



SGI® InfiniteStorage NAS System
Storage Subsystem Guide

007-5597-002

COPYRIGHT

© 2010 SGI. All rights reserved; provided portions may be copyright in third parties, as indicated elsewhere herein. No permission is granted to copy, distribute, or create derivative works from the contents of this electronic documentation in any manner, in whole or in part, without the prior written permission of SGI.

LIMITED RIGHTS LEGEND

The software described in this document is “commercial computer software” provided with restricted rights (except as to included open/free source) as specified in the FAR 52.227-19 and/or the DFAR 227.7202, or successive sections. Use beyond license provisions is a violation of worldwide intellectual property laws, treaties and conventions. This document is provided with limited rights as defined in 52.227-14.

The electronic (software) version of this document was developed at private expense; if acquired under an agreement with the USA government or any contractor thereto, it is acquired as “commercial computer software” subject to the provisions of its applicable license agreement, as specified in (a) 48 CFR 12.212 of the FAR; or, if acquired for Department of Defense units, (b) 48 CFR 227-7202 of the DoD FAR Supplement; or sections succeeding thereto. Contractor/manufacturer is SGI, 46600 Landing Parkway, Fremont, CA 94538.

TRADEMARKS AND ATTRIBUTIONS

SGI and the SGI logo are trademarks or registered trademarks of Silicon Graphics International Corp. or its subsidiaries in the United States and other countries.

LSI Logic is a trademark or registered trademark of LSI Logic Corporation. Internet Explorer, Windows, Windows NT, and Windows 2000/2003/2008 are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. Java and Java Virtual Machine are trademarks or registered trademarks of Sun Microsystems, Inc. Linux is a registered trademark of Linus Torvalds, used with permission by SGI. UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company, Ltd.

The following are trademarks licensed to BlueArc Corporation, registered in the USA and other countries: BlueArc, the BlueArc logo, and the BlueArc Storage System.

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (<http://www.openssl.org/>). Some parts of ADC use open source code from NetApp, Inc. and Traakan, Inc.

The product described in this guide may be protected by one or more U.S. patents, foreign patents, or pending applications.

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

Record of Revision

Version	Description
001	October 2009. Original publication.
002	January 2010. Updated publication, replaces 007-5597-001.

Table of Contents

1 About This Guide

- Audience 1
- Chapter Descriptions 1
- Related Publications 2
- Conventions 2
- Browser Support 3
- Product Support 3
- Reader Comments 4

2 The SGI InfiniteStorage NAS Storage System

- System Overview 5
 - System Management Unit 5
 - Storage Server(s) 5
 - Virtual Servers (EVSs) 6
 - Private Management Network 6
 - Public Data Network 7
 - Storage Subsystem 7
 - Storage Subsystem Characteristics 7
 - Hot Spare Disk 7
- Understanding Tiered Storage 8
 - Fibre Channel Fabric 9
 - About FC Paths 9
 - Load Balancing and Failure Recovery 10
 - Fibre Channel Statistics 11
- Supported Storage Subsystems 11
 - The SGI InfiniteStorage 220 Storage Controller Enclosure 12
 - The SGI InfiniteStorage 220 Drive Shelf Expansion 12

The SGI InfiniteStorage 15000 Controller Enclosure 13
The SGI InfiniteStorage 15000 Drive Shelf Expansion 13

3 Configuring SGI InfiniteStorage Arrays

Specifications 15

Storage Characteristics 15

 RAID Controllers 15

 Hot Spare Disk 15

Managing the Storage Subsystem 16

 Discovering and Adding Racks 16

 Partially Discovered RAID Racks 17

 Listing Installed RAID Racks 18

 Performing a Media Scan on a System Drive 22

 Monitoring Active Tasks 22

 Reviewing Events Logged 23

Monitoring Physical Disks 24

4 System Drives and System Drive Groups

System Drives 27

 System Drive Groups 27

 SD Groups and Read Balancing 28

Creating System Drives 30

Managing System Drive Groups 34

 Creating SD Groups Automatically 35

 Creating SD Groups Manually 36

 Backing up or Restoring SD Groups 38

 Modifying System Drive Groups 39

1 About This Guide

This guide provides an overview of the storage subsystems supported for use with the SGI[®] InfiniteStorage NAS Server (IS-NAS Server). This guide also provides information and instructions for managing the storage subsystems attached to your IS-NAS Server.

License keys are required to enable the usage of storage subsystems, and are also used to control the availability of some system features and functionality. For more information on licenses, refer to the *System Administration Guide* or contact SGI Global Services.

Audience

This guide is written for storage administrators of SGI InfiniteStorage NAS Server systems. It is written with the assumption that the reader has a good working knowledge of storage subsystems (RAID arrays), networking concepts and practices, computers, and computer systems.

Chapter Descriptions

The following topics are covered in this document:

- Chapter 1: "[About This Guide](#)"
Provides an introduction to this guide, lists other documentation resources available for this product, and explains the conventions used in this document.
- Chapter 2: "[The SGI InfiniteStorage NAS Storage System](#)"
Provides an overview of the major elements of the NAS storage server system, including the parts of the system, tiered storage, and supported storage subsystems.
- Chapter 3: "[Configuring SGI InfiniteStorage Arrays](#)"
Describes the storage characteristics and management tasks, and monitoring options of the supported storage subsystems.
- Chapter 4: "[System Drives and System Drive Groups](#)"
Provides information about all aspects of system drives, and system drive groups.

Related Publications




The following documents are relevant to the SGI InfiniteStorage NAS Server:

- SGI InfiniteStorage NAS Server *System Administration Guide*: In PDF format, this guide provides information about the features, configuration, administration, and maintenance of an SGI InfiniteStorage NAS Server or cluster.
- SGI InfiniteStorage NAS Server *Software Installation Guide*: In PDF format, this guide provides information about installing software and firmware, including instructions on how to upgrade and downgrade the storage server and the SMU.
- SGI InfiniteStorage NAS Server *Hardware Installation Guide*: In PDF format, this guide provides information about installing the storage server and connecting it to your network.
- SGI InfiniteStorage NAS Server *Hardware Reference*: This guide (in PDF format) provides an overview of the InfiniteStorage NAS Server hardware, describes how to resolve any problems, and shows how to replace faulty components.
- Titan Server *Hardware Reference*: This guide (in PDF format) provides an overview of the Titan Server hardware, describes how to resolve any problems, and shows how to replace faulty components.
- SGI InfiniteStorage NAS Server *Command Line Reference*: This guide (in HTML format) describes how to administer the system by typing commands at a command prompt.
- SGI InfiniteStorage NAS Server *Release Notes*: This document gives late-breaking news about the system.

Conventions

The following conventions are used throughout this document:

Convention	Meaning
Command	This fixed-space font denotes literal items such as commands, files, routines, path names, signals, messages, and programming language structures.
<i>variable</i>	The italic typeface denotes variable entries and words or concepts being defined. Italic typeface is also used for book titles.
user input	This bold fixed-space font denotes literal items that the user enters in interactive sessions. Output is shown in nonbold, fixed-space font.
[and]	Brackets enclose optional portions of a command or directive line.
...	Ellipses indicate that a preceding element can be repeated.

Convention	Meaning
<p>GUI element</p>	<p>This font denotes the names of graphical user interface (GUI) elements such as windows, screens, dialog boxes, menus, toolbars, icons, buttons, boxes, fields, and lists.</p>
	<p>Tip: A tip contains supplementary information that is useful in completing a task.</p>
	<p>Note: A note contains information that helps to install or operate the system effectively.</p>
	<p>Caution: A caution indicates the possibility of damage to data or equipment. Do not proceed beyond a caution message until the requirements are fully understood.</p>

Browser Support

Any of the following browsers can be used to run Web Manager, the System Management Unit (SMU) web-based graphical user interface.

- Microsoft Internet Explorer: version 7.0 or later.
- Mozilla Firefox: version 1.5 or later.

The following Java Runtime Environment is required to enable some advanced Web Manager functionality: Sun Microsystems Java Runtime Environment: version 5.0, update 6, or later.

Some product documentation is included for download or viewing through Web Manager. The following software is required to view this documentation: Adobe Acrobat: version 7.0.5 or later.

Product Support

SGI provides a comprehensive product support and maintenance program for its products. SGI also offers services to implement and integrate Linux applications in your environment.

- Refer to <http://www.sgi.com/support/>.
- If you are in North America, contact the Technical Assistance Center at +1 800 800 4SGI or contact your authorized service provider.
- If you are outside North America, contact the SGI subsidiary or authorized distributor in your country.

Reader Comments

If you have comments about the technical accuracy, content, or organization of this document, contact SGI. Be sure to include the title and document number of the manual with your comments. (Online, the document number is located in the front matter of the manual. In printed manuals, the document number is located at the bottom of each page.)

You can contact SGI in any of the following ways:

Send e-mail to the following address: techpubs@sgi.com.

- Contact your customer service representative and ask that an incident be filed in the SGI incident tracking system.
- Send mail to the following address:

SGI
Technical Publications
46600 Landing Parkway
Fremont, CA 94538

SGI values your comments and will respond to them promptly.

2

The SGI InfiniteStorage NAS Storage System

System Overview

The InfiniteStorage NAS Server and the Titan Server are highly scalable and modular network attached storage (NAS) servers, with multi-gigabit throughput from network to disk. These systems consist of the following elements:

- System Management Unit (SMU)
- InfiniteStorage NAS Servers and/or Titan Servers
- Virtual Servers (EVSs)
- Private Management Network
- Public Data Network
- Storage Subsystem(s)

System Management Unit

The System Management Unit's (SMU's) Web Manager interface provides front-end server administration and monitoring tools. It supports clustering, data migration, and replication, and acts as the Quorum Device in a cluster. Although integral to the system as a whole, the SMU does not move data between the network client and the storage server.

There are two kinds of SMU; external and internal.

- An external SMU can manage up to eight (8) storage servers/clusters in any combination. Each external SMU can manage both IS-NAS Servers/clusters and Titan Servers/clusters.

An external SMU is a separate device in the storage server system. To eliminate the SMU as a single point of failure, you can configure your system with a second external SMU as a standby SMU.

- An internal SMU can manage a single stand-alone IS-NAS Server (an external SMU is required to manage more than a single IS-NAS Server).

An internal SMU is a service that runs on the IS-NAS Server and provides the same management and monitoring functionality as an external SMU. When using an internal SMU, there is no way to configure a standby SMU.

Storage Server(s)

The patented architecture of the IS-NAS Server and the Titan Server is structured around bi-directional data pipelines and a hardware-based file system. It scales to 4 petabytes, supporting higher sustained access loads

without compromising performance. Each storage server can be configured as a single stand alone server or as a node of a cluster. All network clients communicate directly with the storage server.

The server processes file access requests from network clients via Gigabit Ethernet (GE) or 10 Gigabit Ethernet (10GbE) links, reading and writing from/to one or multiple storage devices, connected through Fibre Channel (FC) links. Storage servers can be configured as stand alone servers or as a cluster with multiple nodes, which share the same storage devices, so that network requests can be distributed across cluster nodes.

IS-NAS Servers support clusters with up to two nodes, and current generation Titan Servers support clusters with up to eight nodes. Should one cluster node fail, its file services and server administration functions are transferred to other nodes.



Note: All nodes of a cluster must be of the same series. A cluster cannot be made up of a different models of storage servers.

Both the IS-NAS Server and the Titan Server are rack mountable and contain three hot-swappable fan assemblies, and two hot-swappable redundant power supplies. The front panel of each storage server displays system status with a green power LED and an amber fault LED. The rear panel has additional status LEDs, and includes connectors (power, Ethernet, Fibre Channel, RS-232). See the *Hardware Reference* for your series of storage server for more information about the storage server hardware.

Virtual Servers (EVSs)

All file services are provided by logical server entities referred to as EVSs (virtual servers). A server or cluster supports up to 64 EVSs. Each EVS is assigned unique network settings and storage resources. In clusters, EVSs are automatically migrated between servers when faults occur to ensure maximum availability. When multiple servers or clusters are configured with common storage access, they are referred to as *server farms*. EVSs can be manually migrated between servers in a Server Farm based on performance and availability requirements.

Private Management Network

To minimize the performance impact of auxiliary devices, a private management network connects the SMU and devices such as FC switches, and uninterruptible power supply (UPS) units. The private management network connects the private management interface of the SMU, the Ethernet management interface on the storage server(s), and all of the Ethernet managed devices that make up the storage system.

The private management network is isolated from the public data network by the SMU, which uses network address translation (NAT) technology. Devices on the private network are only accessible from the public data network through the SMU, which provides NAT, NTP, and email relay services.

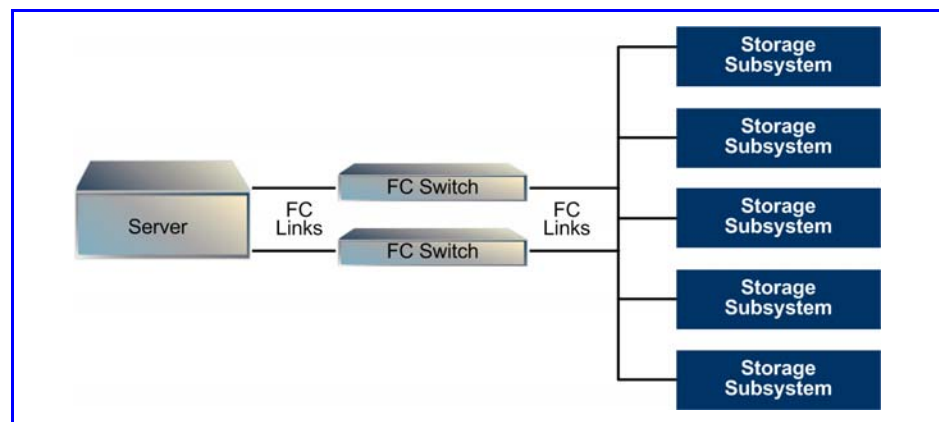
Public Data Network

The public data network, from the storage server perspective, consists of the public Ethernet port on the SMU, and management access can be enabled on individual Gigabit Ethernet (GE) interfaces on the storage server.

Clients connect to the SMU through the public data network, and client connections are made through the public data network. It is also possible to configure storage servers/clusters so that they can be managed using the public data network.

Storage Subsystem

Storage subsystems contain the devices that store the data managed by the storage server. The server allows you to simultaneously connect multiple diverse storage subsystems behind a single server (or cluster), which integrates all physical storage resources into one or more logical file systems.



Each storage subsystem is made up of RAID controllers, storage devices, and the Fibre Channel (FC) infrastructure (such as FC switches and cables) used to connect these devices to a single storage server or cluster.

Storage Subsystem Characteristics

The storage subsystems use hardware RAID controllers, which provide complete RAID functionality and enhanced disk failure management. The number of controllers in the system depends on its storage capacity.

Some storage subsystems use only one type of disk (FC, SATA, or SAS). Some storage subsystems can use different disk technologies, and may even be able to mix disk types within a storage subsystem (but mixing drive types within a storage enclosure is not supported).

Hot Spare Disk

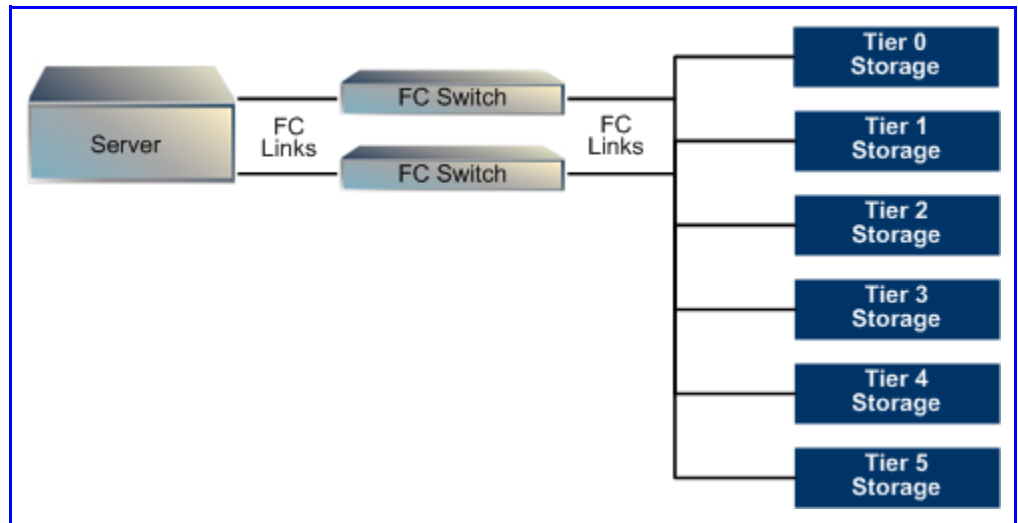
A hot spare disk is a physical disk that is configured for automatic use in the event that another physical disk fails. Should this happen, the system automatically switches to the hot spare disk and rebuilds on it the data that was located on the failed disk. The hot spare disk must have at least as much capacity as the other disks in the system drive. The hot spare disk is a global hot spare, meaning that only one hot spare disk is required per RAID rack, but SGI recommends having several hot spares per rack to maintain a higher margin of safety against hardware failures.

If it is necessary to remove and replace a failed disk, some storage subsystems support “hot swap” operations. In a hot swap, an offline or failed disk is removed and a replacement disk is inserted while the power is on and the system is operating.

Understanding Tiered Storage

Tiered storage allows you to connect multiple diverse storage subsystems behind a single server (or cluster). Using tiered storage, you can match application storage requirements (in terms of performance and scaling) to your storage subsystems. This section describes the concept of tiered storage, and explains how to configure the storage server to work with your storage subsystems to create a tiered storage architecture.

Each server supports up to eight FC ports, independently configurable for either 1-, 2- or 4-gigabit operation. Independent configuration allows you to connect to a range of storage subsystems, which allows you to choose the configuration that will best meet application requirements. The server manages all back-end storage as a single system, through an integrated network management interface.



Based on a storage subsystem’s performance characteristics, it is classified as belonging to a certain tier, and each tier is used differently in the enterprise storage architecture. The currently supported storage subsystems are fit into the tiered storage model as follows:

Tier	Performance	Disk Type	Disk RPM
1	Very high	Dual-ported FC	15,000
		SAS	15,000
2	High	Dual-ported FC	10,000
3	Nearline	SATA or SAS	7,200
		PATA	7,200

Tier	Performance	Disk Type	Disk RPM
4	Archival	SATA or SAS	7,200
5	Long-term storage	N/A (Tape)	NA

The storage server supports tiers of storage, where each tier is made up of devices with different performance characteristics or technologies.

Tiers of storage and storage virtualization are fully supported by Data Migrator, an optional feature which allows you to optimize the usage of tiered storage and remote NFSv3 servers by automatically migrating data among storage subsystems of primary and secondary storage. Based on user-defined policies, Data Migrator monitors file metadata such as size, type, duration of inactivity, access history, and so on. When the criteria of a policy are met, Data Migrator migrates files according to rules specified in the policy as background tasks with minimal impact on server performance. From the perspective of the client workstation, primary versus secondary file location is transparent. Note that Data Migrator does not support migrating data to or from tape library systems. For detailed information about Data Migrator, refer to the *System Administration Guide*.

Fibre Channel Fabric

The server supports fabric FC switches, and when connecting to a FC switch, the server must be configured for N_Port operation. Several FC Switch options are available, contact SGI Global Services for more information on supported switches.

You can manage the FC interface on the server/cluster through the command line interface (CLI), using the following commands:

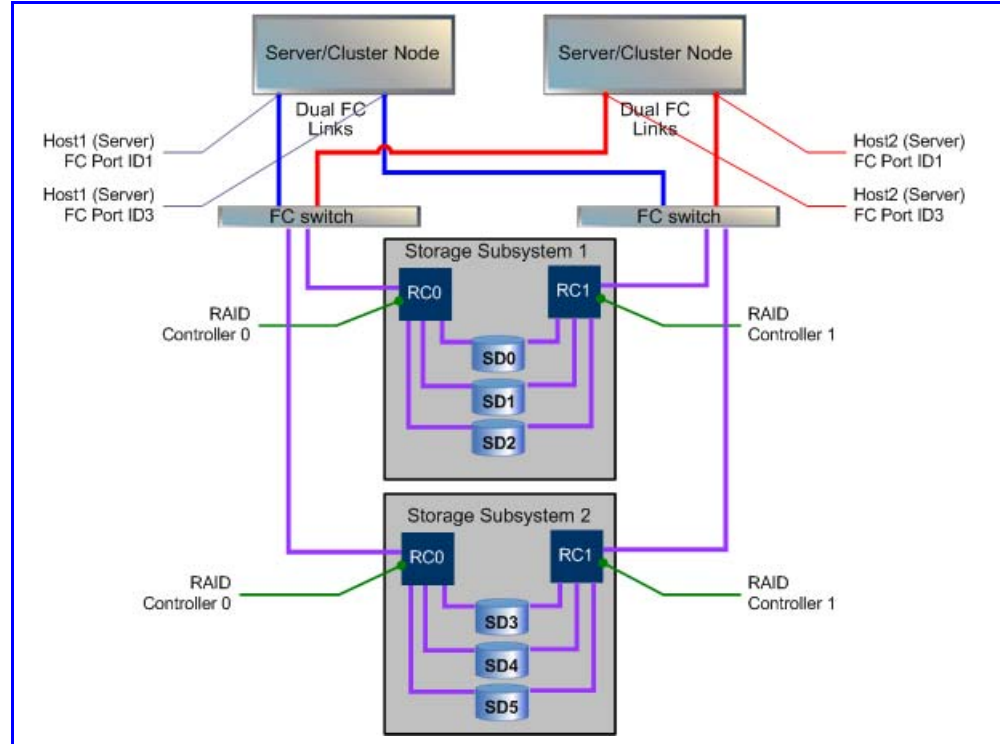
- `fc-link` to enable/disable the FC link.
- `fc-link-type` to change the FC link type.
- `fc-link-speed` to change the FC interface speed.

For more information about these commands, refer to the *Command Line Reference*.

About FC Paths

Each storage subsystem can be accessed through either of two available FC paths. An FC path is made up of the server's host port ID, the storage subsystem port WWN (worldwide name), and the SD identifier (ID). The

following illustration shows a complete path from the server to each of the SDs on the storage subsystem:



You can view information about the FC paths on the server/cluster through the command line interface (CLI), using the `fc-host-port-load`, `fc-target-port-load`, and the `sdpath` commands.

Load Balancing and Failure Recovery

Load balancing on a storage server is a matter of balancing the loads to the system drives (SDs) on the storage subsystems (RAID arrays) to which the storage server is connected. SDs are a logical division of a group of the physical disks of the storage subsystem, and the SD is the basic storage unit of the storage subsystem.

The server routes FC traffic to individual SDs over a single FC path, distributing the load across two FC switches and, when possible, across dual active/active or multi-port RAID controllers.

Following the failure of a preferred path, disk I/O is redistributed among other (non-preferred) paths. When the server detects reactivation of the preferred FC path, it once again redistributes disk I/O to use the preferred FC path.

Default load balancing (load balancing automatically performed by the storage server) is performed based on the following criteria:

- “Load” is defined as the number of open SDs, regardless of the level of I/O on each SD. SDs count towards load at the target if they are open to at least one cluster node; the number of nodes (normally all nodes in a cluster, after boot) is not considered.

- Balancing load on RAID controller target ports takes precedence over balancing load on server FC host ports.
- Balancing load among an subsystem's RAID controllers takes precedence over balancing among ports on those controllers.
- In a cluster, choice of RAID controller target port is coordinated between cluster nodes, so that I/O requests for a given SD do not simultaneously go to multiple target ports on the same RAID controller.

You can manually configure load distribution from the CLI (overriding the default load balancing performed by the server), using the `sdpath` command. When manually configuring load balancing using the using the `sdpath` command:

- You can configure a preferred server host port and/or a RAID controller target port for an SD. If both are set, the RAID controller target port preference takes precedence over the server host port preference. When a specified port preference cannot be satisfied, port selection falls back to automatic selection.
- For the SDs visible on the same target port of a RAID controller, you should either set a preferred RAID controller target port for **all SDs** or for **none of the SDs**. Setting the preferred RAID controller target port for only some of the SDs visible on any given RAID controller target port may create a situation where load distribution is suboptimal.

The `sdpath` command can also be used to query the current FC path being used to communicate with each SD. For more information on the `sdpath` command, run the `man sdpath` command.

Fibre Channel Statistics

The server provides per-port and overall statistics, in real time, at 10-second intervals. Historical statistics cover the period since previous server start or statistics reset. The **Fibre Channel Statistics** page of the Web Manager displays a histogram showing the number of bytes/second received and transmitted during the past few minutes.

Supported Storage Subsystems

The storage server supports the following storage enclosures:

- Current offerings:
 - SGI InfiniteStorage 220
 - SGI InfiniteStorage 15000

Due to the specific capacity and performance characteristics of each storage subsystem, it will typically be used in the tiered storage model as follows:

Enclosure	Typically used in Tier(s)
SGI InfiniteStorage 220	Tier 1, Tier 2, and Tier 3 This enclosure has several configurations, and is suitable for use in several tiers, based on configuration of the individual storage array.
SGI InfiniteStorage 15000	Tier 1, Tier 2, and Tier 3 This enclosure has several configurations, and is suitable for use in several tiers, based on configuration of the individual storage array.

The following table describes the RAID levels supported by each of the currently supported storage subsystems:

Enclosure	RAID Level(s) Supported
SGI InfiniteStorage 220	0/1/ 1+0/5/ 6
SGI InfiniteStorage 15000	0/1/ 1+0/5/ 6

The SGI InfiniteStorage 220 Storage Controller Enclosure

A storage enclosure containing up to 12 SAS (Serial Attached SCSI) or Serial ATA (SATA) disk drives. The InfiniteStorage 220 is available with 15,000 RPM SAS disk drives, as well as 7,200 RPM SATA disk drives. In the subsystem, an InfiniteStorage 220 storage enclosure serves as a storage controller enclosure, housing up to 12 SAS or SATA disks and dual RAID controllers. Each RAID controller has dual RAID host ports, and 2 FC drive loops, and 2 Ethernet ports. The Ethernet ports are used to connect to the environmental sensor modules in the drive shelf enclosures.

The RAID controllers in an InfiniteStorage 220 storage enclosure support a maximum of 4 enclosures, including the InfiniteStorage 220 storage controller enclosure itself, and up to 3 drive shelf enclosures.

The SGI InfiniteStorage 220 Drive Shelf Expansion

A drive shelf enclosure, which can be connected to the InfiniteStorage 220 storage enclosure. Each drive shelf expansion consists of an enclosure fitted with 12 SAS or SATA disks, and dual environmental sensor modules.

Each drive shelf enclosure must contain all SAS or all SATA disk drives; you cannot mix SAS and SATA disks within a drive shelf enclosure.

Each environmental sensor module has two interfaces, which are used to Loop In and Loop Out of the storage controller enclosure. Wiring is routed so that one path leads to the first drive shelf enclosure and the other path leads to the last drive shelf enclosure.

Drive shelf enclosures containing FC disks and drive shelf enclosures containing SATA disks may be connected to the same InfiniteStorage 220 storage controller enclosure. The following configurations are supported:

- All SAS configuration: InfiniteStorage 220 and up to 3 drive shelf expansion enclosures with SAS drives.
- All SATA configuration: InfiniteStorage 220 and up to 3 drive shelf expansion enclosures with SATA drives.
- Mixed SAS and SATA configuration: Up to 3 drive shelf enclosures containing SAS or SATA disks may be connected to the InfiniteStorage 220 storage controller enclosure, as long as each drive shelf enclosure contains only SAS or only SATA drives. You cannot mix drive types within an enclosure.

The SGI InfiniteStorage 15000 Controller Enclosure

A storage controller enclosure, the InfiniteStorage 15000 contains dual RAID controllers, but no disk drives. Each RAID controller has 4 RAID host ports, and 10 back end channels.

The RAID controllers in an InfiniteStorage 15000 storage enclosure support a maximum of 20 enclosures, including up to 20 drive shelf enclosures.

The SGI InfiniteStorage 15000 Drive Shelf Expansion

A drive shelf enclosure, which can be connected to the InfiniteStorage 15000 storage enclosure. Each drive shelf expansion consists of an enclosure fitted with up to 60 SAS or SATA disks.

Each drive shelf expansion enclosure may contain all SAS or all SATA disk drives, and you can mix SAS and SATA disks within a drive shelf enclosure as long as all disks in a RAID group (LUN) are of the same type.

Drive shelf enclosures containing SAS disks and drive shelf enclosures containing SATA disks may be connected to the same InfiniteStorage 15000 storage controller enclosure.

The following configurations are supported for the InfiniteStorage 15000 storage enclosure:

- All SAS configuration: InfiniteStorage 15000 and up to 20 drive shelf expansion enclosures with all SAS drives.
- All SATA configuration: InfiniteStorage 15000 and up to 20 drive shelf expansion enclosures with all SATA drives.
- Mixed SAS and SATA configuration: Up to 20 drive shelf expansion enclosures in any combination. A drive shelf expansion enclosure may be made up of all SAS drives, all SATA drives, or a mix of SAS and SATA drives.



A RAID group (LUN) may not contain a mixture of drive types.

3 Configuring SGI InfiniteStorage Arrays

Specifications

The specifications for each supported RAID controller enclosure or drive shelf expansion may be found as described below:

Enclosure	Enclosure Type	Page
InfiniteStorage 220	RAID Controller	12
InfiniteStorage 220	Drive Shelf Expansion	12
InfiniteStorage 15000	RAID Controller	13
InfiniteStorage 15000	Drive Shelf Expansion	13

Storage Characteristics

The SGI InfiniteStorage family of storage controller enclosures use hardware RAID controllers, which provide complete RAID functionality and enhanced disk failure management.

The drive shelf expansion enclosures attach to the RAID controllers in order to expand the amount of storage available.

RAID Controllers

The RAID controllers operate as an Active/Active (A/A) pair within the same rack. Both RAID controllers can actively process disk I/O requests. Should one of the two RAID controllers fail, the storage server reroutes the I/O transparently to the other controller, which starts processing disk I/O requests for both controllers.

Hot Spare Disk

When the failed disk is replaced, the RAID controller's CopyBack process will automatically move the reconstructed data from the disk that was the hot spare to the replacement disk. The hot spare disk will then be made available for future use.

If it is necessary to remove and replace failed disks, it is possible to perform "hot swap" operations. In a hot swap, an offline or failed disk is removed and



a replacement disk is inserted while the power is on and the system is operating.

Note: When replacing disk drives, allow at least a one-minute interval between disk removal and disk replacement. This interval allows the RAID controller to recognize the change in the RAID rack’s configuration.

Managing the Storage Subsystem

InfiniteStorage RAID racks can be managed using Web Manager. Common operations are:

- Changing the rack name, password, or media scan period.
- Checking the status of media scan and other operations.
- Reviewing events logged by the RAID rack.
- Determining the status of physical disks.

Discovering and Adding Racks

Before creating system drives, RAID racks must be discovered and added to the list of RAID racks for the currently selected managed server.

To Add a RAID rack:

1. Navigate to the **Storage Management** page (**Home > Storage Management > RAID Racks**).
2. In the **RAID Racks** page, click **Discover Racks**.

Storage Management | [Home](#) > [Storage Management](#) > RAID Racks

RAID Racks

Name	Controller 0 (A) Status	Controller 1 (B) Status	Firmware	Rack Status	
<input type="checkbox"/> 2822	Online	Online	05.42.10.00 / 05.42.10.00	● OK	details
<input type="checkbox"/> 2882	Online	Online	06.10.05.90 / 06.10.05.90	● OK	details
<input type="checkbox"/> Rack 0	Online	Online	8.50.00 / 8.50.00	● OK	details

[Check All](#) | [Clear All](#)

Actions: [delete](#) [Discover Racks](#) | [View Physical Disks](#) [View Active Tasks](#)

Shortcuts: [System Drives](#)

3. Select a rack.

In the Choose Racks to Monitor area, check the checkboxes in the Rack Name column to be added to the currently managed server. Use the Rack Password field for racks with configured passwords. Then, click **OK**.

Choose racks to monitor		
Rack Name	IP Address(es)	Rack Password
<input checked="" type="checkbox"/> 2882	192.168.38.33 / 192.168.38.34	<input type="text"/>
<input checked="" type="checkbox"/> 2822	192.168.38.35 / 192.168.38.36	<input type="text"/>

If no racks appear, the SMU was unable to find any supported racks on its network. In this case, verify that the RAID racks have their network settings properly configured.



Note: RAID racks that have already been added to the currently managed server will not be present in the list of discovered RAID racks.

Once a rack is added, the following events occur:

- The selected RAID racks appear on the **RAID Racks** list page and on the System Monitor (for the currently selected managed server).
- The SMU begins logging rack events, which can be viewed through the Event Log link on the **RAID Rack Details** page.
- RAID rack severe events will be forwarded to each managed server that has discovered the rack and included in its event log. This triggers the server's alert mechanism, possibly resulting in alert emails and SNMP traps.
- RAID rack time is synchronized daily with SMU time.
- If system drives are present on the RAID rack, the rack "cache block size" will be set to 16 KB.

Partially Discovered RAID Racks

When discovering RAID racks, it is possible that only one of the controllers' IP addresses might be discovered (for instance, if only one controller is online). In this instance, the RAID rack is considered only "partially discovered." The RAID rack can still be added; however, it will appear on the **RAID Rack** list page with an amber status and will have reduced functionality, because the SMU can communicate with only one controller. If a rack is partially discovered, you should identify and remedy the problem as soon as possible.

When both controllers are back online, the RAID rack should be removed and rediscovered. Rediscovery will allow each controller's IP address to be fully discovered. Having both IP addresses allows the SMU to maintain contact with the RAID rack even if one of the controllers fails.



Note: Deleting a RAID rack only removes it as a device managed by the SMU. It will not affect any configured system drives or file systems.

Listing Installed RAID Racks

To list installed RAID racks:

1. Navigate to the RAID Racks page (Storage Management > RAID Racks).

RAID Racks

Name	Controller 0 (A) Status	Controller 1 (B) Status	Firmware	Rack Status	
<input type="checkbox"/> Rack 0	Online	Online	05.40.04.00 / 05.40.04.00	● OK	details
<input type="checkbox"/> Rack 1	Online	Online	8.60.00 / 8.60.00	● OK	details
<input type="checkbox"/> Rack 2	Online	Online	8.60.00 / 8.60.00	● OK	details

[Check All](#) | [Clear All](#)

Actions: [delete](#) | [Discover Racks](#) | [View Physical Disks](#) | [View Active Tasks](#)

Shortcuts: [System Drives](#)

The following actions are available and apply to racks selected by filling the checkbox for the RAID rack:

- **Delete** removes the RAID rack from the list. Deleting the rack just removes the rack as a managed rack, it does not affect the system drives configured on the storage enclosures in the rack.
- **Discover Racks** allows the server to check for additional RAID racks. It searches for supported storage devices connected to both the public and private management networks. Once a RAID rack has been found, then it can be managed.
- **View Physical Disks** shows the status of the physical disks associated with a RAID rack.
- **View Active Tasks** shows the status of operations, such as media scans, which are in progress for a RAID rack.
- The shortcut for **System Drives** brings up the **System Drives** page in which a system drive can be managed.

In any individual row, clicking the **details** button brings up a **RAID Rack Details** page. This page provides information on the RAID rack.

Storage Management | [Home](#) > [Storage Management](#) > [RAID Racks](#) > RAID Rack Details

RAID Rack Details

Identification

Name:

Rack WWN: 60:0a:0b:80:00:16:48:f3:00:00:00:43:2e:f5:20

Rack Status Summary

Overall Status: ● Severe [View Issues](#)

Configuration

Media Scan Period: days

Controllers

Controller A (192.168.38.31): Status: ● Online Mode: Active Firmware: 06.12.18.00	Controller B (192.168.38.32): Status: ● Online Mode: Active Firmware: 06.12.18.00
--	--

Physical Disks

Overall Physical Disk Status: ● Severe [Severe](#)

Batteries

Battery 0 Status: ● OK Location: Tray 1 Age: 68 Day(s) Life Remaining: 651 Day(s)	Battery 1 Status: ● OK Location: Tray 1 Age: 68 Day(s) Life Remaining: 651 Day(s)
---	---

Power Supplies

PSU 0 Status: ● OK Location: Tray 1	PSU 1 Status: ● OK Location: Tray 1
---	---

ESM Modules

Temperature Sensors

Temperature Sensor 0
Status: ● OK
Location: Tray 1

Fans

Fan 0 Status: ● OK Location: Tray 1	Fan 1 Status: ● OK Location: Tray 1
---	---

Actions: [Change Password](#)

Shortcuts: [Physical Disks](#) [Active Tasks](#) [Event Log](#) [System Drives](#)

The following table describes fields and columns in this page:

Area	Field	Description
Identification		<p>This area displays identification information, including:</p> <ul style="list-style-type: none">• Name. The name of the RAID rack. You can change this name to make it easier to identify the RAID rack.• Rack WWN. Worldwide name for the RAID rack, set by the manufacturer.
Rack Status Summary		<p>The indicator light provides status information about the RAID rack:</p> <ul style="list-style-type: none">• Green – OK• Amber – Warning• Red – Severe <p>When you click the View Issues link, the Rack Issues page is displayed.</p>
Configuration	Media Scan Period	<p>The number of days over which a complete scan of the system drives will occur.</p>
Controllers		<p>The name, IP address (if available), and status information for each RAID controller:</p> <ul style="list-style-type: none">• Status. The status of the RAID controller.• Mode. The mode reports the RAID controller as running in either active or passive mode. By default, both controllers are active.• Firmware. The firmware version installed on the RAID controller.
Physical Disks		<p>The summary of the overall status of the physical disks within the RAID rack:</p> <ul style="list-style-type: none">• Green – OK• Amber – Warning• Red – Severe <p>Click the indicator light or the status to display the physical disks on the RAID Rack page.</p>

Area	Field	Description
Batteries		<p>Batteries maintain data held in the RAID controller's cache, preserving data in the event of a power failure.</p> <p>The information for each battery within the RAID rack:</p> <ul style="list-style-type: none"> • Status. The status of the batteries: <ul style="list-style-type: none"> • Green – OK • Amber – Warning • Red – Severe • Location. The location of the batteries within the RAID rack. • Age. The number of days that the batteries have been in the RAID rack. • Life Remaining. The number of days until the batteries should be replaced.
Power Supplies		<p>The status of each power supply unit (PSU) in the RAID rack.</p> <ul style="list-style-type: none"> • Status. The status of the PSU: <ul style="list-style-type: none"> • Green – OK • Amber – Warning • Red – Severe • Location. The location of the PSU in the RAID rack.
ESM Temperature Sensors		<p>The status and location of the environmental sensor module in the RAID rack.</p> <ul style="list-style-type: none"> • Status. The status of the sensor: <ul style="list-style-type: none"> • Green – OK • Amber – Warning • Red – Severe • Location. The location of the sensor in the RAID rack.
Fans		<p>The status of the fans within the RAID rack.</p> <ul style="list-style-type: none"> • Status. The status of the fan: <ul style="list-style-type: none"> • Green – OK • Amber – Warning • Red – Severe • Location. The location of the fan in the RAID rack.

The following actions are available:

- Click **Apply** to save any changes to the RAID rack identification or configuration.
- Click **Change Password** to display the **RAID Rack Password** page.

The following shortcuts are available:

- **Physical Disk Status.** Click the **Physical Disk Status** link to display the **Physical Disks on RAID Rack** page for the RAID rack.
- **Active Tasks.** Click the **Active Tasks** link to display the **Active Tasks** page for the RAID rack.
- **Event Log.** Click the **Event Log** link to display the **Rack Event Log** page.
- **System Drives.** Click the **System Drives** link to display the **System Drives** page.

Performing a Media Scan on a System Drive

Media Scan provides early warning of system drive failure and reduces the possibility of encountering a media error during host operations. A typical file system comprises frequently used files, rarely used files, and free space. When a disk develops bad spots in seldom-accessed sectors, failures in those locations can remain undetected. Media Scan can detect such drive media errors before they are found during a normal read or write to the system drive.

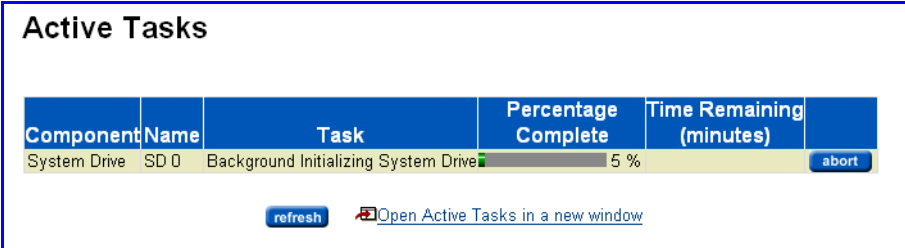
Media Scan is performed as a background task, scanning and comparing all data and parity information on the configured system drives. Media Scan runs on all system drives that are optimal (that is, operating without known failures) and have no modification operations in progress. Errors detected during a scan are reported in the event log.

Monitoring Active Tasks

You can view the status of on-going activity (Media Scan, CopyBack, initializing, and so on) within the RAID rack on the **Active Tasks** page.

1. **Navigate to the RAID Racks page (Home > Storage Management > RAID Racks).**
2. Select a RAID rack.

In the **RAID Racks** page, check the checkbox next to the RAID rack on which to view the Active Tasks, then click **View Active Tasks**.



Component Name	Task	Percentage Complete	Time Remaining (minutes)
System Drive SD 0	Background Initializing System Drive	5 %	

[refresh](#) [Open Active Tasks in a new window](#)

All ongoing activity on the RAID rack are displayed on this page. This page will automatically refresh every 60 seconds.

The following choices are available:

Refresh updates the status of Active Tasks.

Back brings up the **RAID Racks** list page.



Reviewing Events Logged

Note: Some tasks will report a Percentage Complete but not a Time Remaining. In this case, Time Remaining will be shown as Not Known.

The SMU monitors events, such as failure conditions, logged by the RAID rack. The SMU is connected to the storage subsystem through an “out-of-band” Ethernet link (the private management network is typically used for this purpose). Severe events are immediately forwarded to the server’s event log. This will trigger alert notifications (for example, email or SNMP traps) if the server is configured to do so. In addition, if the SMU is unable to connect to the RAID racks, a severe alert will be triggered on the server.

To review events logged by the RAID rack:

1. **Navigate to the RAID Rack Details page (Home > Storage Management > RAID Racks).**

Click the **details** button corresponding to the RAID rack on which to view the RAID Rack Details page.

2. Display the event log.

From the RAID Rack Details page, click **Event Log**.

Storage Management | Home > Storage Management > RAID Racks > Rack Event Log

Rack Event Log

Details		Filter	
Rack Name: 5884		Severity: All <input type="button" value="filter"/>	
Current Controller A Time: 10/10/06 02:36:06PM			
Current Controller B Time: 10/10/06 02:36:06PM			
Events 21-40 of 994 : Page: 1 2 3 4 >> >> >>			
Severity	Date/Time	Component - Location	ID - Message
Info	10/08/06 12:00:32AM	Controller Firmware - None	5023 - Controller return status/function call for ...
Info	10/08/06 12:00:32AM	Controller - Controller in Slot A	5017 - Synchronize controller clock
Info	10/08/06 12:00:32AM	Controller Firmware - None	5023 - Controller return status/function call for ...
Info	10/08/06 12:00:32AM	Controller - Controller in Slot B	5017 - Synchronize controller clock
Info	10/06/06 01:29:40PM	Volume - Volume sd1	2024 - Media scan (scrub) resumed
Info	10/06/06 01:29:38PM	Controller Firmware - None	5023 - Controller return status/function call for ...
Info	10/06/06 01:29:38PM	Channel - Host-side: controller in Slot A port 0	5031 - Deactivate remote volume mirroring
Info	10/06/06 01:29:38PM	Volume - Volume CTL 0 Mirror Repository	2203 - Volume group or volume deleted
Info	10/06/06 01:29:38PM	Volume - Volume CTL 1 Mirror Repository	2203 - Volume group or volume deleted
Info	10/06/06 01:29:37PM	Volume - Volume sd1	2023 - Media scan (scrub) completed
Info	10/06/06 01:29:37PM	Volume - None	6301 - Mirror repository volume deleted
Info	10/06/06 11:24:19AM	Volume - Volume sd1	2022 - Media scan (scrub) started
Info	10/06/06 11:22:32AM	Volume - Volume CTL 1 Mirror Repository	221f - Immediate availability initialization(IAF) ...
Info	10/06/06 11:22:28AM	Volume - Volume CTL 1 Mirror Repository	2026 - Reconstruction completed
Info	10/06/06 11:22:28AM	Volume - Volume CTL 1 Mirror Repository	225a - Immediate availability initialization(IAF) ...
Info	10/06/06 11:22:22AM	Volume - Volume CTL 0 Mirror Repository	221f - Immediate availability initialization(IAF) ...
Info	10/06/06 11:22:20AM	Controller Firmware - None	5023 - Controller return status/function call for ...
Info	10/06/06 11:22:20AM	Channel - Host-side: controller in Slot A port 0	5030 - Activate remote volume mirroring
Info	10/06/06 11:22:19AM	Volume - Volume CTL 0 Mirror Repository	2026 - Reconstruction completed
Info	10/06/06 11:22:19AM	Volume - Volume CTL 0 Mirror Repository	2025 - Reconstruction started
Events 21-40 of 994 : Page: 1 2 3 4 >> >> >>			

Actions:

[Home](#) | [About](#) | [Sign Out](#) | [BlueArc Web Site](#)

The event log is updated every three minutes or when a severe event occurs on the RAID rack. It displays a maximum of 1,000 events (info or

severe). Up to 3,000 events are archived on the SMU and are available for download using the **download** button.

This table can also be filtered to view the event log based on severity level: All, Info, or Severe.

Item/Field	Description
Severity	The level of severity is displayed for each event. Severity might be either Info (informational), or severe (error).
Date/Time	The date and time at which the event was logged.
Component - Location	The component (controller, battery, volume, channel, and so forth) and location of the event, if known.
ID/Message	The event ID and the details about the event.

The Details section provides the rack name and the current (RAID) controllers date and time.

The following actions are available:

- **Refresh** updates the **Event Log** page. The **Event Log** page automatically refreshes every 60 seconds.
- **Download** allows the archived events to be downloaded in a comma-separated values (.csv) provided in a ZIP file. Even though the SMU displays only the most recent 1,000 events, many more are archived on the SMU hard drive. Approximately 2 MB (about 4,000) of the most recent events are archived.
- **Clear All** permanently deletes all the events in the SMU event log and the RAID rack log.

Monitoring Physical Disks

The status of the physical disks associated within a RAID rack can be determined by using Web Manager. Also, the status of the physical disks can be changed if a physical disk needs to be removed or a new hot spare disk has been added.

To check and change the status of physical disks:

1. Navigate to the **RAID Racks** page (**Home > Storage Management > RAID Racks**).
2. Display status for selected physical disks.

Check the checkbox for the RAID rack with physical disks of interest and click **View Physical Disks**.

[Storage Management](#) | [Home](#) > [Storage Management](#) > [RAID Racks](#) > Physical Disks on RAID Rack 2882dt

Physical Disks on RAID Rack 28d

Enclosure	Slot	Capacity	Type	Storage Pool(s)	Status	Hot Spare	Available	Off Line	Manufacturer	Firmware
<input type="checkbox"/>	1	136.73 GB	FIBRE		OK	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	2	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	3	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	4	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	5	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	6	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	7	136.73 GB	FIBRE		OK	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	8	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	9	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	10	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	11	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	12	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	13	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11
<input type="checkbox"/>	14	136.73 GB	FIBRE Test		OK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SEAGATE	MS11

Actions: [Assign as Hot Spare\(s\)](#) [Unassign as Hot Spare\(s\)](#)

[Home](#) | [About](#) | [Sign Out](#) | [BlueArc Web Site](#)

The following table describes the columns in this page:

Item/Field	Description
Enclosure	The ID of the enclosure housing the disk.
Slot	The slot number in the storage enclosure in which the physical disk resides.
Capacity	The storage capacity of the disk.
Type	The type of physical disk in the enclosure, typically either Fibre Channel or SATA.
Storage Pool	The label of the Storage Pool, if the physical disk is in use within a Storage Pool.
Status	The current status of the physical disks within the RAID rack.
Hot Spare	The box is checked if it is assigned as a hot spare.
Available	The box is checked if the physical disk is available.
Offline	The box is checked if the physical disk is offline.
Manufacturer	The name of the disk manufacturer.
Firmware	The firmware version on the physical disk.



Within the **Physical Disk** page, physical disks which are shown as available can be assigned or unassigned as hot spares.

Note: SGI requires that at least one disk be marked as a hot spare by the time the first system drive is created.

4

System Drives and System Drive Groups

System Drives

Logically, system drives (SDs) are the basic storage element used by the storage server. The server assigns each system drive a unique identifying number (ID) and, once assigned, the SD is referenced by that ID number, and the ID number may not be changed.

Physically, each SD is made up of several physical disks. The size of the system drive depends on factors such as the RAID level, the number of disks, and their capacity. See [Supported Storage Subsystems](#), on page 11 for information on the RAID level(s) supported by your storage subsystem(s).

On all IS-NAS Server/clusters, and all Series 3000 Titan Servers/clusters, system drives can be organized into **system drive groups**, which can improve the performance of a storage server or cluster by optimizing reads and writes.

System Drive Groups

With many storage subsystems, system drives (SDs) are limited to 2TB each. However, with today's large physical disks, RAID arrays must be considerably larger than 2TB in order to make efficient use of space. So it is common for system administrators to build large RAID arrays (often called "RAID groups" or "volume groups") and then divide them into system drives (SDs) of 2TB or less. Note that each SD in a RAID group typically uses some space on each disk in the RAID group. SDs are the units of storage that the server sees and manages (the server organizes SDs into Storage Pools, which then contain the file systems).

When performing write operations, if the server were to write simultaneously to multiple SDs in the same RAID group, it would increase head-movement, reducing both performance and the expected life of the disks. So the server has a mechanism to allow it to write to only one SD (LUN) in a RAID group at any one time. By defining SD groups, you tell the server which SDs are in each RAID group and give it the information it needs to optimize write performance.



Note: For Titan Server/clusters, system drive groups are supported on Series 2000 and Series 3000 storage servers, but only Series 3000 Titan Servers/clusters take advantage of this functionality. All IS-NAS Servers/clusters take advantage of system drive groups.

A system drive that is not in any group is treated as if it were in a group of its own.

The SMU cannot group or ungroup system drives that are used in open Storage Pools. A Storage Pool is open if it has any file system that is mounted or is being checked or fixed anywhere on the cluster.

During EVS migration, the SMU automatically copies the groups from the source storage server or cluster and adds them to the target storage server or cluster.

See [Managing System Drive Groups](#), on page 34 for information on creating and modifying system drive groups.

SD Groups and Read Balancing

When performing read operations, if the server can read file system data that has been distributed across the SD groups, the read operations are more efficient and disk head movement may be decreased. These benefits occur because the data is distributed across different physical disks in the SDs making up the SD group instead of being on the same physical disks making up the SDs of the SD group.

Typically, file system data is distributed across the SDs in a Storage Pool during normal write operations due to dynamic write balancing. However, after adding SDs to an SD group or extending a Storage Pool by adding SDs that reside on a new/different physical storage subsystem, you may want to manually initiate a data redistribution so that the file system's data is evenly spread among all SDs in the SD group.

During such a data redistribution, the server's file serving performance may be impacted, because the server is reading and writing data from the data distribution as well as data from external requests.

The file system data redistribution utility is controlled using the `fs-read-balancer` command, and a separate command, `fs-sdg-utilization`, is used to see a report on how a file system is utilizing each SD group in the underlying Storage Pool. For more information about these commands, refer to the *Command Line Reference*.



Note: After a file system has been expanded to use new storage, you can spread the file system's data into the new storage by running the file system data redistribution utility. Spreading the data into the new storage spreads the read load across all of the file system's physical disks, which improves read performance.

Read Balancing Utility Considerations

Running the file system data redistribution utility causes data to be re-written to a new location, which will be the least utilized SD groups (the new storage) resulting in more balanced utilization of SD groups.



Note: The file system data redistribution utility can be run only after expanding the file system into the new storage, but it should be run immediately and it may be run only once. If you run the data redistribution utility more than once, or after an application has written a significant amount of data into the recently added storage, the utility will either refuse to run or produce unpredictable results.

The file system data redistribution utility is designed to operate when a file system is expanded into new storage after SDs have been added to a Storage Pool when the file system is nearly full. However, storage may also be added to a Storage Pool for other reasons:

- To increase performance.
- To prevent the file system from becoming 100% full.

To achieve the desired results in either of these situations, you can use the following process after adding the storage (after adding the SDs to the Storage Pool):

1. Create a dummy file system, using all available space.
2. Expand the Storage Pool into the new SDs.
3. Expand the almost full target file system to use some (or all) of the space added to the Storage Pool. Note that the expansion should be at least 50% of the added storage capacity.
4. Run the file system data redistribution utility.
5. Delete the dummy file system.

Snapshots and the File System Data Redistribution Utility

When the file system data redistribution utility is run and snapshots are enabled, the old data is preserved. As a result, snapshots will grow, consuming a lot of disk space. The space used by these snapshots is not freed until all snapshots present when the file system data redistribution utility was started have been deleted.

There are four options available to recover the space used by snapshots:

1. Allow the snapshots to be deleted according to the snapshot configuration.
This option recovers space the slowest, but could be used in scenarios where the space won't be required immediately after the file system data redistribution utility completes.
2. Manually delete snapshots after running the file system data redistribution utility.
This option recovers space more quickly than option 1.
3. Manually kill snapshots after running the file system data redistribution utility.
This option also recovers space more quickly than options 1 or 2, but it requires that the file system is taken offline.
4. Disable snapshots (and therefore backups) and kill/delete existent snapshots before running the file system data redistribution utility.
This option avoids the snapshot space usage problem altogether.

Creating System Drives

When your system was first installed, SDs were pre-configured by SGI. You can, however, create additional SDs as needed when adding disks to partially-populated enclosures, adding expansion enclosures, upgrading to higher-capacity disks, or when you have unused disks in an enclosure and you need to use that capacity.

After creating additional SDs, you should place them in appropriate SD groups as described in [System Drive Groups](#), on page 27.



Caution: Before creating system drives, set aside at least one disk to be used as a hot spare. The SMU will automatically configure a hot spare of the correct type and size.

1. **Navigate to the System Drives page (Home > Storage Management > System Drives).**
2. **In the System Drives page, click Create.**
3. **Select a rack.**

When the **Select RAID Rack** page is displayed, select a rack, then click **next**.

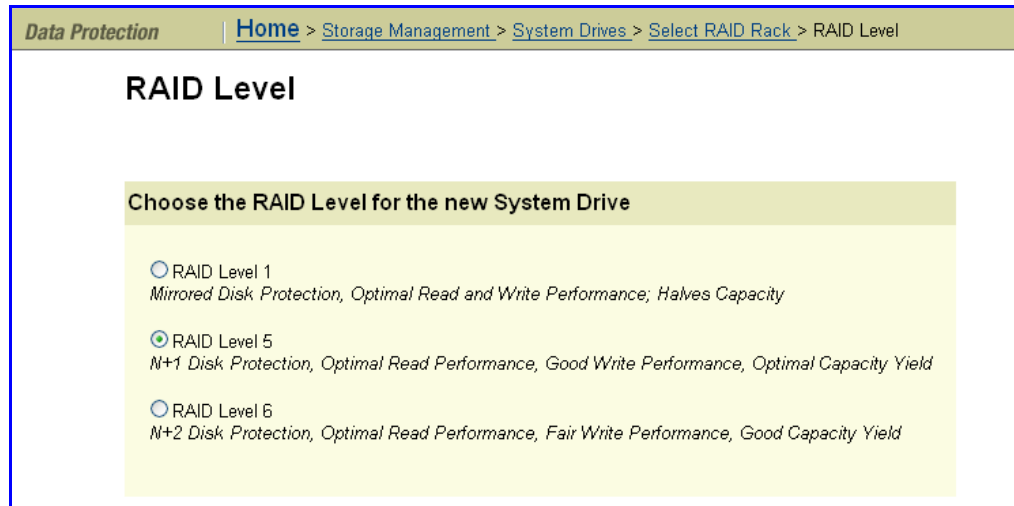
Select RAID Rack

Select the Rack on which to create a System Drive.

Name	Rack Status
<input checked="" type="radio"/> 20000050CC00BFBB	<input checked="" type="radio"/> OK
<input type="radio"/> 50050CC00200423F	<input checked="" type="radio"/> OK

4. **Indicate RAID level.**

In the **RAID Level** page, select a RAID level, then click **Next**.



For information about RAID levels 1, 5 and 6, refer to the following table.

RAID Level	System Drive Size	Description and Failover	Advantages/Disadvantages
1	2 to 32 disks, up to 2 TB or, >2 TB with the proper controller firmware and disk type (contact SGI Global Services for more information)	Provides mirroring and duplexing. If a physical disk fails and a hot spare disk is available, the controller automatically inserts the spare and builds onto it the contents of the failed disk.	Data written to one disk is duplicated to a second for maximum data protection, but with only 50% usable capacity.
5	3 to 16 FC or SAS disks, >2 TB, or 3 to 16 SATA disks up to 2 TB	Provides a combination of striping and parity checking. If a physical disk fails and a hot spare disk is available, the controller automatically inserts the spare and builds onto it the contents of the failed disk, from data and parity information on the remaining disks. For many applications, RAID 5 offers a good compromise between capacity, reliability, and speed.	Parity checking provides redundancy without the overhead of having to double the disk capacity.

RAID Level	System Drive Size	Description and Failover	Advantages/Disadvantages
6	3 to 16 disks, >2 TB	<p>Provides a combination of striping and two disk parity checking.</p> <p>If one or two physical disks fail and a hot spare disk is available, the controller automatically inserts the spare and builds onto it the contents of the failed disk, from data and parity information on the remaining disks. For many applications, RAID 6 offers the best compromise between capacity, reliability, and speed.</p>	<p>All parity data is distributed on two disks (dual parity), and parity checking provides redundancy without the overhead of having to double the disk capacity. Provides fault tolerance for two drive failures, making larger RAID groups more practical, especially for high availability systems.</p> <p>As large-capacity drives become more common, this becomes increasingly important because large-capacity drives lengthen the time needed to recover from the failure of a single drive. Single parity RAID levels are vulnerable to data loss until the failed drive is rebuilt, and the larger the drive, the longer the time to rebuild. Dual parity reduces the amount of time data is volatile during the recovery of the failed drive.</p>

The Create System Drive page will appear.

Storage Management | [Home](#) > [Storage Management](#) > [System Drives](#) > [Select RAID Rack](#) > [Select RAID Level](#) > Create System Drive

Create System Drive - rack Eng-2ef (titans 2e, 2f) RAID Level 5

Create System Drive from an **existing** RAID Group. | Create System Drive in a **new** RAID Group.

Capacity	Number of Physical Disks	Physical Disk Type	Capacity	Number of Physical Disks	Physical Disk Type
<input type="radio"/> 58.41 GB	4	SATA	<input checked="" type="radio"/> 1.36 TB	3	SATA
<input type="radio"/> 58.41 GB	4	SATA			
<input type="radio"/> 58.41 GB	4	SATA			
<input type="radio"/> 58.41 GB	4	SATA			
<input type="radio"/> 58.41 GB	4	SATA			
<input type="radio"/> 58.41 GB	4	SATA			

System Drive Capacity: GB
Do not alter this value without guidance from your support team. If the capacity is less than the available capacity of the selected RAID Group, then a region of unallocated free space will remain in the RAID Group.

Stripe Size:

Unallocated Space: GB

5. Specify the parameters for the drive.

In the Create System Drive page, do the following:

a. Select SD size.

In the Capacity column, click the radio button for the desired size of the SD. (The server calculates SD capacity based on the number of physical disks specified in the Number of Physical Disks column.)



Caution: To ensure optimal performance of the storage subsystem, do not change the value specified in the System Drive Capacity field except under the direction of SGI Global Services.

b. Specify stripe size.

The value specified for the stripe size should always match the value configured at the RAID controller. The recommended stripe size is automatically selected when you select the system drive capacity.

The stripe size, also referred to as the segment size, defines the size of the data patterns written to individual disks in a system drive.

6. Click create.

A RAID system drive will now be created, with the specified label and the selected stripe size.

The RAID controller performs a initialization of the system drive to check for bad sectors and set up the RAID parity. The lights on the disks flicker during this process, and the Active Tasks dialog box shows the progress of the initialization.

7. Verify the SD.

Navigate to the **System Drives** page. (**Home > Storage Management > System Drives.**)

The newly added system drive should appear in the list.

Storage Management | Home > Storage Management > System Drives

System Drives

Licensing
 Current capacity used: 14.55 TB
 Limit: 510.00 TB

Filter
 No Filtering Applied [filter](#)

ID	Capacity	Manufacturer	Label	Comment / Rack Name	Storage Pool	Allow Access	Status	
<input type="checkbox"/> 0	204.35 GB	Manufacturer	Not known		iscsiTest	Allowed	OK	details
<input type="checkbox"/> 1	204.35 GB	Manufacturer	Not known		iscsiTest	Allowed	OK	details
<input type="checkbox"/> 2	204.35 GB	Manufacturer	Not known		iscsiTest	Allowed	OK	details
<input type="checkbox"/> 3	204.35 GB	Manufacturer	Not known		iscsiTest	Allowed	OK	details
<input type="checkbox"/> 4	204.35 GB	Manufacturer	Not known		iscsiTest	Allowed	OK	details
<input type="checkbox"/> 5	204.35 GB	Manufacturer	Not known		iscsiTest	Allowed	OK	details
<input type="checkbox"/> 6	10.00 GB	Manufacturer	Not known		p_test	Allowed	OK	details
<input type="checkbox"/> 7	1.05 TB	Manufacturer	Not known		span1	Allowed	OK	details
<input type="checkbox"/> 8	1.05 TB	Manufacturer	Not known		span1	Allowed	OK	details
<input type="checkbox"/> 9	1.05 TB	Manufacturer	Not known		span1	Allowed	OK	details
<input type="checkbox"/> 10	1.05 TB	Manufacturer	Not known		span1	Allowed	OK	details
<input type="checkbox"/> 11	1.05 TB	Manufacturer	Not known		span1	Allowed	OK	details
<input type="checkbox"/> 12	1.05 TB	Manufacturer	Not known		span2	Allowed	OK	details
<input type="checkbox"/> 13	1.05 TB	Manufacturer	Not known		span2	Allowed	OK	details
<input type="checkbox"/> 14	1.05 TB	Manufacturer	Not known		span2	Allowed	OK	details
<input type="checkbox"/> 15	1.05 TB	Manufacturer	Not known		span2	Allowed	OK	details
<input type="checkbox"/> 16	1.05 TB	Manufacturer	Not known		span2	Allowed	OK	details
<input type="checkbox"/> 17	1.05 TB	Manufacturer	Not known		iSCSI-pool	Allowed	OK	details
<input type="checkbox"/> 18	1.05 TB	Manufacturer	Not known		iSCSI-pool	Allowed	OK	details
<input type="checkbox"/> 19	407.20 GB	Manufacturer	Not known		3992-sp	Allowed	OK	details
<input type="checkbox"/> 20	407.20 GB	Manufacturer	Not known		3992-sp	Allowed	OK	details

[Check All](#) | [Clear All](#)

Actions: [allow access](#) [deny access](#) [forget](#) | [create](#) [Refresh Status](#)

Shortcuts: [RAID Racks](#) [Storage Pools](#) [Active Tasks](#) [System Drive Groups](#)

Format and initialization data can be viewed in real time under the Status column.

8. Create a new file system.

Once the system drive has been initialized, a file system can be created. Refer to the *System Administration Guide*.

Managing System Drive Groups

After system drives (SDs) are created, they should be placed into system drive groups to optimize system performance. If your SD is large enough to use all the space in the RAID group, you cannot have multiple SDs in that RAID group, and you must create a system drive group containing just the one SD. You can create a system drive group either automatically or manually. (See [Creating SD Groups Automatically](#), on page 35 or [Creating SD Groups Manually](#), on page 36.)



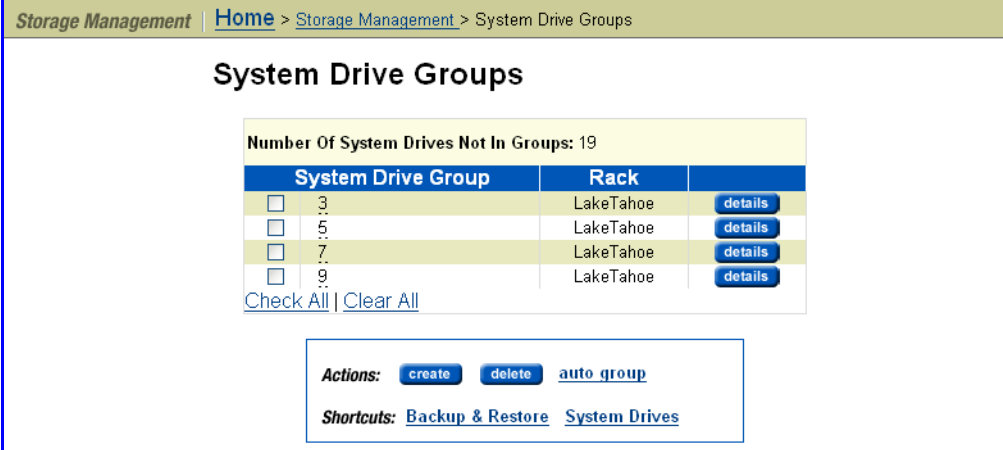
Caution: When creating a system drive group, you must ensure that the system drives being grouped are located on the same physical devices (or RAID group). Otherwise system performance can be adversely affected.

Creating SD Groups Automatically

The following steps describe how to create system drive groups automatically.

1. **Navigate to the System Drives Group page (Home > Storage Management > System Drive Groups).**

The **System Drive Groups** page indicates the number of SDs not in groups and lists the SD groups that have already been created.



Storage Management | Home > Storage Management > System Drive Groups

System Drive Groups

Number Of System Drives Not In Groups: 19

System Drive Group	Rack	details
<input type="checkbox"/> 3	LakeTahoe	details
<input type="checkbox"/> 5	LakeTahoe	details
<input type="checkbox"/> 7	LakeTahoe	details
<input type="checkbox"/> 9	LakeTahoe	details

[Check All](#) | [Clear All](#)

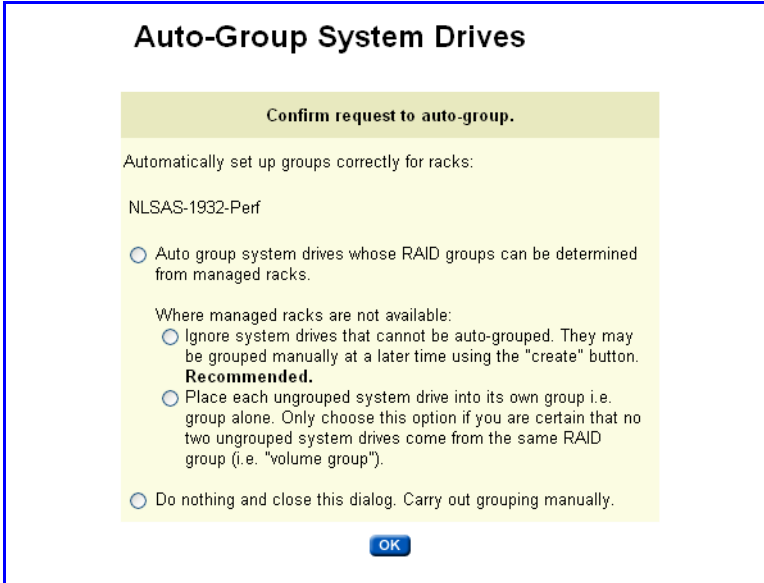
Actions: [create](#) [delete](#) [auto_group](#)

Shortcuts: [Backup & Restore](#) [System Drives](#)

2. **Start auto-grouping the system drives.**

In the **System Drives Group** page, click the **auto_group** link (if the "Number of System Drives Not In Groups" is zero, the auto_group link is unavailable because no groups can be created).

The **Auto-Group System Drives** page opens, allowing you to automatically group system drives.



Auto-Group System Drives

Confirm request to auto-group.

Automatically set up groups correctly for racks:

NLSAS-1932-Perf

Auto group system drives whose RAID groups can be determined from managed racks.

Where managed racks are not available:

Ignore system drives that cannot be auto-grouped. They may be grouped manually at a later time using the "create" button. **Recommended.**

Place each ungrouped system drive into its own group i.e. group alone. Only choose this option if you are certain that no two ungrouped system drives come from the same RAID group (i.e. "volume group").

Do nothing and close this dialog. Carry out grouping manually.

[OK](#)



Note: Grouping system drives automatically can be performed only when the SMU can determine the physical location of the system drives along with the RAID configuration. Currently, the SMU can determine the physical location of the system drives only for some storage subsystems. Con-

tact SGI Global Services for more information about which storage subsystems can provide that information to the SMU. Without such information, you will need to group system drives manually (see [Creating SD Groups Manually](#), on page 36). The storage server cannot automatically group SDs on certain storage subsystems, because those SDs (LUNs) are not directly managed by the storage server. Because of this, when you use the auto-group feature of the storage server, each SD on these storage subsystems will be placed into a separate group.

The SMU can determine the physical location of the system drive if it is managing the rack. The list of managed racks can be found in the **RAID Racks** page (**Home > Storage Management > RAID Racks**).

When you click **OK** to confirm the auto-group request, the SMU automatically groups the system drives located on storage being managed by the SMU. The SMU does the following:

- Checks each existing system drive group to make sure it is correct, adding missing system drives, and removing system drives that should not be in that group.
- Places all system drives not in groups into correct groups based on their location on the physical disks.
- If you fill the "Place each system drive in an individual group" checkbox on the Auto-Group dialog, the SMU places each SD that is present, but in racks not being managed by the SMU, into its own group. To do this, the SMU first checks the SDs to see if it is capable of managing the storage. If it can, the SMU sets those SDs aside and does nothing to them. (When the storage is added to the SMU for management, you can then auto-group those SDs.) The SMU places the remaining SDs (if any) into individual groups.



If none of the RAID/volume groups on the storage subsystem use more than a single SD, select "**Place each system drive in an individual group**".

1. **Click OK to automatically create the SD groups.**

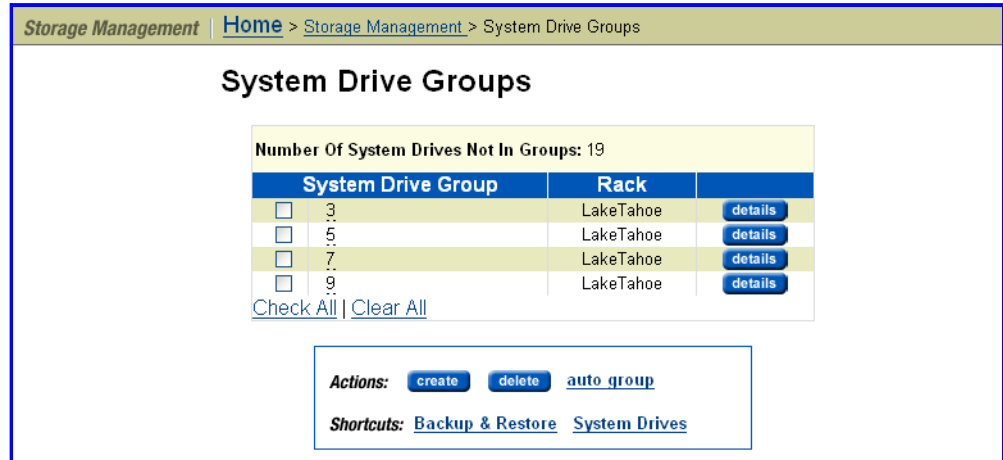
Creating SD Groups Manually

Before creating SD groups manually, you must first determine the physical location of the system drives along with the RAID configuration. This information can be obtained using various third-party applications. With this information, you can determine the SDs that are in the same RAID group, and group them together.

The following steps describe how to create system drive groups manually.

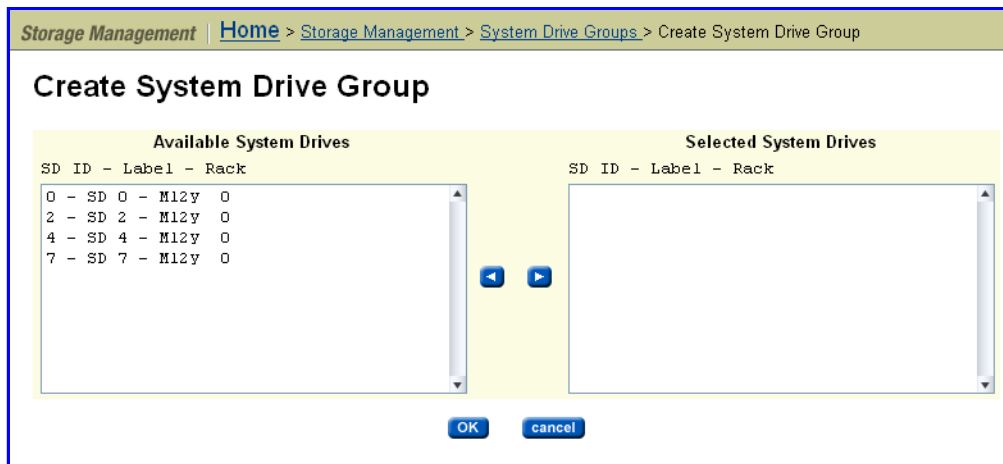
1. **Navigate to the System Drives Group page (Home > Storage Management > System Drive Groups).**

The **System Drive Groups** page indicates the number of SDs not in groups and lists the SD groups that have already been created.



2. Create the System Drives Group.

In the **System Drives Group** page, click **create** (if the "Number of System Drives Not In Groups" is zero, the **create** button is unavailable because no groups can be created). The **Create System Drive Group** page opens, allowing you to manually group system drives.



The **Available System Drives** list shows all system drives in the storage server or cluster that are not already in a group.

Select the SDs you want to group together from the Available System Drives list and click the right arrow to move them to the Selected System Drives list.

3. Create the new System Drive Group.

In the **Create System Drives Group** page, click **OK** to create a new group containing the SDs in the selected list. Repeat these steps as needed to create the desired SD groups.

Backing up or Restoring SD Groups

Backing up and restoring SD groups is a simple, quick, and error-free method of transferring SD group definitions among clusters or servers that share storage or when moving storage between servers/clusters. SD groups are backed up as a part of the normal server/cluster configuration backup process, so backing up SD groups is not necessary for failure recovery.

Backing up SD groups imposes a very slight management overhead, and may cause other management functions to slow down slightly during the short time the backup is being made. There should be no noticeable effect on file serving throughput.

As with backing up SD groups, restoring saved SD groups will not cause a significant overhead, but there are rules about when you can and cannot import groups. For example, you cannot import a group that includes SDs that are used in a storage pool that has currently mounted file systems. For more information on these restrictions, see the `sd-group` command in the *Command Line Reference*.

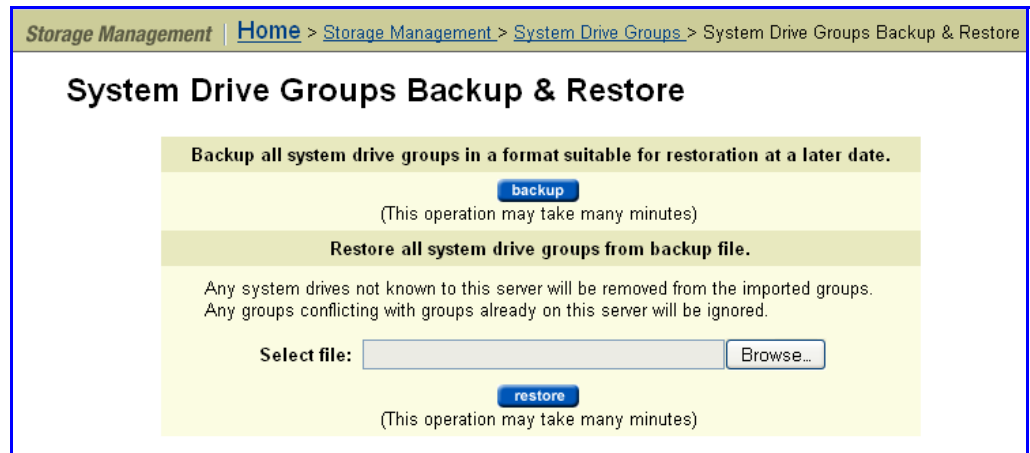


Note: When moving storage between stand-alone servers or clusters, be aware that SD device IDs usually change, which may make it seem as if groups have been imported incorrectly.

To back up or restore SD Groups:

1. Navigate to the System Drive Groups Backup & Restore page.

From the **Storage Management** page, select **System Drive Groups**, then click **Backup & Restore** to display the **System Drive Groups Backup & Restore** page.



The screenshot shows a web browser window with the following content:

- Navigation bar: [Storage Management](#) | [Home](#) > [Storage Management](#) > [System Drive Groups](#) > System Drive Groups Backup & Restore
- Section title: **System Drive Groups Backup & Restore**
- Section 1: **Backup all system drive groups in a format suitable for restoration at a later date.**
 - button: **backup**
 - text: (This operation may take many minutes)
- Section 2: **Restore all system drive groups from backup file.**
 - text: Any system drives not known to this server will be removed from the imported groups. Any groups conflicting with groups already on this server will be ignored.
 - Form: **Select file:**
 - button: **restore**
 - text: (This operation may take many minutes)

2. Backup or restore:

- **To backup:** Click **backup**. In the browser, specify the name and location of the backup file, then click **OK/Save** (the buttons displayed and the method you use to save the backup file depend on the browser you use).

A backup file name is suggested, but you can customize it. The suggested file name uses the syntax:

SD_GROUPSyyyy-mm-dd_time-UTC-offset.txt, where the following example illustrates the appropriate syntax: SD_GROUPS2008-04-30_1729-0700.txt

- **To restore:** Click **Browse** to display a dialog you can use to choose the backup file, navigate to the directory where the backup file is stored, select the backup text file (SD_GROUPS2008-04-30_1729-0700.txt) for the specific export(s) you want to restore, then click **Open**.

When the **System Drive Groups Backup & Restore** page displays the name and location of the selected file, click **restore**.

Modifying System Drive Groups



After system drive groups are created, you can modify a group by adding or removing system drives. Before modifying SD groups, you must first determine the physical location of the system drives along with the RAID configuration. This information can be obtained using various third-party applications. Contact SGI Global Services for information about which third-party applications are supported.

Caution: When modifying a system drive group, you must ensure that the system drives being added to a group are located on the same physical devices (or RAID group) as the system drives already in the group. Otherwise system performance can be adversely affected.

The following steps describe how to modify system drive groups manually.

1. **Navigate to the System Drives Group page (Home > Storage Management > System Drive Groups).**

The **System Drive Groups** page indicates the number of SDs not in groups and lists the SD groups that have already been created.

Storage Management | [Home](#) > [Storage Management](#) > System Drive Groups

System Drive Groups

Number Of System Drives Not In Groups: 19

System Drive Group	Rack	
<input type="checkbox"/> 3	LakeTahoe	details
<input type="checkbox"/> 5	LakeTahoe	details
<input type="checkbox"/> 7	LakeTahoe	details
<input type="checkbox"/> 9	LakeTahoe	details

[Check All](#) | [Clear All](#)

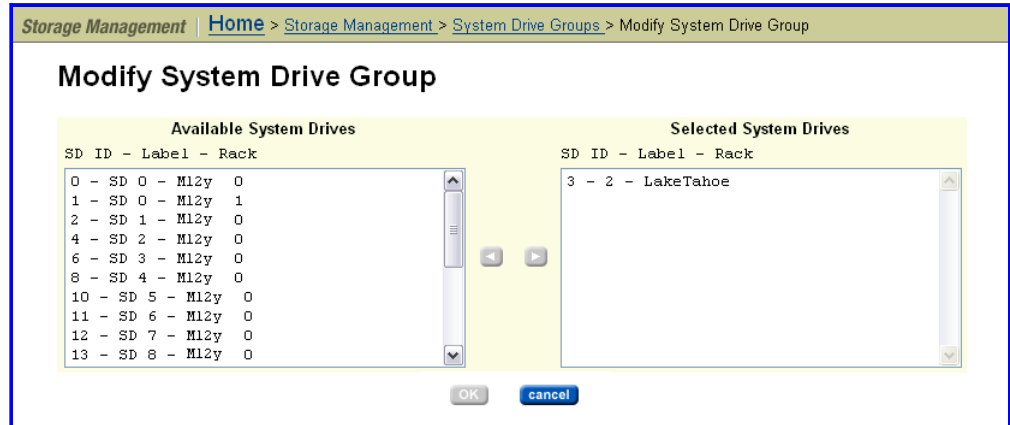
Actions: [create](#) [delete](#) [auto_group](#)

Shortcuts: [Backup & Restore](#) [System Drives](#)

2. **Select the system drive group you want to modify.**

In the **System Drives Group** page, click **details** for the system drive group you want to modify. The **Modify System Drive Group** page opens,

displaying the system drives that are available to be grouped and the SDs in the selected group.



The **Available System Drives** list shows all SDs in the storage server or cluster that are not already in a group. The **Selected System Drives** list shows the SDs that are already included in the selected group.

3. Modify the SD group.

You can add SDs to the SD group by selecting them in the **Available System Drives** list and clicking the right arrow to move the SDs to the **Selected System Drives** list.

You can remove SDs from the SD group by selecting them in the **Selected System Drives** list and clicking the left arrow to move the SDs to the **Available System Drives** list.

4. Review and save your changes.

In the **Modify System Drives Group** page, ensure that you have the desired SDs in the **Selected System Drives** list, and click **apply** to save the SD group with the modified content. Repeat these steps as needed to modify the desired SD groups.