6920 System

The Sun StorEdge 6920 system comes with Sun Storage Automated Diagnostic Environment software pre-installed to monitor the health of the system. To effectively use the monitoring features of the Sun StorEdge 6920 system, perform the following tasks:

- Configure local email notification
- Maintain the Alarms page

Prerequisites

Before you configure email notification and use the Alarms page, perform the following tasks:

- Verify site information
- Ensure that company and contact information is correct
- Manually initiate the agent run to verify system health

Note

- When setting up new email notification recipients on the Administration > Notification > Email Notification > Add Email Notification page, select Critical and above or Major and above. Selecting All or Informational causes notification to be sent for all events, including those that do not indicate a fault.

Procedure: Setting Up Email Notification

1. In the system's browser interface, click Sun Storage Automated Diagnostic Environment > Administration > Notification > Email.
2. Click New.
3. Fill out the Add Email Notification Form, using the following guidelines:
   - Type: Enter email or pager, depending on how you plan on receiving your alerts.
   - Email Address: Fill in your email address.
   - Categories: Select All Categories.
   - Priority: Select Critical and above. Any other priority saturates the email address with noncritical events.
   - Active: Select Yes.
   - Apply Email Filters: Select Yes.

Note – To set up email filters, use the Administration > Notification > Email Filters page.

   - Skip Components of Aggregated Events: Select No.
4. Click Save.

Procedure: Maintaining the Alarms Page

When the Sun Storage Automated Diagnostic Environment software generates an alarm, the administrator should acknowledge it and delete the alarm. If alarms are not deleted after problems are addressed, the current health of the system might not be apparent from the Alarm Summary page.

1. In the system's browser interface, click Sun Storage Automated Diagnostic Environment > Alarms > Summary
2. Read Alarm Details.
3. Do the following:
   - When an alarm is resolved, you may acknowledge it.
   - When an alarm condition is reviewed, you may delete it.

Example 8-2. Upgrading Software and Firmware Using the Monitoring and Diagnostic Software

The Sun StorEdge 6920 system enables you to upgrade the entire system while you are performing I/Os. During the upgrade, the DSP and the Sun StorEdge 6020 array controllers are taken offline one at a time. In this process, you might see a degradation in throughput. Additionally, external arrays might experience the interruption of individual links during the upgrade of the DSP.

Prerequisites

Before upgrading the system, answer the following questions:

- Will applications be hindered with I/O performance degradation during the upgrade?
- Are there any alarms that must be rectified before the upgrade?
- Are all system components in an optimal state?
- Is failover software correctly configured and operational on the host systems?

Notes

- Performing an upgrade when the I/O load is heavy is not recommended.
- Generate a new inventory before performing an upgrade. This ensures that only components with lower versions are targeted for upgrade.
- Upgrading a system that contains several arrays can take many hours due to numerous actions that the Sun StorEdge 6920 system must take to protect data availability. Upgrading each component individually can reduce the effect on host systems through better time management, but requires more administrative actions.
- External arrays without dual connections lose connectivity during the upgrade process.
- The back-out patch facility of the Storage Automated Diagnostic Environment software can take you down only to the previously installed patch level, so it should be thought of as an undo mechanism. If several patches were installed the last time, all of those patches are backed out.

Procedure

To upgrade software and firmware using the monitoring and diagnostic software:

1. Login to the Java Web Console as the "storage" user. and click Sun Storage Automated Diagnostic Environment.

   The Alarm Summary page displays a list of current alarms.

2. Check each of the alarms and take corrective action where needed to bring the system to its optimal state.

3. Set up your preferences for revision analysis:

   a. Click the Administration tab.

   b. Under the Revision Analysis Setup section, add the information that is required to set up the source from which you want to update your system. It is a best practice to use the Revision Analysis Server setting, as this gives you the latest baseline available for the system.

   c. When you have verified your settings, click Save.

4. Generate a current inventory:

   a. Click the Inventory tab.

   b. From the Actions list, select Generate New Inventory.

   c. Click Generate.

      This generates a list of the current firmware version on all the components.

   d. When the inventory has been gathered, click Save.

   e. Click the Rack Inventory link.

      This takes you back to the Inventory page.

5. From the Actions list, select Upgrade.

6. Click Generate Upgrade Report to identify which components must be upgraded.

7. When the report is displayed, upgrade the components that are recommended–one at a time.

   It is best practice to upgrade the Storage Service Processor first, then the internal arrays, and finally the DSP.

   You can view the progress of the upgrade process by clicking the Jobs tab and viewing the Revision Maintenance Upgrade report.

8. Generate an inventory again to capture the latest versions of the system.

Result

All recommended components within the Sun StorEdge 6920 system are upgraded. Your storage system is at the recommended baseline at the optimal software and firmware levels. No actions are taken on any external arrays attached to the system.

Troubleshooting Sun Storage Automated Diagnostic Environment Software

Following are issues that might occur with the Sun Storage Automated Diagnostic Environment software and recommended actions toward resolution.

About System Log Files

Messages related to the Sun StorEdge 6920 system are gathered in a number of log files. Sun StorEdge 6920 system-related log files include the following:

- The `/var/adm/messages` file on production hosts – Collects messages issued by the Solaris OS, including messages from Sun StorEdge Traffic Manager.
- The `/var/adm/messages.dsp` file on the Storage Service Processor – Collects messages from the Data Services Platform.
- The `/var/adm/messages.array` file on the Storage Service Processor – Collects messages from the Sun StorEdge 6020 system arrays.
- The `/var/tmp` log files on the Storage Service Processor – Java technology log files that are read by the `sscs` command. The Sun Storage Automated Diagnostic Environment software processes these log files to create lists of events and alarms, and to provide email and pager notifications.

You can also obtain system event information on the following tabs of the browser interface:

- The Event Log (Administration > Event Log) – Provides access to message file administration, event and alert log search capabilities, and Sun Storage Automated Diagnostic Environment agent error messages. Sun Storage Automated Diagnostic Environment uses event log files located in the `/var/opt/SUNWstade/DATA` directory.
- The Alarms tab – Provides access to a list of current alarms and alarm maintenance functions.
- The Administration tab – Provides access to administration functions for local and remote email notifications. It also provides access to event filters.

Diagnosing FRU Errors

To diagnose a FRU error within the Sun StorEdge 6920 system:

1. Gather information from one or more of the following messages:
   - Sun Storage Automated Diagnostic Environment alerts or email messages
   - Log files monitored by the Sun Storage Automated Diagnostic Environment, including the `/var/adm/messages.array` file on the Storage Service Processor
   - The `/var/adm/messages.dsp` file on the Storage Service Processor
   - The `/var/adm/messages` file on the Storage Service Processor
   - The `/var/adm/messages` file on production hosts
   - The Sun StorEdge 6920 system `/var/tmp` log files on the Storage Service Processor
   - Storage Service Processor log messages (obtained through the `sscs list log` command)

2. Determine the probable location of the problem using the following:
   - Review the Sun Storage Automated Diagnostic Environment Topology view.
   - Review the FC path status on production hosts by issuing the `luxadm` command (if using Sun StorEdge Traffic Manager).
   - Use the Sun Storage Automated Diagnostic Environment Revision Analysis Setup, Revision Maintenance, and Upgrade Report functions to check patch revisions in place on Sun StorEdge 6920 system components.

   Check the status of specific components in the Sun StorEdge 6920 system by doing the following:

   - Review the Sun Storage Automated Diagnostic Environment Event Advisor for errors by clicking an individual event or alarm details link.
   - Check the LED states on the Sun StorEdge 6920 system.
   - Correlate the LED states to LED status codes.
   - Check the appropriate documentation for status code definitions.

Verifying Storage Configuration

When isolating problems, compare the configuration information from the `cfgadm`, `luxadm`, and `format` commands from the production hosts with configuration information from the Sun StorEdge 6920 system. The monitoring and diagnostic tool provides a utility to extract configuration information from the Sun StorEdge 6920 system. When enabling the Sun StorEdge Traffic Manager or reconfiguring the Solaris OS, it is helpful to issue the `devfsadm -vC` command to clean up the unused paths in the device tree.

CHAPTER 9

Tuning System Performance

Tuning your Sun StorEdge 6920 system to optimally meet performance, capacity, growth, and availability requirements can be a challenging task. This chapter is intended to help you tune your system and to provide information to help you prevent issues or resolve them when they occur.

This chapter contains the following sections:

Understanding Performance Tuning

Properly sizing and configuring a modern storage system to optimally meet the performance, capacity, growth, and availability requirements of a business-critical application can be a challenging task, especially given the flexibility and opportunities to truly consolidate applications within monolithic-like storage systems. The Sun StorEdge 6920 system combines the flexibility of modular storage systems with the manageability and data services benefits of monolithic systems. These features help enterprises consolidate applications and better utilize storage resources. The Sun StorEdge 6920 system consists of three distinct components:

- Data Services Platform (DSP)
- Volume management
- Data services

Volume management and data services are provided by Sun StorEdge Storage Pool Manager and Sun StorEdge Data Snapshot software, respectively.

At a high level, the DSP of the Sun StorEdge 6920 system provides the ability to separate the physical structure of the storage devices (arrays, LUNs, disks) from the presentation layer (what the hosts that use the storage see). By masking the physical layout from the presentation layout, storage and system administrators can work around the historical problem of being unable to use all storage effectively. This model also enables the deployment of advanced storage area network (SAN)-based storage and data management applications (for example, SAN-based snapshots). There is no longer a direct connection between the application server view of storage and the physical installation (storage virtualization). As a result, applications like snapshots, data replication, or transparent data migration can be added without consuming host resources.

Monitoring Using the Sun StorEdge Remote Response Service

The Sun StorEdge Remote Response service provides remote monitoring of the Sun StorEdge 6920 system. It can be configured to monitor on a 24-hour basis, collecting information that enhances the reliability, availability, and serviceability (RAS) of the Sun StorEdge 6920 system. This service is delivered with the Sun StorEdge 6920 system. Upon activation of the Sun StorEdge Remote Response service provides preemptive and preventive support for the Sun StorEdge 6920 system. The primary goal of the service is to provide early detection of component or system anomalies, allowing faster response and reaction to alarms. In addition, the service provides remote troubleshooting, diagnosis, and repair of Sun StorEdge products.

The Storage Service Processor continually monitors the messages sent to the system log by the software and firmware in the subsystems. If a message is determined to contain an event that requires an alert, the modem within the Storage Service Processor accessory tray is initialized to send a message to a Sun Service Center. Then, the call-home data is analyzed and a case is generated.

All hardware and software required for using the SSRR is included with your system. The modem installed in the base cabinet is qualified for use in most countries with no modification required.

Note – If remote monitoring by means of the service is not possible, configure RAS telemetry for your system. The type of data sent includes information such as firmware levels, WWNs, serial numbers, event logs, and system uptime. Customer data is not sent. The telemetry data tells Sun if your system requires implementation of a Field Information Notice (FIN), Field Change Order (FCO), or firmware upgrade. To set up encrypted email for RAS telemetry, refer to the instructions in "Setting Site Information for Diagnostics and Monitoring" on page 169 to configure the Network Storage Command Center (NSCC) remote provider service.

For more information, see "Sun StorEdge Remote Response Service" on page 33.

Monitoring Using Storage Automated Diagnostic Environment Software

Storage Automated Diagnostic Environment Enterprise Edition is a distributed diagnostic monitoring tool for SAN devices, including data hosts and external storage devices connected to the Sun StorEdge 6920 system. You can configure the software to monitor on a 24-hour basis, collecting information that enhances the RAS of the storage devices.

Use the Storage Automated Diagnostic Environment Enterprise Edition software, installed on a data host or management station, to monitor the health of external storage devices. Storage Automated Diagnostic Environment Enterprise Edition is delivered on the Host Installation Software CD. Select Sun StorEdge 6920 Remote Management Host Software to install the remote scripting CLI client.

The remote scripting CLI client provides access to the same configuration and monitoring tasks that you perform through the browser interface. You can also write scripts using the `sscs` commands to automate configuration and management tasks. You can install the remote scripting CLI client on a data host or an external management host. See the `sscs` (1M) man page for information about the remote scripting CLI commands.

The remote scripting CLI client is also available for the Microsoft Windows, Red Hat Linux, AIX, and HP-UX operating environments.

Note – Only encrypted services are provided on the Sun StorEdge 6920 system; the monitoring and diagnostic software as well as the management software support only encrypted (HTTP) communications.

Note – The application data path is used exclusively to process I/O between the application host and disks. No configuration or monitoring is performed over the data path, other than normal SCSI inquiry requests.

For more information about the Storage Automated Diagnostic Environment, see Chapter 8.

Online Transaction Processing (OLTP)

Databases for online transaction processing (OLTP) are usually characterized by very high amounts of small, random, nonsequential read and write transactions within concentrated time periods. These transactions make the applications very sensitive to latency delays, which impose a time penalty on each transaction and result in slow user response times. To alleviate these delays, configure the system with ample amounts of RAID controller cache to provide fast write operations. In addition, I/O can be spread across multiple spindles (or disk drives) to optimize parallel I/O for both reads and writes. Generally speaking, 36-gigabyte 15,000-RPM drives are used for OLTP-type databases.

Decision support services (DSS) databases are usually characterized by very high amounts of large sequential I/O. This makes them very cache-dependent and bandwidth-dependent. The Sun StorEdge 6920 system provides up to 28 gigabytes of mirrored cache and 800 megabytes per second of bandwidth. Generally speaking, 73-gigabyte or 146-gigabyte 10,000-RPM drives are used for DSS-type databases.

Note – For both OLTP and DSS databases, RAID-5 is used for read-heavy environments, and RAID-1 (1+0) is used for write-heavy environments.

Read-Ahead Settings

The Read-Ahead setting is applied at the storage array level and applies to the array controller firmware:

- Read Ahead = on: For a block of I/O coming in, read ahead is always performed (the entire cache buffer is read), regardless of whether subsequent blocks of I/O are sequential.
- Read Ahead = off: For a block of I/O coming in, two subsequent blocks of I/O are checked to see if they are sequential. If so, read ahead is performed (the entire cache buffer is read).

When Read-Ahead is enabled, the Sun StorEdge 6020 array continues to read data to fill out the remainder of the cache block associated with the user data blocks that are requested. Larger cache blocks frequently execute more aggressive read-aheads, but they utilize cache space less efficiently than small blocks.

Any I/O of a sequential nature can benefit from read-aheads. Since most data is backed up frequently, and backup is generally sequential, it is advisable to set Read-Ahead to on in nearly every situation. Because the Sun StorEdge 6020 arrays make this setting controller-wide, it must be the same for all drives serviced by a particular controller tray pair.

Messaging Applications

Email applications combine the workload characteristics of OLTP and DSS databases. I/O is highly random and therefore unpredictable, and record sizes range from small text-only messages to large messages with file attachments. These environments require a storage system capable of performing both transaction-heavy and throughput-heavy I/O.

The Sun StorEdge 6920 system can be scaled to optimize I/O request operations, data movement, or a combination of both. I/O optimization does not sacrifice data protection, making this system an ideal storage solution for messaging applications.

I/O Queue Depth

One limit on a host's demand for I/Os is its I/O queue depth. Queue depth is the number of I/O requests waiting to be completed, also known as outstanding I/Os. A higher queue depth can significantly increase overall storage performance because of the storage system's ability to "hide" queued I/O activity from the application. The application does not have to wait for each I/O to complete before continuing on with its next operation.

Further, if server I/O activity is single-threaded, performance is gated by each individual I/O request, which typically has a response time of 1 millisecond. Thus, the storage system produces only 1000 IOPS.

To use a storage system to its potential, you need a much more robust I/O generator tool. Sun StorEdge vdbench is an I/O generation application that has many options and can be programmed to drive a storage system with almost endless variations and load combinations.

Note – When a detailed level of analysis and display of data is required of a storage system, use the Sun StorEdge Workload Analysis Tool (SWAT) rather than the UNIX command `sar`. SWAT has the ability to display pie charts of controller activity.

Example 9-1. Using OLTP to Tune Performance

OLTP can take many forms. It can be an Oracle Database application, Microsoft Exchange Mail Server, or Sybase application. These applications have a mixed workload of many small-block transfers and a few large-block transfers. Such a mixed workload is challenging to configure. In addition, the response time of certain files can limit the overall application performance. Files such as redo logs, temp files, and archive files frequently affect application performance.

This example describes a best practice for using OLTP to tune the performance of your Sun StorEdge 6920 system.

Single-Threaded Small-Block Transfers

The Sun StorEdge 6920 system performs well with large-block transfers and many concurrent I/Os. With a large number of controllers and drives in a Sun StorEdge 6920 system, it takes a large number of concurrent I/Os to maintain activity to all the controllers and drives.

Redo logs and application temp files are often small-block transfers and single-threaded (only one outstanding I/O transaction at any time). The throughput to these files is driven by the response time from the storage. If the storage has a response time of 2 milliseconds, throughput to the file is around 500 I/O transactions per second (IOPS). With a response time of 0.5 milliseconds, throughput is 2000 IOPS. For optimal throughput, put these small files on solid-state disks or on very fast, low-latency direct attached storage.

If you are unable to acquire dedicated fast storage for these files, place them on their own controller in the Sun StorEdge 6920 system to avoid contention with the data portions of the application.

Controller Striping

Striping is a technique of mapping data so that the data is interleaved among two or more physical disks. More specifically, data is allocated in equal-sized units (called stripe units) that are interleaved between the disks. Each stripe unit is a set of contiguous blocks on a single physical disk, as shown in FIGURE 9-1.

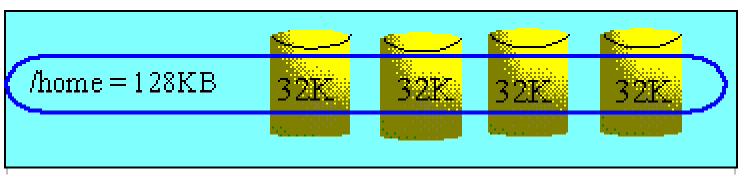

FIGURE 9-1 Disk Striping at the Controller Level

Striping is useful if large amounts of data must be written to or read from the physical disks quickly; using many parallel data transfers to multiple disks does this. Striping is also helpful for balancing the I/O demands of multithreaded workloads across multiple disks.

Drive Transfer Rates and Mechanical Latency

Drive transfer rate is another variable that affects the performance of a storage system. The newer 15,000-RPM disk drives are designed to improve mechanical latency, and this has improved their data transfer rates up to 57 to 85 megabytes per second.

The larger the size of the chunk of data being transferred to these 15,000-RPM drives, the more efficient the transfer process is. This is because it takes a fixed amount of time to get to the data and put the head of the drive at that point (seek and latency), regardless of the size of the chunk of data being transferred. FIGURE 9-2 depicts the effect of drive stripe size on the performance of data transfers.

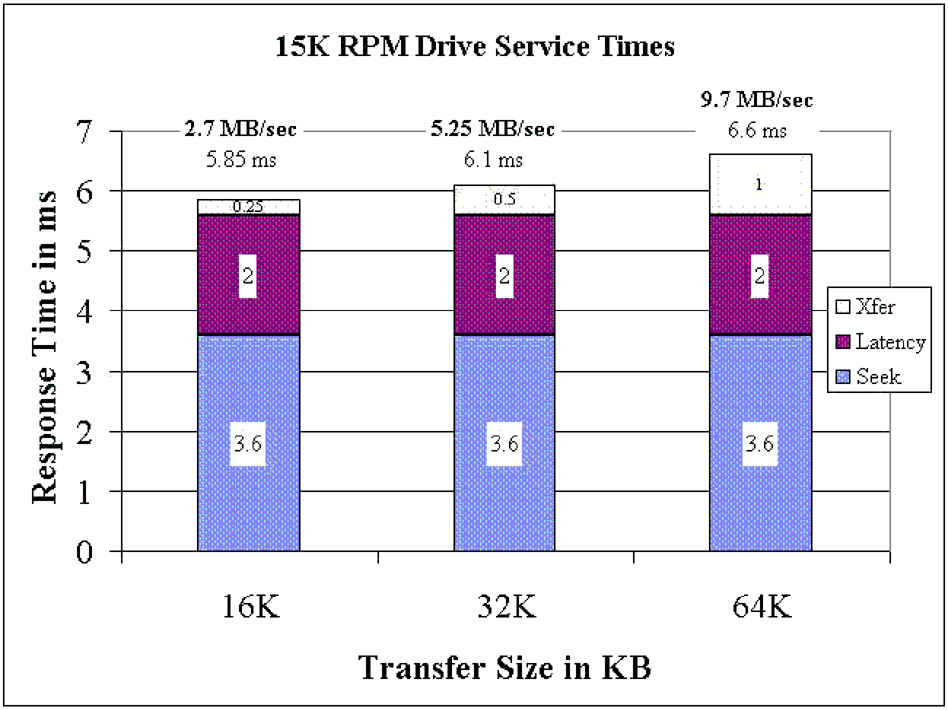

FIGURE 9-2 Drive Service Times

In addition to selecting large database block sizes, you can configure the segment size and stripe size to support large-block transfer sizes.

- Configure segment sizes to 32 kilobytes or 64 kilobytes.
- Configure the stripe size to be one of the following:
  - For RAID-0 and RAID-1: 1 to 2 times the number of data drives times the segment size.
  - For RAID-5: 2 to 4 times the number of data drives times the segment size.

To increase concurrency, many applications offer application concurrency variables. For example, Oracle 10 has variables such as `recovery_parallelism`, `job_queue_processes`, `parallel`, and `db_file_multiblock_read_count` to increase concurrency and increase data transfer sizes.

To minimize I/Os in the first place, consider increasing cache sizes as much as server resources allow.

Backup Strategies Using Sun StorEdge SAM-FS Software

Sun StorEdge SAM-FS software is an advanced storage file system that provides users easy access to all their files, regardless of whether the files are physically stored on disk or tape. Online, nearline, or offline files all appear to the user in the file system and can be retrieved from their designated storage medium upon request.

With built-in, enterprise scalability, Sun StorEdge SAM-FS software supports rapid growth in large-scale environments. It enhances traditional backups to improve storage resource utilization for applications in which data needs to be available continuously and quickly restored in the event of a business disruption. Sun StorEdge SAM-FS software's continuous archiving services enable customers to manage growing volumes of data based on the value of the data to the business, while dramatically increasing service levels, improving IT efficiency, reducing complexity, and lowering costs to improve the bottom line. The software increases customer control over the data environment by enabling centralized, policy-based management of files across storage media and geographic sites.

Sun StorEdge SAM-FS software offers the following advantages to help overcome the problems of traditional backup strategies:

- Rapid recovery

  The software recovers quickly from application and system downtime without data corruption or loss.

- Improved data protection

  Up to four copies are made automatically when a file is closed. These copies can be stored on local or remote disk, tape, or optical storage based on customer-defined policies.

- No interruption to file system

  Sun StorEdge SAM-FS software continuously, automatically, and transparently makes copies of new or changed files to tape or other secondary media.

- Data protection for easy recovery without a full backup

  Copies of only new and changed files are made to media on local or remote tape and magneto optical libraries.

- Multi-threaded architecture

  Clients and servers remain in the same configuration they had before. The only change is that the backup application is reconfigured so that it directs all backup files into a Sun StorEdge SAM-FS file system instead of going directly to tape.

Sun StorEdge SAM-FS software makes copies (up to four) of a file only when it is new or if it has changed. The software automatically backs up files as they are changed, eliminating the need for a traditional backup. Users and applications continue to have access to the file system while files are copied. Because new and changed files are copied only once and not repeatedly, system overhead and media usage are much more efficient than in common backup strategies.

The software stores metadata in a location separate from the data files. In the event of a system failure, the entire file system structure can be reconstructed in minutes. This represents a tremendous advance over traditional file system and data protection software products that might require hours or even days to rebuild the entire file system. Files can be transparently restored from near-line media, providing on-demand restore capabilities. The data portion of files not immediately accessed can be restored in the background while users and applications are working normally in the file system.

Upgradable Components

The Sun StorEdge 6920 system has the following upgradable components:

- Data Services Platform (DSP) image
- Sun StorEdge 6020 array controller and loop card image
- Sun StorEdge 6020 array disk image
- Monitoring and diagnostic software image
- Storage Service Processor image
- Router and firewall image

For optimal operation of the system, upgrade all components when patches are available. Once you begin to upgrade the system, continue until all components are updated.

Note – If a failure occurs when you are upgrading firmware on a system component, attempt the upgrade again. If the second upgrade attempt fails, all components that were previously upgraded should be downgraded.

The monitoring and diagnostic software tracks version levels of Sun StorEdge 6920 system components to ensure that appropriate firmware is installed and necessary configuration information is restored after a new Sun StorEdge 6920 system component or field-replaceable unit (FRU) replacement is added. The monitoring and diagnostic software discovers the new hardware and invokes the software that tracks version levels of individual components within the Sun StorEdge 6920 system. After confirmation from the user, this software upgrades or downgrades the component firmware as necessary and then runs an installation script to restore the configuration information.

Note – The monitoring and diagnostic software's back-out patch facility can take you down only to the previously installed patch level, so it should be thought of as an undo mechanism. If several patches were installed the last time, all of those patches will be backed out.

The Sun StorEdge 6920 system supports online firmware upgrades of the components listed above, except the Sun StorEdge 6020 array disk image. There is full data availability during the firmware upgrade operation.

During an upgrade of the disk image firmware, the applicable Sun StorEdge 6020 array must be taken offline. During the upgrade process, I/O exchanges can occur and access to data is not possible.

- **Remote lights-out power management** – The Sun StorEdge 6920 system supports remote lights-out power management with two forms of shutdown:
  - **Remote partial shutdown** – The DSP and Sun StorEdge 6020 arrays can be remotely shut down. The Storage Service Processor, accessory tray, and Ethernet hub are kept running to allow for a subsequent remote power-on operation.
  - **Remote full shutdown** – A full shutdown, including the Storage Service Processor, can be initiated. However, powering on of the system must be done manually at the site location of the system, the system cannot be restored remotely.
- **Booting** – The Sun StorEdge 6920 system supports warm booting. After a partial shutdown, the DSP reboots because its management interface card (MIC) is still up.
- **Depopulated trays** – The Sun StorEdge 6920 system supports depopulated trays. Each tray can have a minimum of seven drives.

Note – It is a best practice to add drives in increments larger than one.

- **Addition of trays to existing storage arrays** – You can add controller-less trays, referred to as expansion trays, to existing Sun StorEdge 6020 arrays. Only 2x2 and 2x4 arrays can accommodate expansion trays.

  When expanding arrays, you must shut down the arrays before the conversion is performed. This is not an online procedure.

Mixed Drive Sizes Within Trays

In the Sun StorEdge 6920 system, different drive sizes can be utilized within each tray. However, the same drive size should be used within a virtual disk. Each tray can have a maximum of two virtual disks.

Note – If different drive sizes are used within a virtual disk, the lowest drive size dictates the usable capacity of each drive in the virtual disk. For example, if a seven-drive virtual disk has six 73-gigabyte drives and one 36-gigabyte drive, only 36 gigabytes of capacity on each of the seven drives is recognized. In addition, the drive size of the hot-spare (if a hot-spare is used), whether an array or dedicated hot-spare, must be greater than or equal to the smallest drive sizes in the virtual disk in the tray.

Solutions for Storage Bottlenecks

Reducing storage system bottlenecks can improve overall end-to-end performance and assist in meeting overall business performance requirements. This section addresses common storage system bottlenecks. The best practices discussed can help you do the following:

- Increase end-to-end performance by taking advantage of multithreaded I/Os to the storage system controller and its drives
- Deliver predictable storage system performance by matching the storage system configuration to workload requirements
- Increase storage system throughput by distributing a workload across the storage system

In today's competitive business world, you need to obtain optimum end-to-end performance from your computing solutions (with end-to-end defined as the servers, the storage, and the applications in between). To achieve optimal performance, you must configure each computing resource in the end-to-end solution to meet application workload needs and specific business performance requirements. To accomplish this requires understanding the most critical pieces of the end-to-end computing solution puzzle: the storage system.

Storage system performance typically has a significant impact on the overall application and computing system performance. Consequently, typical storage system performance criteria (for example, I/O rate, response time, and data rate) determined by the storage system workload.

A typical storage system is a hierarchical layering of software and hardware. The hierarchy starts at the server (hardware) level where the applications (software) reside. Applications must access data, and they have a multitude of ways of doing so, enabled by a file system or raw device files in the server operating system (OS).

A file system is a physical partition on a logical unit number (LUN), which is a combination of disks or storage systems visible to the server. A LUN is created by an OS command and includes metadata that the OS uses to identify where to write files and other important file information, such as inodes, blocks, and super blocks. An additional benefit of a file system is that it contains a file system cache that speeds access to data. The cache is employed to access frequently used data, thus bypassing the slower physical disk drive.

A raw device file is a physical partition on the disk drive that has no file system. However, the data on a raw device cannot be viewed or accessed by users as can be done with a file system. A raw device file is most commonly implemented to improve I/O performance and is generally used by databases.

File systems can become very large, and the best way to manage them is to implement a volume manager. A volume manager is an application that manages disks by combining disk sectors into volumes, which can then be subdivided or combined into RAID sets for redundancy and performance. Additionally, the volume manager is typically used to help spread data across many devices, thereby increasing performance and redundancy.

Sun StorEdge Workload Analysis Tool (SWAT)

The Sun StorEdge Workload Analysis Tool (SWAT) is a graphical user interface (GUI) Java application that collects, processes, and reports storage performance information about disk I/O workloads. In the Solaris Operating System, the tool uses the Solaris trace normal form (TNF) utility. In the Microsoft Windows operating system, the tool uses the Tracelog utility. The information can be displayed at a variety of levels and uses many different graphs and charts to identify recurring performance problems. In the Solaris Operating System, the tool can also capture iostat/kstat level storage performance data that can be used for a less detailed analysis than can be obtained directly from I/O trace data. The information gathered is stored on the system where the data has been collected.

This data, which includes transfer sizes, queue depths, cache hit rates, and skew across devices, helps administrators understand the behavior of storage system workloads. In turn, this provides valuable insight into how best to configure the storage system to meet the demands of the workload.

Sun StorEdge Vdbench

Sun StorEdge vdbench is a command-line, Java-based, synthetic I/O driver that is portable across multiple platforms. It can be used to replicate and approximate workload performance and throughput. Vdbench replaces tools such as Veritas vxbench.

Vdbench can be used to do the following:

- Validate software package installation and verify connections by generating I/O through those connections
- Benchmark by varying variables in a workload
- Play back a real customer workload captured through SWAT
- Validate performance and throughput capabilities of a storage system
- Set I/O rate to maximum, exponential, or fixed inter-arrival time
- Set transfer size to fixed or varied
- Control many aspects of a workload, including the following:
  - Percent read – Controls workload read-to-write ratio.
  - Queue depth – Controls the number of active threads for a given workload.
  - Percentage of read hit – Controls and the number of reads taken from the storage system's controller cache compared the slower mechanical disk drives.
  - Percentage of write hit – Controls and the number of writes to the storage system's controller cache, compared with writing to the slower mechanical disk drives
  - Percent random – Determines the percentage of I/O requests involving mechanical disk overheads known as "seek" and "latency," which significantly affect system performance.

Another benefit of vdbench is its ability to provide real-time feedback during benchmark execution. This enables the system under test to provide immediate benchmark progress. Vdbench displays the following information:

- I/O rate
- Response time
- Data rate
- Read-to-write ratio
- Maximum service time
- Service time standard deviation

Finally, vdbench provides other valuable data as text and HTML files that can be easily imported into spreadsheet packages, allowing graphing of virtually any parameters. This permits further analysis and facilitates the sharing of performance information with others.

Single-Threaded and Multithreaded I/O Requests

The UNIX command `dd` is single-threaded because it requests one I/O at a time. A way to increase storage system I/O demand is to request multiple simultaneous transactions; this is commonly referred to as multithreading. With multithreading, the storage system can supply more I/O transactions to the server, increasing its overall performance.

FIGURE 9-3 shows an example of an application requesting 8 chunks of data in single-threaded format, compared with one requesting 8 chunks of data in multithreaded format. The single-threaded request takes 8 milliseconds to complete, while the multithreaded request requires only 1 millisecond to complete.

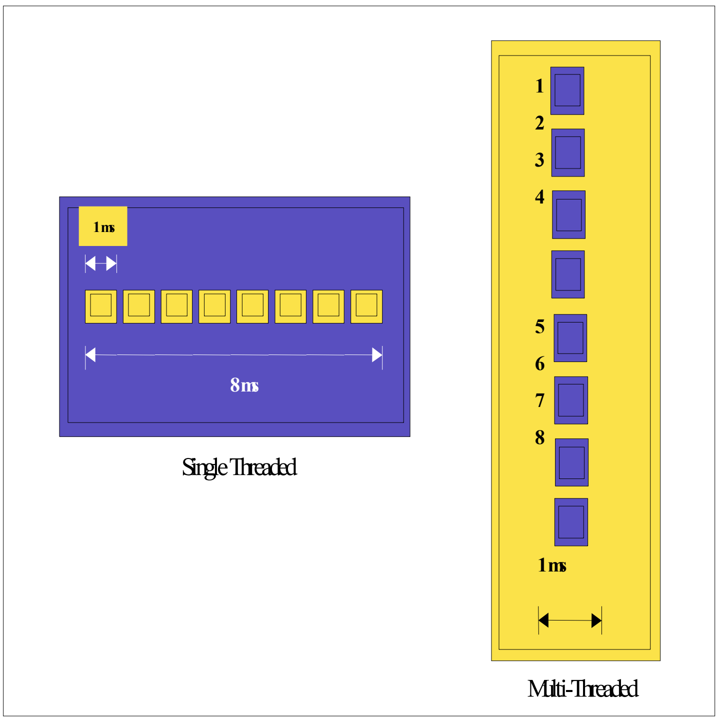

FIGURE 9-3 Single-Threaded and Multi-Threaded I/O Requests

To demonstrate the effects of multithreading I/O requests, an experiment was run using Sun's I/O generator tool, the Sun StorEdge vdbench, and a Sun StorEdge 3510 system with a 6-drive LUN configuration. Several simple benchmarks were run, starting with a demand for only a single thread (I/O) of the storage system. This was subsequently increased to 8 threads, and finally to 16 threads. Requesting a single thread produced 146 IOPS, 8 threads produced 707 IOPS, and 16 threads produced 967 IOPS, at a 562 percent performance improvement. Increasing the thread count, within reason, can dramatically improve storage system performance.

Transfer and Drive Access

To understand the effects that small storage controller stripe sizes have on large block write requests, you must understand the disk drive configuration itself. As an example, if the disk drive is configured with small stripe sizes such as 16 kilobytes, and a write request of 128 kilobytes is made, 8 drives must be accessed to complete that write (128/16=8). If the stripe size is set to 32 kilobytes, then 4 drives must be accessed. Similarly, if the stripe size is set to 64 kilobytes, then only 2 drives need be accessed to complete the 128-kilobyte write.

Impact of Stripe Size

To substantiate this theory of the impact of disk stripe size, a lab exercise was performed with a storage system that was configured with an 8-drive RAID-5. The configuration started with a disk drive stripe size of 16 kilobytes. Next, a workload of writes was applied to the storage system. The workload started with writes of 16-kilobyte blocks or smaller. It was then increased to between 16-kilobyte blocks and 32-kilobyte blocks and so on, until the workload demand was above 500-kilobyte block writes. The exercise was repeated for 32-kilobyte stripe sizes and 64-kilobyte stripe sizes.

The results showed that use of a larger physical stripe size seems to be the most efficient configuration to use, regardless of the I/O request size. Further, a large I/O request response time is much better with large stripe sizes than with small stripe sizes, since fewer drives need to be accessed to complete the I/O requests.

Small Host-Based Stripe Sizes

Creating small stripe sizes across many controllers might hinder the storage controllers' ability to perform. One of these improvements that is typically negated by many small stripe sizes is "parity-on-the-fly."

Parity-on-the-fly is a specially designed performance optimization of the storage controller's RAID-5 algorithm. It is used for write operations, and it enables parity to be calculated without the disk drives being accessed. Parity-on-the-fly is activated for a write operation when all the data in a RAID-5 group is stored in the storage controller cache. The cache algorithm collects the full RAID-5 group data from the cache, and then calculates the new parity from that data rather than taking time to access the disk drives, to retrieve the old data and old parity, and then to calculate the new parity.

Large Host-Based Volume Stripe Sizes

If you choose too large a stripe size for a host-based volume, all I/Os are forced to funnel through a single controller causing an unbalanced load. For example, if a file of 16 megabytes needs to be written to disk, and the volume it belongs to has been created with stripe sizes of 16 megabytes, then all I/O activity goes through one, and only one, controller, causing 100 percent skew. A single controller does all the work while all other controllers remain inactive.

Best Practices: Volume Configuration

To increase the efficiency of how the host and storage controller volume managers work together, configure them using the following guidelines:

- **RAID-5** – Configure the host volume manager stripe size using sizes from 2 to 4 times larger than the storage controller stripe size.
- **RAID-0 and RAID-1** – Configure the host volume manager stripe size using sizes from 1 to 2 times larger than the storage controller stripe size.

APPENDIX **A**

System Usage Limits

TABLE A-1 lists maximum values for elements of the Sun StorEdge 6920 system.

TABLE A-1 Sun StorEdge 6920 System Limits

System Attribute	Maximum
Volumes per system	1024
Virtual disks per tray	2
Volume partitions per virtual disk	32
Mirrored volumes	128 (256 mirrored components)
Components in a mirror	4, including the primary component
Legacy volumes	128
Snapshots per volume	8
Initiators that can communicate with the system	256 **Note:** If a data host-side HBA port sees *N* ports, the system sees *N* initiators. The 256-initiator limit translates to a maximum of 128 dual-path data hosts, where each data host HBA port can see one port of the system.
Initiators per DSP port	126
Data host HBA ports that can communicate with one system port	126
Volumes that can be mapped to a single data host HBA port World Wide Name (WWN)	256

TABLE A-1 Sun StorEdge 6920 System Limits *(Continued)*

System Attribute	Maximum
Storage domains	14 (1 system defined; 13 available for user definition)
Storage pools	64
Storage profiles	14 system-defined storage profiles; no limit for user-defined profiles

Glossary

Definitions obtained from the Storage Networking Industry Association (SNIA) Dictionary are indicated with “(SNIA)” at the end. For the complete SNIA Dictionary, go to www.snia.org/education/dictionary.

agent The component of the system monitoring and diagnostic software that collects health and asset information about the system.

alarm A type of event that requires service action. See also event.

array Multiple disk drives that function as a single storage device. A high-availability (HA) array configuration has multiple controller and expansion trays of disk drives. See also dedicated hot-spare and hot-spare.

array hot-spare A spare disk within an array that is used for failover when any of the active disks fail. See also dedicated hot-spare and hot-spare.

array type An internal storage array configuration that is defined by the number of controller units and the total number of storage trays. For example, a 2x4 storage array configuration consists of a total of four storage trays, two of which are controller trays.

asynchronous queue In the context of data replication, a queue used to store writes that are to be replicated to the remote site. After the writes have been put into the queue, the writes are acknowledged to the application and then forwarded to the remote site as network capabilities permit. The asynchronous queue is a persistent queue, so in the event of a disaster at the primary site, the data in the asynchronous queue is not lost.

asynchronous replication A form of data replication in which application write operations are written to the primary site and to the asynchronous queue on the primary site. The asynchronous queue forwards queued writes to the secondary site as network capabilities permit. The write operations to the primary site are confirmed, regardless of when, or whether, they are replicated successfully to the secondary site. Deferring the secondary copy removes long-distance propagation delays from the I/O response time. See also synchronous replication.

autosynchronization An option enabled at the primary site that attempts to synchronize replication sets or consistency groups whenever a link is established. With autosynchronization, synchronization continues even if there are link errors, for example.

base cabinet The system's main cabinet, which contains a Data Services Platform (DSP), storage trays, a Storage Service Processor with a USB flash disk, a Storage Service Processor accessory tray, a service panel, and power distribution units (PDUs). The base cabinet is pre-wired with Ethernet, Fibre Channel, and power cables and can be connected to one or two expansion cabinets. See also expansion cabinet.

break To remove a mirror component from the mirror, voiding its relationship with the other mirror components. The broken mirror component becomes a standalone volume in the system, and synchronization with the other mirror components is not maintained. See also mirror component and split.

captive storage See internal storage.

combo card See storage I/O (SIO) card.

component See mirror component.

concatenation A storage allocation method in which sequential blocks on a disk are linked together to form a larger logical device. This method combines the storage potential of several physical devices. See also striping.

consistency group A collection of replication sets grouped together to ensure write order consistency across all the replication sets' primary volumes. An operation on a consistency group applies to all the replication sets within the consistency group, and consequently their volumes.

controller pair A pair of controller units that services a group of storage trays.

controller tray A storage tray with an installed RAID controller and up to 14 disk drives. In a Sun StorEdge 6920 system, a pair of controller trays is the smallest possible storage array configuration, the 2x2 array type. See also expansion tray.

controller unit The card that manages RAID functions and failover characteristics for an array.

control path The route used for communication of system management information, usually an out-of-band connection. See also out-of-band management.

copy on write A technique for maintaining a point in time copy of a collection of data by copying only data which is modified after the instant of replicate initiation. The original source data is used to satisfy read requests for both the source data itself and for the unmodified portion of the point in time copy. (SNIA) See also snapshot reserve space.

customer LAN See site LAN.

DAS See direct attached storage (DAS).

data host Any host that uses the system for storage. A data host can be connected directly to the system (direct attached storage, or DAS) or can be connected to an external switch that supports multiple data hosts (storage area network, or SAN).

data path The route taken by a data packet between a data host and the storage array. See also in-band management.

data replication A disaster recovery and business continuance method in which a primary volume at the local site and a secondary volume at a remote site contain the same data on an ongoing basis, thereby protecting user data.

Data Services Platform (DSP) The controller component of the Sun StorEdge 6920 system, which consolidates and virtualizes storage so that all storage in the system can be managed as a single scalable entity.

data striping See striping.

dedicated hot-spare A disk that serves as the hot-spare for one and only one virtual disk in a storage tray.

degraded The condition of a volume in which one or more input or output data paths are not operating properly. Although the redundant failover paths are still intact, a degraded volume holds no significant value for the storage configuration and should therefore probably be deleted from the system.

dependent copy See snapshot.

direct attached storage (DAS) A storage architecture in which the systems that store data are connected physically to storage elements.

disk A physical, non-volatile, rewritable data storage device. See also virtual disk.

domain See storage domain.

DSP See Data Services Platform (DSP).

electro-magnetic interference Radiated electro-magnetic signals that can interfere with the normal transmission of information.

EMI See electro-magnetic interference.

event Any condition reported by a device to the system monitoring agent. See also alarm.

expansion cabinet A cabinet, pre-wired with Ethernet, Fibre Channel, and power cables, connected to the base cabinet to increase storage capacity. A power management cable connects the service panels of the cabinets, and the Fibre Channel/Ethernet cable connects the I/O panels of the cabinets. See also base cabinet.

expansion tray A storage tray that has up to 14 disk drives, but does not have a RAID controller. This type of storage tray is used to expand the capacity of an array and must be attached to a controller tray to function. See also controller tray.

extend In the context of managing volumes, to increase a volume's capacity.

extent A set of contiguous blocks with consecutive logical addresses on a physical or virtual disk.

external storage A physical disk or storage array connected to the Sun StorEdge 6920 system located outside of the base or expansion cabinets.

fabric A Fibre Channel switch or two or more Fibre Channel switches interconnected in such a way that data can be physically transmitted between any two N_Ports on any of the switches. (SNIA)

failover and recovery The process of changing the data path automatically to an alternate path.

Fast Start operation An option of the suspend operation and the procedure in which a method such as a backup tape is used to copy data from the primary volume to the secondary volume. This procedure is used to avoid the initial step in which you send the primary volume's data over the physical link. For example, network bandwidth might justify a Fast Start procedure. See also resume operation and suspend operation.

fault signature analysis An algorithm applied by the diagnostic and monitoring software to a specific set of events that can be correlated by time and locality to a specific cause. Fault signature analysis assumes that the most significant event is the most probable cause and aggregates the remaining events underneath that event to improve the signal/noise ratio when a single probable cause might result in a multitude of events.

FC See Fibre Channel (FC)**.**

FC port See Fibre Channel (FC) port.

FC switch See Fibre Channel (FC) switch.

Fibre Channel (FC) A set of standards for a serial I/O bus capable of transferring data between two ports at up to 100 MBytes/second, with standards proposals to go to higher speeds. Fibre Channel supports point to point, arbitrated loop, and switched topologies. Fibre Channel was completely developed through industry cooperation, unlike SCSI, which was developed by a vendor and submitted for standardization after the fact. (SNIA)

Fibre Channel (FC) port A port on the I/O panel that connects data hosts, external storage, or internal storage to the Sun StorEdge 6920 system. See also host port and storage port.

Fibre Channel (FC) switch A networking device that can send packets directly to a port associated with a given network address in a Fibre Channel storage area network (SAN). Fibre Channel switches can be used to expand the number of data host or external storage device connections. Each switch is managed by its own management software.

field-replaceable unit (FRU) An assembly component that is designed to be replaced on site, without the system having to be returned to the manufacturer for repair.

flash card or disk See USB flash disk.

force break To remove a mirror component (mirrored volume) from the mirror before the resilvering process is complete, causing the condition of its data to be degraded. Just as with a break operation, the mirror component for which you implement a force break becomes a standalone volume in the system; however, it is inaccessible and should probably be deleted. See also break, degraded, mirror component, and resilvering.

free The state of a volume that is not mapped to initiators.

FRU See field-replaceable unit (FRU).

FSA See fault signature analysis.

full synchronization A resume operation in which a complete volume-to-volume copy occurs. Unlike a normal resume operation, in which a copy of differences between the primary and secondary volumes occurs, a full synchronize operation copies the entire contents of the volume. The system performs a full synchronize operation the first time you resume data replication on a replication set. See resume operation and synchronization.

HBA See host bus adapter (HBA).

host bus adapter (HBA) An I/O adapter than connects a host I/O bus to a computer's memory system. (SNIA) See also initiator.

host port A port on the I/O panel that connects to a data host. See also storage port.

hot-spare The drive used by a controller to replace a failed disk. See also array hot-spare and dedicated hot-spare.

in-band management Transmission of a protocol other than the primary data protocol over the same medium as the primary data protocol. Management protocols are a common example of in-band transmission. (SNIA)

The Sun StorEdge 6920 system uses an in-band management path between hosts and the storage arrays to transport both data and management traffic. See also out-of-band management.

independent copy See mirror component and mirroring.

initiator A system component that initiates an I/O operation over a Fibre Channel (FC) network. If allowed by FC fabric zoning rules, each host connection within the FC network has the ability to initiate transactions with the storage array. Each host in the FC network represents a separate initiator, so if a host is connected to the system through two host bus adapters (HBAs), the system identifies two different initiators (similar to multi-homed, Ethernet-based hosts). In contrast, when multipathing is used in round-robin mode, multiple HBAs are grouped together, and the multipathing software identifies the group of HBAs as a single initiator.

internal storage An array physically housed in the Sun StorEdge 6920 system base cabinet, or expansion cabinet, and managed by the system management software. See also external storage.

I/O panel The portion of the service panel that provides Fibre Channel (FC) port connections for data hosts, internal and external storage, gigabit Ethernet ports for remote replication, power connections for up to two expansion cabinets, and Ethernet and FC connections for expansion cabinets. See also service panel and Service Processor panel.

IOPS A measure of transaction speed, representing the number of input and output transactions per second.

LAN See local area network (LAN).

legacy volume An entire LUN on an external storage array that you can use in specific ways as if it were any other local volume, while preserving the user data on that external storage array. You can apply the system's data services to a legacy volume, however, you cannot extend a legacy volume.

local area network (LAN) A communications infrastructure designed to use dedicated wiring over a limited distance (typically a diameter of less than five kilometers) to connect a large number of intercommunicating nodes.

logical unit number (LUN) The SCSI identifier of a logical unit with a target. (SNIA)

In the Sun StorEdge 6920 system, a LUN is the number that is associated with mapping of a volume to an initiator.

LUN See logical unit number (LUN).

LUN mapping The process by which volume permissions (read only, read/write, or none) are assigned to an initiator.

LUN masking The process by which a mapped initiator is filtered.

MAC address See media access control (MAC) address.

management host A host with an in-band and/or out-of-band network connection to the system that is used to manage the system. A management host can have monitoring software installed, such as a remote CLI package, Sun StorEdge Enterprise Storage Manager, Sun Storage Automated Diagnostic Environment (Enterprise Edition), or a third-party monitoring program. A management host can also be used to run management software on another machine through a network connection.

management interface card (MIC) The card on which the management software resides. Each Data Services Platform has two MICs, which operate in a master/alternate master mode.

management path See out-of-band management.

master agent In Sun Storage Automated Diagnostic Environment, Enterprise Edition, the primary health and monitoring agent designated by the user to act as the aggregation point for one or more slave agents. The responsibilities of the master agent include analyzing the events forwarded from the slave agents, generating alarm notification to local and remote recipients, and provisioning the user interface to the slave agents. See also master / alternate master and slave agent.

master / alternate master In Sun Storage Automated Diagnostic Environment, Enterprise Edition, a slave agent designated by the user to act as the temporary master agent if the master agent fails to send a heartbeat signal to the alternate master during a defined time period. The alternate master does not assume all the responsibilities of the master agent; it simply acts as the aggregation point for the remaining slave agents, collecting their events, until the master agent returns to active service. See also master agent and slave agent.

media access control (MAC) address The physical address identifying an Ethernet controller board. The MAC address, also called an Ethernet address, is set at the factory and must be mapped to the IP address of the device.

MIC See management interface card (MIC).

mirror A special type of volume in the Sun StorEdge 6920 system, consisting of up to four separate yet equal mirror components you can access independently (through the mirror) and use to track changes to and update your data set and manage your data migration strategy.

mirror component One of up to four individual copies of the same data set that constitute a mirror in the Sun StorEdge 6920 system. When you perform an action on a mirror component, the resilvering process synchronizes the mirror so that each component is an equal yet separate copy of the same data set. You can perform a variety of operations on a mirror component. See also break, force break, rejoin, reverse rejoin, resilvering, and split.

mirrored volume See mirror and mirror component.

mirroring A form of storage – also called RAID Level 1, independent copy, and real-time copy – whereby two or more independent, identical copies of data are maintained on separate media. Typical mirroring technologies enable the cloning of data sets to provide redundancy for a storage system. The Sun StorEdge 6920 system enables you to create and manipulate up to four equal mirror components for each mirror and to access each component individually in order to track changes to and preserve the integrity of your data. See also mirror and mirror component.

mirror log Area of the storage pool used to track the state of mirror components (such as resilvering progress) relative to the mirror as a whole.

multipathing A design for redundancy that provides at least two physical paths to a target.

Network Storage Command Center (NSCC) A repository and application comprised of a database and a browser-based user interface, designed and maintained by Sun, to collect and analyze health and performance data from supported storage devices.

non-captive storage See external storage.

NSCC See Network Storage Command Center (NSCC).

notification The process performed by the master agent when one or more events require fault reporting to a configured notification recipient such as a local email address, an SNMP port, or a remote service such as the Sun StorEdge Remote Response service. The master agent constructs the event and assembles the information necessary to transmit the event by way of the user-configured transport mechanism. The monitoring and diagnostic software supports notification of one or more notification recipients to satisfy the desired level of notification. Note: The Sun StorEdge 6920 array contains a single agent, which effectively acts as the master agent for the array.

original volume The starting point for a mirrored volume or data migration. In the context of mirroring, see also primary component.

out-of-band management Transmission of management information for Fibre Channel components outside of the Fibre Channel network, typically over an Ethernet network. (SNIA) The Sun StorEdge 6920 system is managed over an out-of-band network by way of an Ethernet connection between the service panel and the local area network (LAN). See also in-band management.

parallel monitoring A monitoring control that allows the agent to examine more than a single device in parallel during an agent polling cycle. The positive effect of increasing the value for this control beyond the default (1) is to speed up the monitoring of devices when a large population exists The negative effect of increasing the value of this control is that the agents will consume more memory and CPU cycles.

parent volume The volume for which a snapshot is taken. See also snapshot.

patch A software or firmware update for a storage device or device component.

PDU See power distribution unit (PDU).

point in time copy A fully usable copy of a defined collection of data that contains an image of the data as it appeared at a single point in time. The copy is considered to have logically occurred at that point in time, but implementations may perform part or all of the copy at other times (e.g., via database log replay or rollback) as long as the result is a consistent copy of the data as it appeared at that point in time. Implementations may restrict point in time copies to be read-only or may permit subsequent writes to the copy. (SNIA)

In the context of data replication, the storage system that houses the primary replication set, which is the remote counterpart of the secondary site. Also known as the local site copy. See also snapshot.

policy A rule or guideline that can result in an automatic response to a system event.

pool See storage pool.

power distribution unit (PDU) The assembly that provides power management for the system. The Sun StorEdge 6920 system provides two PDUs that enable the management software to control the distribution of power to system components for Lights-Out Management (LOM) and field-replaceable unit (FRU) service actions.

primary peer One of a pair of physically separate systems on which the primary replication set resides. The primary peer copies user data to its counterpart, which is the remote, secondary peer.

primary component The first component created to establish a mirrored volume. See also mirror, mirror component, and mirrored volume.

primary volume The volume that contains the original user data that the primary peer replicates to the secondary peer.

profile See storage profile.

provisioning The process of allocation and assignment of storage to hosts.

RAID An acronym for Redundant Array of Independent Disks, a family of techniques for managing multiple disks to deliver desirable cost, data availability, and performance characteristics to host environments. (SNIA)

real-time copy See mirroring.

reconstruction The process of rebuilding of lost data on a replacement disk after a disk failure.

rejoin To move a split mirror component back into the mirror such that, when the resilvering process is complete, the mirror component has been made identical to all other mirror components in the mirror. See also mirror component, reverse rejoin, and split.

remote scripting CLI client A command-line interface (CLI) that enables you to manage the system from a remote management host. The client communicates with the management software through a secure out-of-band interface, HTTPS, and provides the same control and monitoring capability as the browser interface. The client must be installed on a host that has network access to the system.

replication See data replication.

replication bitmap The bitmap that tracks changes to the primary volume. Writes issued to the primary peer are noted in the replication bitmap. The replication set at the secondary peer also includes a replication bitmap that tracks changes if a role reversal assigns the secondary volume the role of primary.

replication link A logical connection associated with a Gigabit Ethernet port that transports data and replication control commands between primary and secondary sites. The Gigabit Ethernet ports at both sites must be enabled for data replication and be configured with the remote site's IP information.

replication peer One of a pair of complimentary components that are on physically separate systems. For example, user data is copied to a remote system, which is the counterpart, or remote peer, of the system on which that user data resides.

replication set A local volume paired with reference to a single remote volume on a remote peer. A replication set works in conjunction with an identically configured replication set on the remote peer to provide an instance of replication. The local volume within a replication set is associated with a replication bitmap and, depending on the set's attributes, with an asynchronous queue.

resilvering The synchronization of mirror components such that, when the process is complete, the mirror comprises equal, independent copies of the same data set. Resilvering occurs when you have performed a rejoin or reverse rejoin operation on a mirror component.

resnap To create a snapshot again and replace the old with the new.

resume operation In the context of data replication, a synchronization operation to establish an identical copy of the primary volume's user data on the secondary volume. The data is synchronized when replication occurs. Synchronization can be initiated by either the user or the system. See also autosynchronization, suspend operation, and synchronization.

reverse rejoin To move a split mirror component back into the mirror such that, when the resilvering process is complete, all mirror components in the mirror have been made identical to the previously split mirror component. See also break, mirror component, rejoin, resilvering, and split.

reverse synchronization See role reversal.

revision analysis In Sun Storage Automated Diagnostic Environment, Enterprise Edition, the process of collecting the current revision information for the software and firmware elements of the system and comparing them against a set of acceptable levels. See also revision maintenance.

revision backout The removal of a patch update on a storage device or device component. See also revision upgrade.

revision maintenance A system process that combines performing revision analysis on the elements of the system and locating, acquiring, and installing the necessary deliverables to bring elements up to an acceptable revision level. See also revision analysis.

revision upgrade The installation of a patch update on a storage device or device component. See also revision backout.

role reversal In the context of data replication, a procedure in which the secondary host is assigned the role of primary host within an established replication set, and the primary volume is updated with the contents of the secondary volume. Role reversal is a failover technique used when the primary site fails and for disaster rehearsal.

rollback The process by which a volume's data is reset to become identical to a snapshot taken of that volume.

SAN See storage area network (SAN).

secondary peer One of a pair of physically separate systems on which the secondary replication set resides. The secondary peer is the recipient of user data from its counterpart, which is the primary peer.

secondary volume The remote counterpart of the primary volume. The secondary volume is the replicated copy of the primary volume. You can map or create a volume snapshot of a secondary volume. You cannot read from or write to a secondary volume unless it is in suspend mode or you change its role to primary.

Service Advisor A diagnostic tool component that provides tools and procedures for performing service on storage device.

service panel A group of input and output connections located at the back of the base cabinet that provides the cabling interface to both control path functions and data path functions. The service panel consists of the Service Processor panel and the I/O panel. See also I/O panel and Service Processor panel.

Service Processor panel The portion of the service panel that provides a modem connection, LAN connections, serial ports, and an AUX port for connection of the Data Services Platform (DSP) management interface card (MIC).

SFC See switch fabric card (SFC).

Simple Mail Transfer Protocol (SMTP) A protocol for sending e-mail messages between servers. Most e-mail systems that send mail over the Internet use SMTP to send messages from one server to another; the messages can then be retrieved with an email client using either Post Office Protocol (POP) or Internet Message Access Protocol (IMAP). In addition, SMTP is generally used to send messages from a mail client to a mail server. This is why you need to specify both the POP or IMAP server and the SMTP server when you configure your email application. (Webopedia)

Simple Network Management Protocol (SNMP) An IETF protocol for monitoring and managing systems and devices in a network. The data being monitored and managed is defined by a Management Information Base (MIB). The functions supported by the protocol are the request and retrieval of data, the setting or writing of data, and traps that signal the occurrence of events. (SNIA)

SIO card See storage I/O (SIO) card.

site LAN The local area network at your site. The system connects to your LAN through the USER LAN port on the service panel. When the system is connected to your LAN, the system can be managed through a browser from any host on the LAN.

slave agent In Sun Storage Automated Diagnostic Environment, Enterprise Edition, a health and monitoring agent that collects health and performance data from the devices it has been designated to monitor and forwards events to the master or alternate master agent for evaluation and notification processing. Slave agents cannot function completely without a master or alternate master agent. See also master agent and master / alternate master.

SMTP See Simple Mail Transfer Protocol (SMTP).

snapshot An instantaneous copy of volume data at a specific point in time. Snapshots are stored in snapshot reserve space on the (parent) volume for which they are taken.

snapshot reserve space Storage space, taken from a pool, where the system stores copies of the parent volume's original data before it is overwritten. See also storage pool.

split To separate a mirror component from the mirror, with the intent to rejoin it later. A split component is counted toward the limit of four mirror components per mirror and continues to be tracked by the system as part of the mirror. See also component, rejoin, reverse rejoin, and snapshot.

SRC See storage resource card (SRC).

SSRR See Sun StorEdge Remote Response service.

storage area network (SAN) An architecture in which the storage elements are connected to each other and to a server that is the access point for all systems that use the SAN to store data.

storage domain A secure container that holds a subset of the system's total storage resources. Multiple storage domains can be created to securely partition the system's total set of storage resources. This enables you to organize multiple departments or applications into a single storage management infrastructure.

storage I/O (SIO) card A card that provides the Fibre Channel (FC) ports for the Data Services Platform (DSP). This card is always paired with a storage resource card (SRC). The Sun StorEdge 6920 system supports two types of SIO cards. The SIO-8 card has eight FC ports and the SIO COMBO card has six FC ports and one Gigabit Ethernet port. See also storage resource card (SRC) set.

storage pool A container that groups physical disk capacity (abstracted as virtual disks in the browser interface) into a logical pool of available storage capacity. A storage pool's characteristics are defined by a storage profile. You can create multiple storage pools to segregate storage capacity for use in various types of applications (for example, high throughput and online transaction-processing applications).

storage port A port on the I/O panel that connects to internal storage. See also host port.

storage profile A defined set of storage performance characteristics such as RAID level, segment size, dedicated hot-spare, and virtualization strategy. You can choose a predefined profile suitable for the application that is using the storage, or you can create a custom profile.

storage resource card (SRC) A card that provides the storage processors for the Data Services Platform (DSP). An SRC is always paired with a storage I/O (SIO) card. See also storage I/O (SIO) card.

storage resource card (SRC) set Two cards in the Data Services Platform (DSP) that together provide the Fibre Channel (FC) and Gigabit Ethernet interfaces: the storage resource card (SRC) and the storage I/O (SIO) card. Two to four SRC sets can be installed in the DSP to provide the ports for connecting data hosts to the system. See also storage I/O (SIO) card.

Storage Service Processor (SSP) A management device integrated into the system that provides unified management access to system components, as well as local and remote management and maintenance. The Storage Service Processor also supports automatic upgrades of patches, firmware, and software.

Storage Service Processor (SSP) accessory tray The portion of the Storage Service Processor that contains a modem, router with a firewall, Ethernet hub, and network terminal concentrator (NTC).

storage tray An enclosure containing disks. A storage tray with a RAID controller is called a controller tray; a storage tray without a controller is called an expansion tray. See also controller tray and expansion tray.

store-and-forward See asynchronous replication

stripe size The number of blocks in a stripe. A striped array’s stripe size is the stripe depth multiplied by the number of member extents. A parity RAID array’s stripe size is the stripe depth multiplied by one less than the number of member extents. (SNIA) See also striping.

striping Short for data striping; also known as RAID Level 0 or RAID-0. A mapping technique in which fixed-size consecutive ranges of virtual disk data addresses are mapped to successive array members in a cyclic pattern. (SNIA) See also concatenation.

SunMC See Sun Management Center (SunMC).

Sun Management Center (SunMC) An element management system for monitoring and managing the Sun environment. Sun Management Center also integrates with the leading enterprise management software, including the Storage Automated Diagnostic Environment, to provide customers with a unified management infrastructure. The base package of Sun Management Center is free and provides hardware monitoring. Advanced applications (add-ons) extend the monitoring capability of the base package.

Sun StorEdge Remote Response service A remote support solution for Sun StorEdge series storage systems that proactively identifies operational anomalies to help prevent them from becoming business problems. Through around-the-clock monitoring, connectivity with Sun, and remote support, the Sun StorEdge Remote Response service helps maximize a storage system's availability.

suspend operation In the context of data replication, an operation in which replication set or consistency group activity is temporarily stopped, and an internal bitmap tracks write operations to the volume rather than sending the write operations over the physical link to the secondary volume. This method tracks write operations that have not been remotely copied while access to the secondary peer is interrupted or impaired. The software uses this replication bitmap to reestablish data replication through an optimized update synchronization rather than through a complete volume-to-volume copy. See also Fast Start operation and resume operation.

switch fabric card (SFC) A board that provides the central switching function for the Data Services Platform (DSP).

synchronization The act of aligning or making entries be equivalent at a specified point in time. (SNIA).

In the context of mirroring, see resilvering.

synchronous replication A replication technique in which data must be committed to storage at both the primary site and the secondary site before a write to the primary volume is acknowledged. See also asynchronous replication.

target The system component that receives a SCSI I/O command. (SNIA)

A target in the Sun StorEdge 6920 system can be an initiator or logical unit number (LUN).

thin-scripting client See remote scripting CLI client.

topology A graphical depiction of a storage network or storage system.

tray See storage tray.

USB flash disk A disk connected to the Storage Service Processor that stores system characteristics for the Storage Service Processor and Data Services Platform (DSP). The disk connects to the USB port on the Storage Service Processor.

virtual disk A set of disk blocks presented to an operating environment as a range of consecutively numbered logical blocks with disk-like storage and I/O semantics. The virtual disk is the disk array object that most closely resembles a physical disk from the operating environment's viewpoint. (SNIA)

In the Sun StorEdge 6920 system, the system itself is the operating environment.

virtualization The act of integrating one or more (back end) services or functions with additional (front end) functionality for the purpose of providing useful abstractions. Typically virtualization hides some of the back end complexity, or adds or integrates new functionality with existing back end services. Examples of virtualization are the aggregation of multiple instances of a service into one virtualized service, or to add security to an otherwise insecure service. Virtualization can be nested or applied to multiple layers of a system. (SNIA)

The Sun StorEdge 6920 system uses virtualization attributes to create and manage storage pools. See also concatenation and striping.

virtualization strategy Selection of the technique used when virtualizing data across multiple virtual disks. See also concatenation and striping.

volume A logically contiguous range of storage blocks allocated from a single pool and presented by a disk array as a logical unit number (LUN). A volume can span the physical devices that constitute the array, or it can be wholly contained within a single physical disk, depending on its virtualization strategy, size, and the internal array configuration. The array controller makes these details transparent to applications running on the attached server system.

World Wide Name (WWN) A unique identifier for a port, initiator, virtual disk, or volume, assigned by the system. The WWN of an object does not change throughout its lifetime and is never reused to name another object.

write order consistency Preservation of write ordering across all volumes in a consistency group or in replication sets.

write ordering The process by which write operations that are directed to the secondary volume occur in the same order as write operations to the primary volume.

WWN See World Wide Name (WWN).

zone A collection of Fibre Channel N_Ports and/or NL_Ports (that is, device ports) that are permitted to communicate with each other via the fabric. (SNIA)

zoning A method of subdividing a storage area network (SAN) into disjoint zones, or subsets of nodes on the network. SAN nodes outside a zone are invisible to nodes within the zone. With switched SANs, traffic within each zone can be physically isolated from traffic outside the zone. (SNIA) See also zone.

Index

F

G

H

I

L

M

N

O

P

Q

T

U

V

W