CHALLENGE® RAID
Owner's Guide

CONTRIBUTORS

Written by Carolyn Curtis
Illustrated by Cheri Brown, Dan Young, and Carolyn Curtis
Production by Heather Hermstad
Engineering contributions by Sammy Wilborn, Albert Lui, Dale Witt, Brad Eacker,
    Carl Strasen, and Paul Tsien
Cover design and illustration by Rob Aguilar, Rikk Carey, Dean Hodgkinson,
    Erik Lindholm, and Kay Maitz

CHALLENGE® RAID Owner's Guide
Document Number 007-2532-006

Contents

List of Figures

List of Tables

About This Guide

The CHALLENGE® RAID storage system provides a compact, high-capacity, high-availability source of disk storage for the complete line of Silicon Graphics® CHALLENGE servers running IRIX™ 5.3 and 5.3 with XFS™: CHALLENGE S, CHALLENGE DM, CHALLENGE L, and CHALLENGE XL.

Note: For clarity, CHALLENGE is written as Challenge in the balance of this guide.

The Challenge RAID storage system uses high-availability disk storage in as many as 20 disk modules. For even more storage, the Challenge RAID rack storage system offers up to four RAID chassis assemblies, each with as many as 20 disk modules. The chassis assemblies in a Challenge RAID rack can be connected to one or more SCSI buses on Challenge servers separately or in combination.

RAID levels 0, 1, 1_0 (0+1), 3, and 5 are supported, as well as disks configured as individual disk arrays or as hot spares. In addition, a basic Challenge RAID storage system provides storage-system caching.

You administer the Challenge RAID storage system through a command-line interface (CLI) or a graphical user interface. This guide explains how to perform all Challenge RAID operations on both interfaces.

Structure of This Guide

This guide contains the following chapters:

- Chapter 1, "Features of the Challenge RAID Storage System," introduces the main Challenge RAID components and summarizes RAID levels and data availability and performance features.

- Chapter 2, "Storage System Configurations," explains Challenge RAID configurations in detail: basic, dual-interface/dual-processor, and split-bus.

- Chapter 3, "Operating the Storage System," describes how to check status, identify failing components, and start and shut down the storage system.

- Chapter 4, "Configuring Disks," explains how to use the command-line interface to group disks into RAID-5 groups and how to display or change information on groups of disks.

- Chapter 5, "Maintaining Disk Modules," explains how to replace a failed disk module and add a disk module array.

- Chapter 6, "Identifying Failed System Components," explains how to get status information on failed components other than disk modules.

- Chapter 7, "Caching," explains how to determine, set up, and change caching parameters.

- Chapter 8, "Reconfiguring," explains how to change LUN configuration and parameters, transfer control of a LUN, and update Licensed Internal code and PROM code.

- Appendix A, "Technical Specifications," summarizes technical information for the Challenge RAID deskside storage system.

- Appendix B, "The raidcli Command-Line Interface," lists and explains all parameters of the *raid5* command.

- Appendix C, "Storage-Control Processor Event-Log Error Codes," lists and explains the error codes that appear in the SP event log.

An index completes this guide.

Conventions

In command syntax descriptions and examples, square brackets ([ ]) surrounding an argument indicate an optional argument. Variable parameters are in italics. Replace these variables with the appropriate string or value.

In text descriptions, IRIX filenames are in italics. The names of keyboard keys are printed in boldface typewriter font and enclosed in angle brackets, such as `<Enter>` or `<Esc>`.

Messages and prompts that appear on-screen are shown in fixed-width type. Entries that are to be typed exactly as shown are in boldface fixed-width type.

Compliance Statements

This section lists various domestic and international hardware compliance statements that pertain to the system.

FCC Warning

This equipment has been tested and found compliant with the limits for a Class A digital device, pursuant to Part 15 of the FCC rules. These limits are designed to provide reasonable protection against harmful interference when the equipment is operated in a commercial environment. This equipment generates, uses, and can radiate radio frequency energy and if not installed and used in accordance with the instruction manual, may cause harmful interference to radio communications. Operation of this equipment in a residential area is likely to cause harmful interference, in which case the user will be required to correct the interference at personal expense.

VDE 0871/6.78

This equipment has been tested to and is in compliance with the Level A limits per VDE 0871.

European Union Statement

This device complies with the European Directives listed on the "Declaration of Conformity" which is included with each product. The CE mark insignia displayed on the device is an indication of conformity to the aforementioned European requirements.

International Special Committee on Radio Interference (CISPR)

This equipment has been tested to and is in compliance with the Class A limits per CISPR publication 22.

Canadian Department of Communications Statement

This digital apparatus does not exceed the Class A limits for radio noise emissions from digital apparatus as set out in the Radio Interference Regulations of the Canadian Department of Communications.

Attention

Le present appareil numerique n'emet pas de bruits radioelectriques depassant les limites applicables aux appareils numeriques de Classe A prescrites dans le Reglement sur le Brouillage Radioelectrique etabli par le Ministere des Communications du Canada.

Japanese Compliance Statement

　この装置は、第一種情報装置（商工業地域において
使用されるべき情報装置）で商工業地域での電波障害
防止を目的とした情報処理装置等電波障害自主規制協
議会(VCCI)基準に適合しております。
　従って、住宅地域またはその隣接した地域で使用す
ると、ラジオ、テレビジョン受信機等に受信障害を与
えることがあります。
　取扱説明書に従って正しい取り扱いをして下さい。
VCCI-1、 CISPR-22　Class　A

Features of the Challenge RAID Storage System

RAID (redundant array of inexpensive disks) technology provides redundant disk resources in disk-array configurations that make the storage system more highly available and improve reliability and performance. RAID was first defined by D. Patterson, G. Garth, and R. Katz of the University of California, Berkeley, in their 1987 paper, "A Case for Redundant Arrays of Inexpensive Disks (RAID)"[1]. That paper defines various levels of RAID.

This chapter introduces the Challenge RAID disk-array storage system. It explains

- Challenge RAID storage systems
- Challenge RAID storage system components
- data availability and performance
- RAID levels
- the RAID hot spare
- using the Challenge RAID command-line interface
- using the RAID graphical user interface RAIDGUI

Challenge RAID Storage Systems

Figure 1-1 is an external view of the deskside version of the Challenge RAID storage system.

---

[1] University of California, Berkeley, Report No. UCB/CSD/87/391.
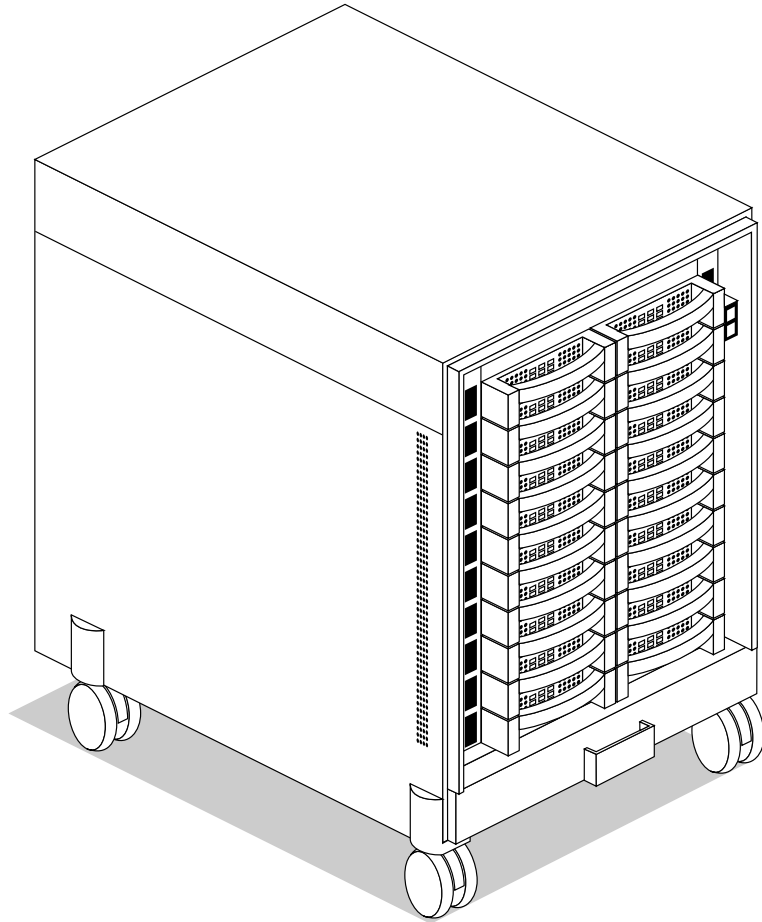
Figure 1-1   Challenge RAID Storage System, Deskside Version: Front View

Note: In Figure 1-1, the front cover is removed for clarity.

Figure 1-2 is an external view of the Challenge RAID rack, with the maximum of four chassis assemblies installed. Each chassis assembly in a Challenge RAID rack corresponds to one deskside Challenge RAID chassis.
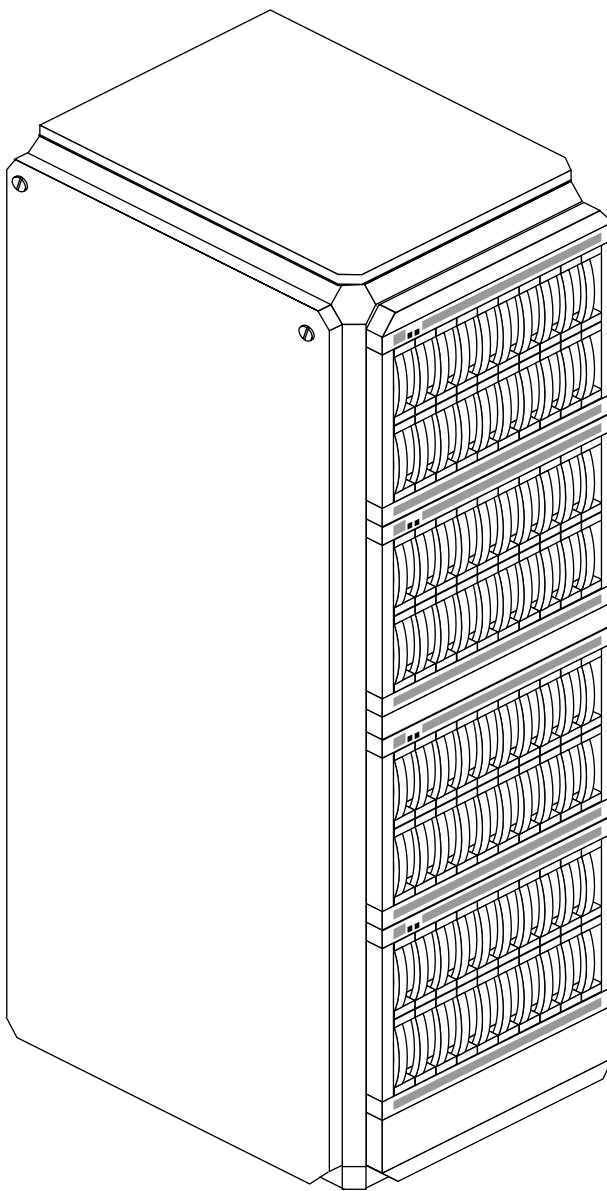
Figure 1-2   Challenge RAID Rack

The Challenge RAID storage system provides a compact, high-capacity, high-availability source of disk storage for your Challenge S, DM, L, or XL server system. The storage system offers a large capacity of high-availability disk storage in multiple disk modules that you can replace when the storage system is turned on.

The Challenge RAID storage system connects by a small computer system interface (SCSI-2) differential bus to a SCSI-2 interface in a Challenge server.

Only Silicon Graphics service personnel authorized to open the computer cabinet and replace parts should install or replace the SCSI-2 interface. You can replace the disk modules by following instructions in Chapter 5 in this guide.

A Challenge server can support multiple Challenge RAID storage systems. The various storage system configurations, along with their availability and performance features, are explained in Chapter 2 in this guide.

The number of Challenge RAID deskside storage systems or rack chassis assemblies that can be connected on one SCSI bus is limited by the recommended SCSI bus length limit of 60 feet, as diagrammed in Figure 1-3.
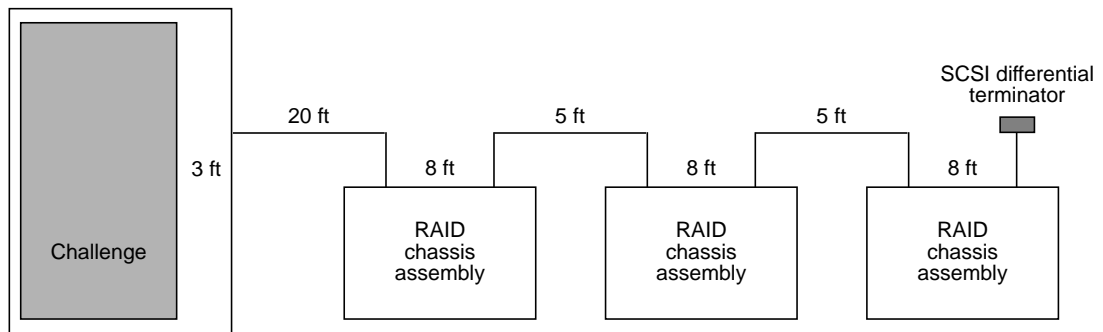
Total length in this example: 57 feet
Recommended maximum length: 60 feet

Figure 1-3   Computing SCSI Cable Length Example: Single-Host Configuration Only

Caution: Although the SCSI bus absolute length limit is 80 feet, exceeding 60 feet on the SCSI bus is not recommended. When cable lengths exceed 60 feet, problems can occur on the SCSI mezzanine card, the IO4B board, or both.

Storage System Components

The Challenge RAID deskside storage system, or each chassis assembly in the Challenge RAID rack, consists of these components:

- one or more host SCSI-2 interfaces that are either

  - native to the POWER Channel™ 2 I/O controller (IO4 board) in the Challenge server

  - HIO add-on cards (mezzanine cards) on the POWER Channel 2 I/O controller

- one or two storage-control processors (SPs)

- 5 to 20 disk modules in groups of five

- one fan module containing six fans, wired in two groups of three fans each

- two or three power supplies (VSCs, or voltage semi-regulated converters)

- one battery backup unit (BBU) for storage system caching (optional)

The Challenge server holds the SCSI-2 interface(s); the Challenge RAID storage system chassis holds the other components.

SCSI-2 Interface

The SCSI-2 interface transfers data between host memory and the SCSI-2 differential bus. A cable connects the SCSI-2 interface to the SP(s) in the storage-system cabinet, as diagrammed in Figure 1-4. The SCSI-2 interface has an operating system-specific device name and ID number.

Challenge RAID Storage-Control Processor

The storage-control processor (SP) is a printed circuit board with memory modules that resides in the storage-system cabinet. It controls the disk modules in the storage system through a synchronous SCSI-2 bus. An SP has five internal fast/narrow SCSI buses, each supporting four disk modules, for a total of 20 disk modules.

Figure 1-4 diagrams a Challenge RAID storage system with one SP.

Challenge server

SCSI-2 interface

SCSI-2 bus

SP A
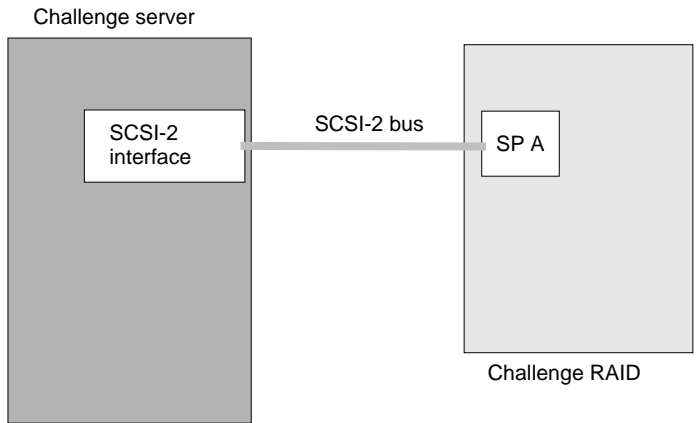
Challenge RAID

Figure 1-4   Challenge RAID Server With One SP

For higher performance, a Challenge RAID storage system can support an additional SP. The second SP provides a second path to the storage system, so both SPs can connect to the same host or two different hosts, as diagrammed in Figure 1-5 and Figure 1-6. With two SPs, the storage system can support storage system caching, whereby each SP temporarily stores modified data in its memory and writes the data to disk at the most expedient time.

Note that the disk modules are owned by the SP, and not by the host, the SCSI-2 bus, or the storage system.
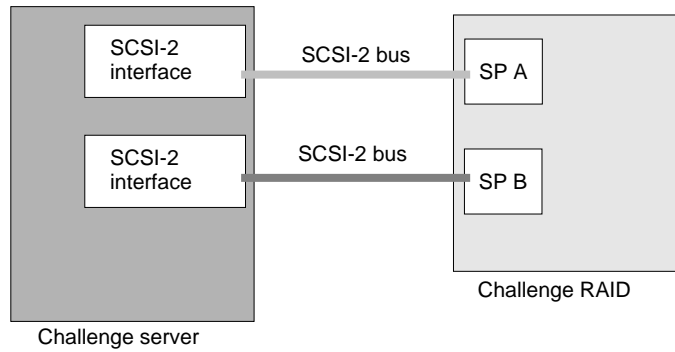
Figure 1-5   SPs Connected to the Same Challenge Chassis
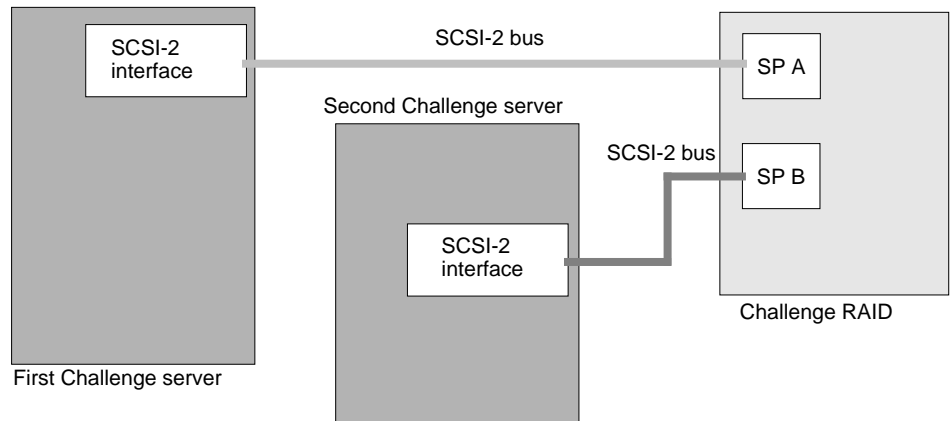
Figure 1-6   SPs Connected to Different Challenge Chassis

Storage System Chassis

The Challenge RAID storage system chassis contains compartments for disk modules, SPs, fan module, power supplies, and battery backup unit. The disk modules face front, and the SP(s), power supplies, battery backup unit, and fan module are accessible from the back.

Each disk module has an ID (the module ID) based on its position in the storage system. The disk modules are inserted in the following order:

- modules A0, B0, C0, D0, and E0 (array 0)

- modules A1, B1, C1, D1, and E1 (array 1)

- modules A2, B2, C2, D2, and E2 (array 2)

- modules A3, B3, C3, D3, and E3 (array 3)

Figure 1-7 diagrams this placement. Individual disk modules have disk position labels attached.
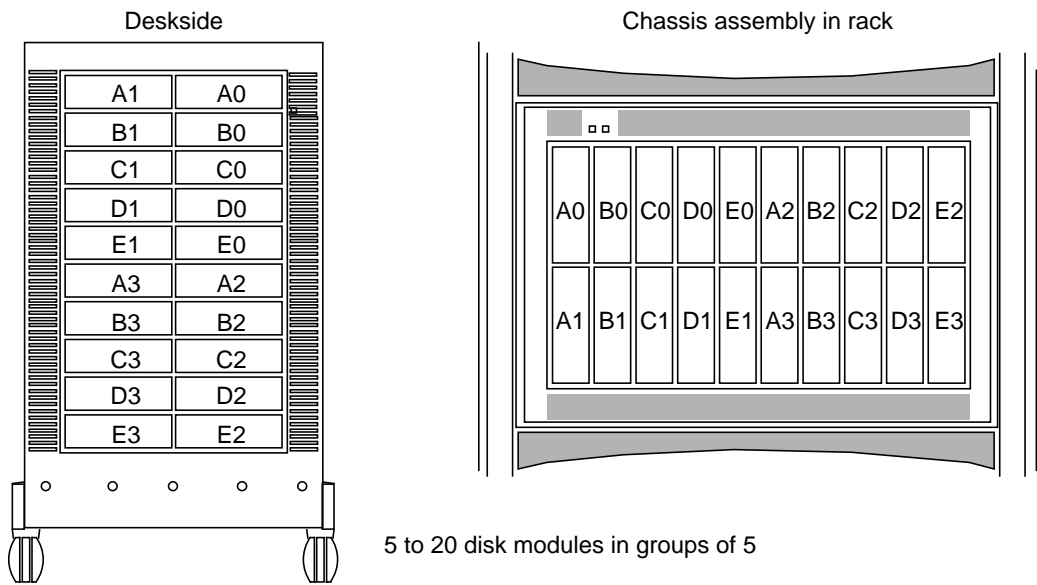
Figure 1-7   Disk Module Locations (Chassis Front View)

Through the SP, the SCSI-2 bus is split into five internal fast/narrow SCSI buses—A, B, C, D, and E—that connect the slots for the disk modules. For example, internal bus A connects the modules in slots A0, A1, A2, and A3, in that order. Figure 1-8 diagrams this configuration.
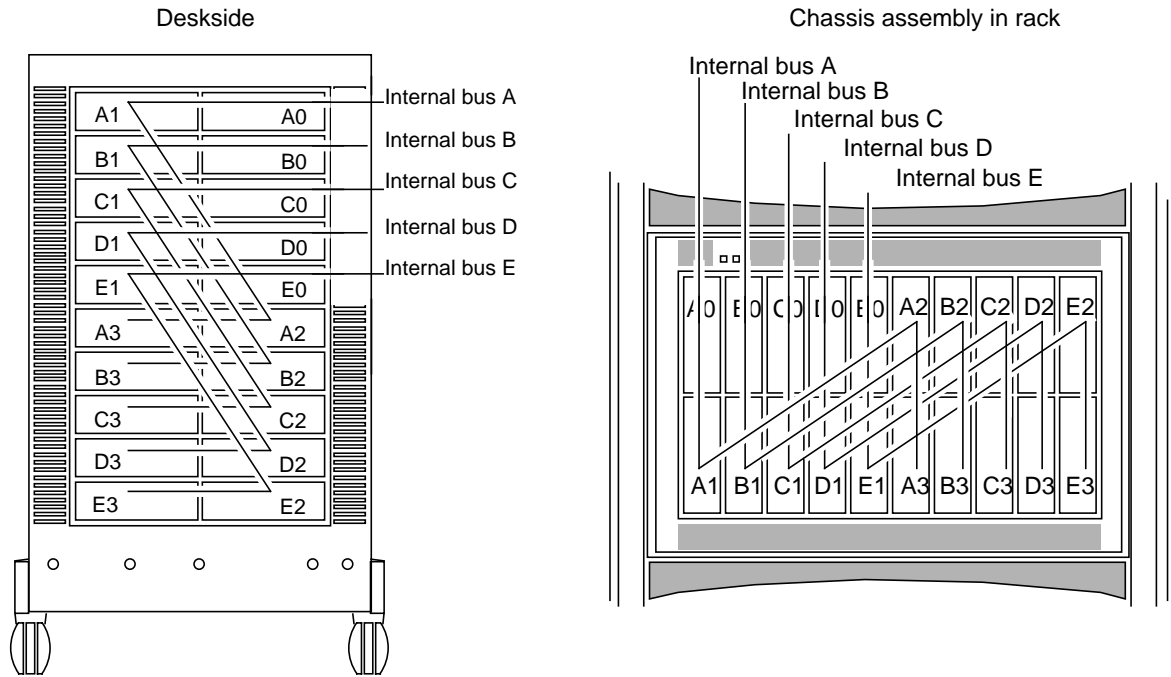
Figure 1-8   SCSI-2 Bus and Internal Buses (Front View)

Disk Modules

A disk module, also called a disk drive module, consists of a disk drive, a power regulator board, internal cabling, and a plastic carrier. The carrier has a handle for inserting and removing the module. Figure 1-9 indicates disk modules in the Challenge RAID chassis and their status lights.
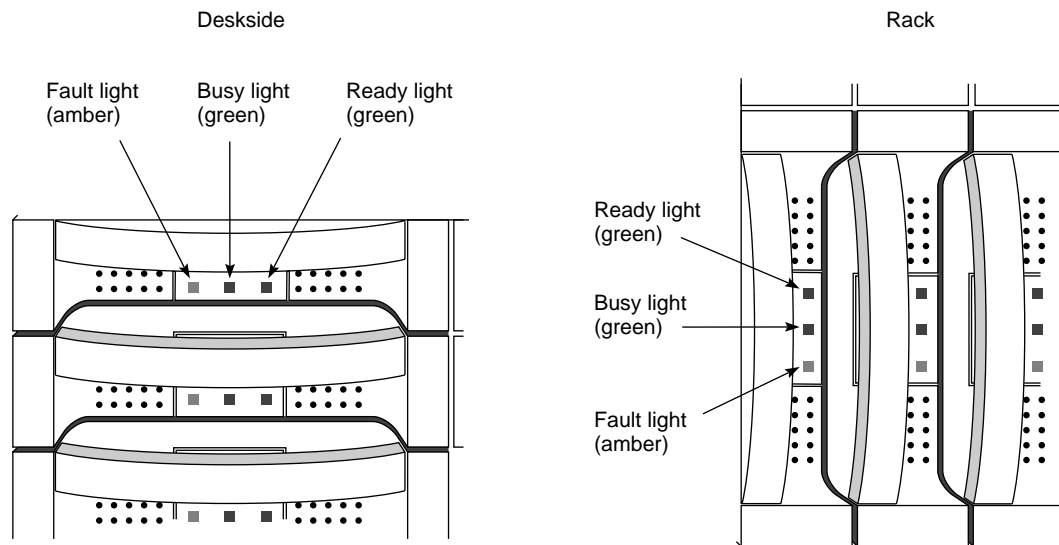
Figure 1-9   Disk Modules and Status Lights

Three status lights on the module indicate the following:

- Ready light (green): lights while the disk module is powered up and ready for use

- Busy light (green): lights while the drive is in use, for example, during formatting or user I/O operations

- Fault light (amber): lights when the module is shut down by the SP because the module failed; also lights after you replace the drive, while the replacement drive spins up to speed

A label attached to the carrier's side shows the disk module's model number and capacity. You can also determine the capacity of a disk module and other features using the command-line interface; see Chapter 3 in this guide.

Data Availability and Performance

The Challenge RAID storage system hardware implements data availability and performance enhancements in these ways:

- data redundancy

- enhanced performance: disk striping

- enhanced performance: storage system caching

- data reconstruction and rebuilding after disk module failure

This section discusses these features.

Data Redundancy

RAID technology provides redundant disk resources in disk-array and disk-mirror configurations that make the storage system more highly available. Data redundancy varies for the different RAID levels supported by Challenge RAID: RAID-0, RAID-1, RAID-1_0, RAID-3, and RAID-5.

A RAID-3 and RAID-5 group maintains parity data that lets the disk group survive a disk module failure without losing data. In addition, the group can survive a single SCSI-2 internal bus failure if each disk module in the group was bound on an independent SCSI-2 internal bus.

A RAID-1 mirrored pair, or a RAID-1_0 group, which uses RAID-1 technology, duplicates data on two groups of disk modules. If one disk module fails, the other module provides continuing access to stored information. Similarly, a RAID-1 mirrored pair or RAID-1_0 group can survive a single SCSI internal bus failure if you bind each disk module on an independent SCSI internal bus.

Enhanced Performance: Disk Striping

In disk striping, the SP lays out data records, usually large data records or a number of small records for the same application, across multiple disks. For most applications, these disks can be written to or read from simultaneously and independently. Because multiple sets of read/write heads work on the same task at once, disk striping can enhance performance.

The amount of information read from or written to each module makes up the stripe element size (for example, 128 sectors). The stripe size is the number of data disks in a group multiplied by the stripe element size. For example, assume a stripe element size of 128 sectors (the default). If the RAID-5 group has five disks (four data disks and one parity disk), multiply by 4 the stripe element size of 128 to yield a stripe size of 512 sectors.

Enhanced Performance: Storage System Caching

Caching is available for Challenge RAID storage systems that have two SPs, each with at least 8 MB of memory, a battery backup unit, and disk modules in slots A0 through E0. With storage system caching enabled, each SP temporarily stores requested information in its memory.

Caching can save time in two ways:

- For a read request, if data is sought after the request is already in the read cache, the storage system avoids accessing the disk group to retrieve the data.

- For a write request, if the information in the write cache is modified by the request and thus must be written to disk, the SP can keep the modified data in the cache and write it back to disk at the most expedient time instead of immediately. Write caching, in particular, can enhance storage system performance by reducing write time response.

To ensure data integrity, each SP maintains a mirror image of the other SP's caches. If one SP fails, the data in its caches is available from the other SP.

As explained in Chapter 7, "Caching," you can enable storage system caching and specify basic cache parameters, and enable or disable read and write caches for individual disk units.

Note: The optional battery backup unit must be present in the Challenge RAID chassis for systems using cache to ensure that data is committed to disk in the event of a power failure.

Data Reconstruction and Rebuilding After Disk Module Failure

All RAID levels except RAID-0 provide data redundancy: the storage system reads and writes data from and to more than one disk at a time. Also, the system software writes parity information that lets the array continue operating if a disk module fails. When a disk module in one of these RAID levels fails, the data is still available because the SP can *reconstruct* it from the surviving disk(s) in the array.

Data *rebuilding* occurs when

•   a hot spare (dedicated replacement disk module) is available

•   the failed disk module is replaced with a new disk module

If a disk module has been configured (bound) as a hot spare, it is available as a replacement for a failed disk module. (See "RAID Hot Spare" later in this chapter.) When a disk module in any RAID level except RAID-0 fails, the SP automatically writes to the hot spare and rebuilds the group using the information stored on the surviving disks. Performance is degraded while the SP rebuilds the data and parity on the new module. However, the storage system continues to function, giving users access to all data, including data stored on the failed module.

Similarly, when a new disk module is inserted to replace a failed one, the SP automatically writes to it and rebuilds the group using the information stored on the surviving disks. As for the hot spare, performance is degraded during rebuilding, but data is accessible.

The length of the rebuild period, during which the SP re-creates the second image after a failure, can be specified when RAID levels are set and disks are bound into RAID units. These processes are explained in "Binding Disks Into RAID Units" in Chapter 4.

RAID Levels

The Challenge RAID system supports these levels of RAID:

- RAID-0 group: nonredundant array
- RAID-1: mirrored pair
- RAID-1_0 group: mirrored RAID-0 group
- RAID-3 group: parallel access array
- RAID-5 group: individual access array
- individual disk unit

Caution: Use only Challenge RAID disk modules to replace failed disk modules. Challenge RAID disk modules contain proprietary firmware that the storage system requires for correct functioning. Using any other disks, including those from other Silicon Graphics systems, can cause failure of the storage system. Swapping disk modules within a Challenge RAID storage system is also not recommended, particularly disk modules in slots A0, B0, C0, and A3, which contain the licensed internal code, and those in slots D0 and E0, which serve with A0, B0, and C0 as the storage system cache vault.

RAID-0 Group: Nonredundant Array

Three to sixteen disk modules can be bound as a RAID-0 group. A RAID-0 group uses striping; see "Enhanced Performance: Disk Striping," earlier in this chapter. You might choose a RAID-0 group configuration when fast access is more important than high availability. On IRIX 5.3 with XFS you can software-mirror the RAID-0 group to provide high availability.

Caution: The hardware does not maintain parity information on any disk module for RAID-0 the way it does for other RAID levels. Failure of a disk module in this RAID level results in loss of data.

RAID-1: Mirrored Pair

In the RAID-1 configuration, two disk modules can be bound as a mirrored pair. In this disk configuration, the SP duplicates (mirrors) the data records and stores them separately on each disk module in the pair. The disks in a RAID-1 pair cannot be split into individual units (as can a software mirror composed of two individual disk units).

Features of this RAID level include

*   fault tolerance

*   automatic mirroring: no commands are required to initiate it

*   physical separation of images

*   faster write operation than RAID-5

With a RAID-1 mirrored pair, the storage system writes the same data to both disk modules in the mirror, as shown in Figure 1-10.
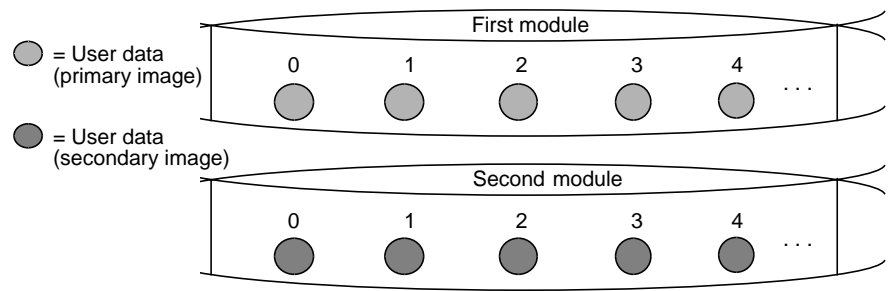
Figure 1-10   RAID-1 Mirrored Pair (Hardware Mirrored Pair)

To achieve the maximum fault tolerance, configure the mirror with each disk module on a different internal SCSI bus; for example, the primary image on A0, the secondary image on B0, and so on.

RAID-1_0 Group: Mirrored RAID-0 Group

A RAID-1_0 configuration mirrors a RAID-0 group, creating a primary RAID-0 image and a secondary RAID-0 image for user data. This arrangement consists of four, six, eight, ten, twelve, fourteen, or sixteen disk modules. These disk modules make up two mirror images, with each image including two to eight disk modules. A RAID-1_0 group uses striping and combines the speed advantage of RAID-0 with the redundancy advantage of mirroring.

Figure 1-11 illustrates the distribution of user data with the default stripe element size of 128 sectors (65,536 bytes) in a six-module RAID-1_0 group. Notice that the disk block addresses in the stripe proceed sequentially from the first mirrored disk modules to the second mirrored disk modules, to the third mirrored disk modules, then from the first mirrored disk modules, and so on.

A RAID-1_0 group can survive the failure of multiple disk modules, providing that one disk module in each image pair survives. Thus, for highest availability and performance, the disk modules in an image pair must be on a different SCSI bus from the disk modules in the other image pair. For example, the RAID-1_0 group shown in Figure 1-11 has three disk modules in each image of the pair.
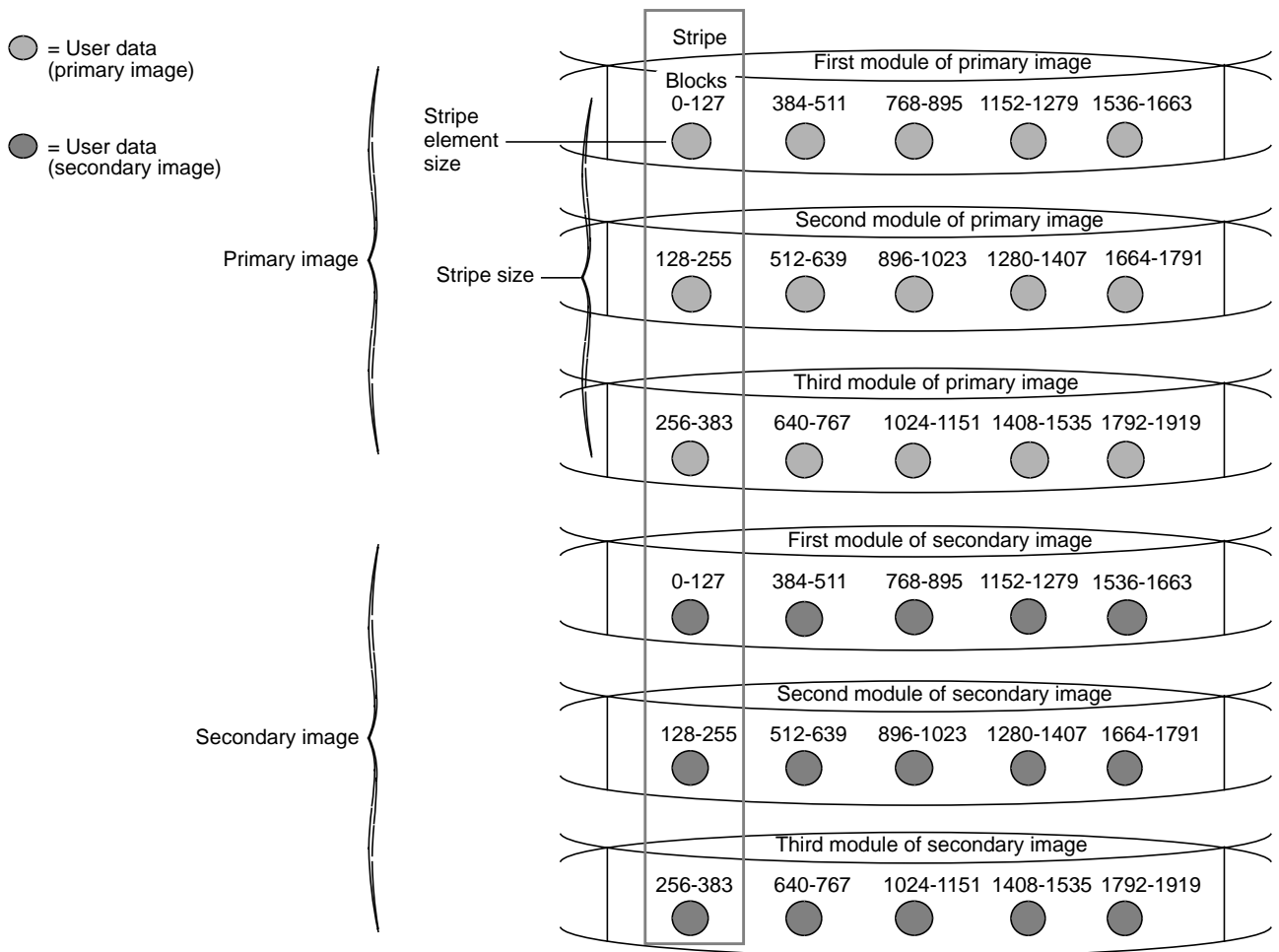
Figure 1-11   Distribution of User Data in a RAID-1_0 Group

When you bind disk modules into a RAID-1_0 group, you must select them in this order: p1, s1, p2, s2, p3, s3, and so on, with primary (p1, p2, p3) and secondary (s1, s2, s3) disk modules on separate internal SCSI buses.

RAID-3: Parallel Access Array

A RAID-3 configuration always consists of five disk modules (four data, one parity) bound as a RAID-3 group. In a RAID-3 group, the hardware always reads from or writes to all its disk modules. A RAID-3 group uses disk striping; see "Enhanced Performance: Disk Striping," earlier in this chapter for an explanation of this feature. RAID-3 striping has a fixed stripe size of one sector.

The Challenge RAID storage system writes parity information that lets the group continue operating if one of the disk modules fails. When you replace the failed module, the SP can rebuild the group using the information stored on the working disks. Performance is degraded while the SP rebuilds the data or parity on the new module. However, the storage system continues to function and gives users access to all data, including data that had been stored on the failed module.

RAID-3 differs from RAID-5 in several important ways:

- In a RAID-3 group, the hardware processes disk requests serially, whereas in a RAID-5 group the hardware can interleave disk requests.

- In a RAID-3 group, the parity information is stored on one disk module; in a RAID-5 group, it is stored on all disks.

- In a RAID-3 group, the I/O occurs in small units (one sector) to each disk.

A RAID-3 group works well for single-task applications that use I/Os of one or more 2 KB blocks, aligned to start at disk addresses that are multiples of 2 KB from the beginning of the logical disk.

When you plan a RAID-3 group for highest availability, configure the disk modules on different internal SCSI buses (A, B, C, D, and E). Figure 1-12 illustrates user and parity data with a data block size of 2 KB within a RAID-3 group. Notice that the byte addresses proceed from the first module to the second, third, and fourth, then to the first, and so on.

Data block

= User data

= Parity data

Stripe element size

Stripe

Bytes

First module

0-511	2048-2559	4096-4607	6144-6655	8192-8603
				…

Second module

512-1023	2560-3071	4608-5519	6656-7167	8604-9115
				…

Third module

1024-1535	3072-3583	5520-5631	7168-7679	9116-9627
				…

Fourth module

1536-2047	3584-4095	5632-6143	7680-8191	9628-10139
				…

Fifth module

Parity	Parity	Parity	Parity	Parity
				…
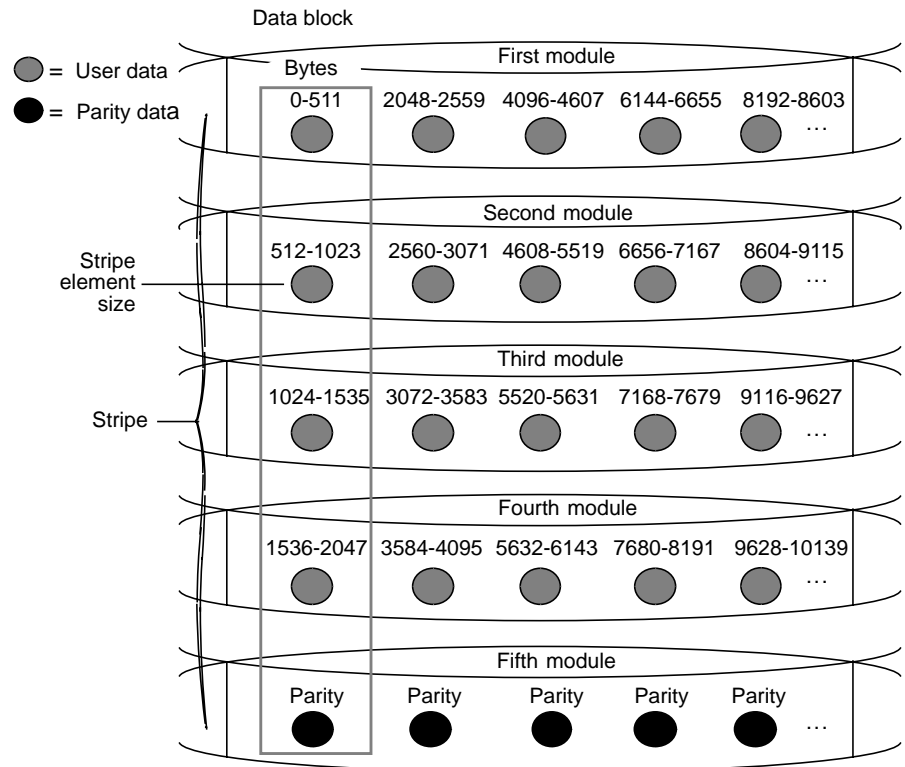
Figure 1-12   Distribution of User and Parity Data in a RAID-3 Group

The storage system performs more steps writing data to a RAID-3 group than to all the disks in a RAID-1 mirrored pair or a RAID-0 group, or to an individual disk unit. For each correctly aligned 2 KB write operation to a RAID-3 group, the storage system performs the following steps:

1.   Calculate the parity data.

2.   Write the new user and parity data.

If the write is not a multiple of 2 KB or the starting disk address of the I/O does not begin at an even 2 KB boundary from the beginning of the logical disk, the storage system performs the following steps:

1. Read data from the sectors being written and parity data for those sectors.

2. Recalculate the parity data.

3. Write the new user and parity data.

RAID-5: Individual Access Array

This configuration usually consists of five disk modules (but can have three to sixteen) bound as a RAID-5 group. Because there are five internal SCSI-2 buses in the Challenge RAID system, an array of five disk modules (or fewer) provides the greatest level of data redundancy.

A RAID-5 group maintains parity data that lets the disk group survive a disk module failure without losing data. In addition, in Challenge RAID storage systems, the group can survive a single SCSI-2 internal bus failure if each disk module in the group was bound on an independent SCSI-2 internal bus. For highest data availability for a RAID-5 group, the disk modules making up the group should be on different SCSI internal buses (A, B, C, and so on).

With RAID-5 technology, the hardware writes parity information to each module in the array. If a module fails, the SP can reconstruct all user data from the user data and parity information on the other disk modules. After you replace a failed disk module, the SP automatically rebuilds the disk array using the information stored on the remaining modules. The rebuilt disk array contains a replica of the information it would have contained had the disk module never failed.

A RAID-5 group uses disk striping; see "Enhanced Performance: Disk Striping," earlier in this chapter for an explanation of this feature. Figure 1-13 illustrates user and parity data with the default stripe element size of 128 sectors (65,536 bytes) in a five-module RAID-5 group. The stripe size comprises all stripe elements. Notice that the disk block addresses in the stripe proceed sequentially from the first module to the second, third, fourth, and fifth, then back to the first, and so on.
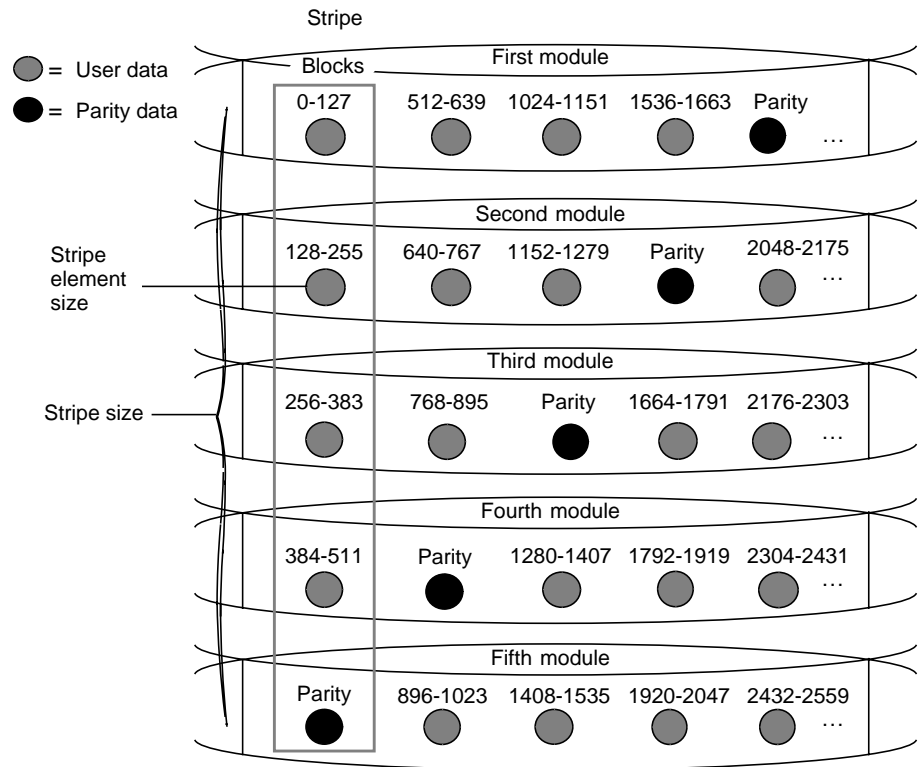
Figure 1-13   Distribution of User and Parity Data in a RAID-5 Group

For each write operation to a RAID-5 group, the Challenge RAID storage system must perform the following steps:

1. Read data from the sectors being written and parity data for those sectors.

2. Recalculate the parity data.

3. Write the new user and parity data.

Individual Disk Unit

An individual disk unit is a disk module bound to be independent of any other disk module in the cabinet. An individual unit has no inherent high availability, but you can make it highly available by software-mirroring it with another individual unit, preferably one on a different internal SCSI bus.

RAID Hot Spare

A hot spare is a dedicated replacement disk unit on which users cannot store information. The capacity of a disk module that you bind as a hot spare must be at least as great as the capacity of the largest disk module it might replace.

Note: The hot spare is not available for RAID-0, because this RAID level does not provide data redundancy.

If any disk in a RAID-5 group, RAID-3, RAID-1 mirrored pair, or RAID-1_0 group fails, the SP automatically begins rebuilding the failed disk module's structure on the hot spare. When the SP finishes rebuilding, the disk group functions as usual, using the hot spare instead of the failed disk. When you replace the failed disk, the SP starts copying the data from the former hot spare onto the replacement disk. When the copy is done, the disk group consists of disk modules in the original slots, and the SP automatically frees the hot spare to serve as a hot spare again.

Note: The SP finishes rebuilding the disk module onto the hot spare before it begins copying data to the new installed disk, even if you replace the failed disk during the rebuild process.

A hot spare is most useful when you need the highest data availability. It eliminates the time and effort needed for someone to notice that a module has failed, find a suitable replacement module, and insert it.

You can have one or more hot spares per storage system. Any module in the storage system can be configured as a hot spare except for modules A0, B0, C0, D0, E0, and A3, because they can store the licensed internal code or serve as the storage system's cache vault.

For example, assume that the modules in slots A0-E0 are a RAID-5 group, those in slots A1 and B1 are a RAID-1 mirrored pair, and the module in A2 is a hot spare, as shown in Figure 1-14. If module D0 fails, the SP immediately begins rebuilding the RAID-5 group using the hot spare. When it finishes, the RAID-5 group consists of disk modules A0, B0, C0, A2, and E0.

When you replace the failed module in D0, the SP starts copying the structure on A2 to D0. When it finishes, the RAID-5 group once again consists of modules A0-E0 and the hot spare becomes available for use if any other module fails. A similar sequence would occur if, for example, module A1 in the mirrored pair failed.
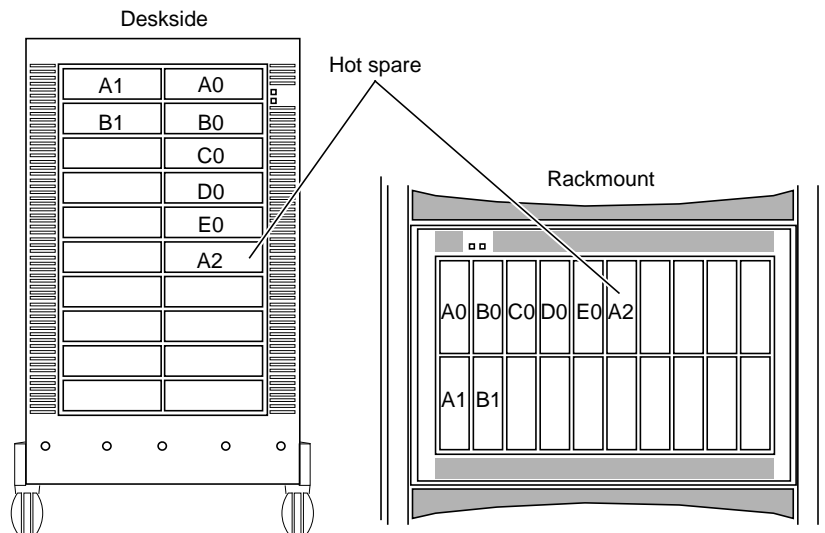
Figure 1-14   Hot Spare Example

Using the Challenge RAID Command-Line Interface

Run the command-line interface, */usr/raid5/raidcli*, in an IRIX window on your Challenge server to

- bind (group) or unbind physical disks into a RAID-0, RAID-1, RAID-1_0, RAID-3, or RAID-5 unit, individual disk unit, or hot spare

- change parameters on a currently bound group (logical unit number, or LUN)

- get names of devices controlled by the SP

- change or get information about the caching environment

- get information about the SP's configuration

- get information on all CRUs (customer-replaceable units, that is, components other than disk modules)

- display status information on disks

- display the SP log

- display information about a group of disks

- perform housekeeping operations, such as clearing the error log or updating firmware

The relevant parameters of the command-line interface are explained for each task in the rest of this guide. Appendix B is a complete guide to the command-line interface.

Using the RAID Graphical User Interface

The RAID graphical user interface RAIDGUI is provided for managing multiple Challenge RAID disk-array storage systems connected to multiple servers on a network. You manage the storage systems using a graphics console on a host, called a client, anywhere on the network. You can manage storage systems on different servers from the same client or the same storage system from different clients. A single host can be both a client and a server, just a client, or just a server.

RAIDGUI's features include

- remote configuration of the storage systems

- notification of the status of the storage systems

- remote management of the storage systems

- electronic distribution of Licensed Internal Code (LIC) to the storage systems

RAIDGUI simplifies the management of storage-system chassis configuration, and fault and problem monitoring:

- chassis configuration management

  Use RAIDGUI to configure the storage-system chassis on local and remote servers. You can bind disk modules into physical disk units (also called LUNs) and change the user-accessible parameters of LUNs, such as their rebuild time, and storage-control processor (SP) owner. You can also set up storage-system caching and update the Licensed Internal Code (LIC) and PROM code that controls a storage system.

- fault and problem monitoring

  To determine the status of the storage-system chassis under its management, the RAID graphical user interface periodically polls the chassis, and gathers any alarms or events that are generated. When the chassis is operating normally, the chassis icon contains a drawing of an intact chassis. When a failure occurs in a chassis, the chassis icon changes or an error popup message appears. You can convert the chassis icon to a window to determine which LUN or component failed.

A client system must run RAIDGUI; a server system must run the storage-system agent, hereafter called the *agent*. The agent and RAIDGUI communicate with each other over a network using remote procedure calls (RPCs). The agent and the Licensed Internal Code (LIC) in the storage system communicate with each other over a SCSI bus. Figure 1-15 diagrams this architecture.
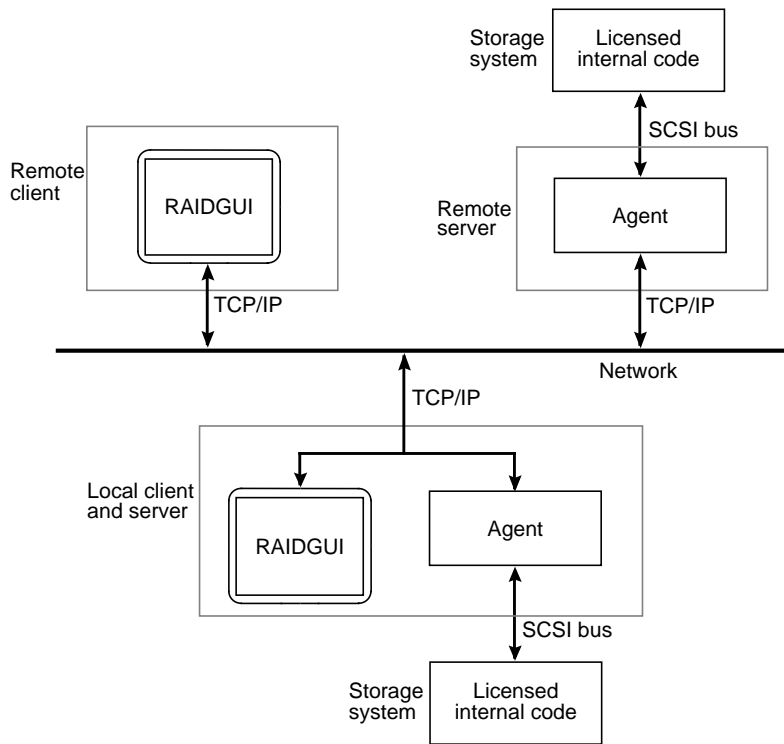
Figure 1-15   RAIDGUI Management Environment Architecture

Storage System Configurations

This chapter explains the various Challenge RAID configurations. Use it to plan your storage system or whenever you contemplate changes in your storage system or physical disk configuration.

A Challenge RAID storage system is configured on two levels:

- storage-system configuration: the number of storage-control processors and SCSI-2 interfaces

- disk configuration within the storage system

Before you can plan your disk configuration, you must understand storage system configuration. Several storage system configurations are available for Challenge RAID storage systems. Table 2-1 lists the hardware components making up each configuration and summarizes the features of each.

This chapter discusses these configurations in separate sections. Each section explains the error recovery features of the configuration.

Table 2-1   Challenge RAID Configurations

Configuration	Host	SCSI-2 Interface	SCSI-2 Bus	SPs	Feature
Basic	1	1	1	1	Applications can continue after failure of any disk module, but cannot continue after failure of SCSI-2 interface or SP.
Dual-interface/dual-processor	1	2	2	2	Provides highest availability and best storage system performance for single-host configurations. Applications can continue after any disk module fails.
Split-bus	2	2 (1 per server)	2 (1 per server)	2	Resembles two basic configurations side by side. Each host and its applications can continue after any disk module fails. The host using a failed SCSI-2 interface or SP cannot continue after failure, but the other host can. If one host, SCSI-2 adapter, or SP fails, the other host can take over the failed host's disks with system operator intervention.
Dual-bus/dual-initiator	2	4 (2 per server)	2 (1 per server)	2	Provides highest availability and best storage-system performance for dual-host configurations. With RAID of any level other than 0, applications can continue after failure of any disk module. If one host, SCSI-2 adapter, or SP fails, the other host can take over the failed host's disk units with system operator intervention.
					This configuration is required for Silicon Graphics FailSafe™ and Oracle Parallel Server™ (OPS™).

Basic Configuration

The basic configuration has one host with one SCSI-2 interface connected by a SCSI-2 bus to the SP in the storage system.

The system can survive failure of a disk module within a redundant RAID group, but it cannot continue after failure of a SCSI-2 interface or SP. Table 2-2 lists the error recovery features of the basic configuration.

Table 2-2   Error Recovery: Basic Configuration

Failing Component	Continue After Failure?	Recovery
Disk module	Yes	Applications continue running. System operator replaces module.
Storage-control processor	No	Storage system fails. System operator replaces SP and restarts operating system.
Fan module	Yes	Applications continue running. System operator replaces module.
Power supply	Yes	If redundant power supply module is present, applications continue running; otherwise, storage system fails. Service provider replaces power supply.
SCSI-2 interface	No	I/O operations to storage system disk units fail. Authorized service provider replaces interface, and system operator restarts operating system and applications.
SCSI-2 cable	No	I/O operations to storage system fail. System operator replaces cable, and restarts operating system and applications.

Dual-Interface/Dual-Processor Configuration

The dual-interface/dual-processor configuration has one host with two SCSI-2 interfaces, each connected by a SCSI-2 bus to a different SP in the storage system.

For better performance with this configuration, you can bind some physical disk units on one SP and some other physical disk units on the other SP. The SP that binds a physical disk unit is the default owner of that physical disk unit.

The storage system can continue running after failure of a disk module within a redundant RAID group. It cannot continue after a SCSI-2 interface or an SP fails unless you manually transfer disk ownership. Table 2-3 lists these features.

Table 2-3   Error Recovery: Dual Interface/Dual-Processor Configuration

Failing Component	Continue After Failure?	Recovery
Disk module	Yes	Applications continue running. System operator replaces module.
Storage-control processor	Yes	I/O operations fail to the disk units owned by a failing SP. System operator can transfer control of the failed SP's disk units to the working SP, shut down the host, power off and on the storage system, reboot the host, and, when convenient, replace the SP and transfer control of disk units to the replacement SP.
Fan module	Yes	Applications continue running. Silicon Graphics SSE or other authorized service provider replaces module.
Power supply	Yes	If redundant power supply module is present, applications continue running; otherwise, storage system fails. Service provider replaces power supply.

Table 2-3 (continued)   Error Recovery: Dual Interface/Dual-Processor
Configuration

Failing Component	Continue After Failure?	Recovery
SCSI-2 interface	Yes	I/O operations fail to storage system disk units owned by the SP attached to the failed interface. System operator can transfer control of the failed SP's disk units to the SP on the surviving interface, shut down the host, power off and on the storage system, and reboot the host. When convenient, the Silicon Graphics SSE or other authorized service provider can replace the interface and the system operator can transfer control of disk units to the replacement SP.
SCSI-2 cable	Yes	I/O operations fail to storage-system disk units owned by the SP attached to the failed cable. System operator can transfer control of these disk units to the other SP, shut down the host, power off and on the storage system, reboot the host, replace the cable, and transfer control of disk units to the replacement SP.

In the example diagrammed in Figure 2-1, one group of five disk modules is bound by storage-control processor A (SP A) and another group of five disk modules is bound by SP B.
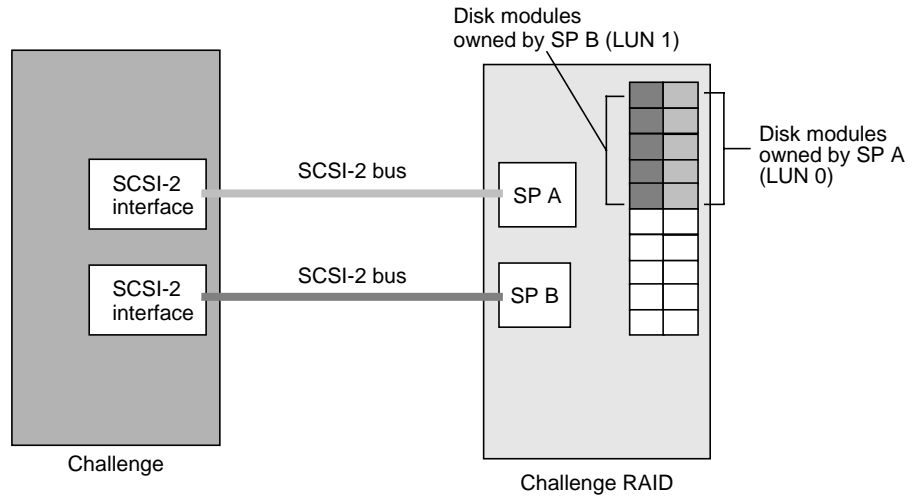
Figure 2-1   Dual-Interface/Dual-Processor Configuration Example

In this example, if one SP or SCSI-2 interface fails, stored data in either LUN is available through the alternate path. Automatic path switching in the event of an SP or SCSI-2 interface failure is possible if XLV volumes and applicable patches are used. For information on XLV volumes, see *Getting Started With XFS Filesystems.*

Note: Only qualified Silicon Graphics System Service Engineers can replace SPs or SCSI-2 interfaces.

Split-Bus Configuration

The split-bus configuration has two hosts, each with a SCSI-2 interface connected by a SCSI-2 bus to a storage-control processor in the storage system. Each host uses its own disks in the storage system independently.

The split-bus configuration resembles two basic configuration systems side by side. This configuration can be used for sites requiring high availability because either host can continue after failure of any disk module within a disk array, and a host can take over a failed host's disks. A host cannot continue after a SCSI-2 interface or an SP fails unless you manually transfer disk ownership.

Table 2-4 lists the error recovery features for this configuration.

Table 2-4   Error Recovery: Split-Bus Configuration

Failing Component	Continue After Failure?	Recovery
Disk module	Yes	Applications continue running. System operator replaces module.
Storage-control processor	Yes	I/O operations fail to the disk units owned by a failing SP. System operator can transfer control of failed SP's disk units to the working SP, shut down the host, power off and on the storage system, and reboot the host. Silicon Graphics SSE or other authorized service provider replaces the SP and transfers control of disk units to the replacement SP.
Fan module	Yes	Applications continue running. Silicon Graphics SSE or other authorized service provider replaces the module.
Power supply	Yes	If redundant power supply module is present, applications continue running; otherwise, storage system fails. Service provider replaces power supply.

	Table 2-4 (continued)	Error Recovery: Split-Bus Configuration
Failing Component	**Continue After Failure?**	**Recovery**
SCSI-2 interface	Yes	I/O operations fail to storage-system disk units owned by the SP attached to the failed interface. System operator can transfer control of the failed SP's disk units to the SP on the interface in the other host, shut down the other host, power off and on the storage system, and reboot the other host. Silicon Graphics SSE or other authorized service provider replaces the interface.
SCSI-2 cable	Yes	I/O operations fail to storage-system disk units owned by the SP attached to the failed cable. System operator can transfer control of these disk units to the other SP, shut down the host, power off and on the storage system, reboot the host, replace the cable, and transfer control of disk units to the replacement SP.

In the example diagrammed in Figure 2-2, one group of five disk modules is bound by storage-control processor A (SP A), which is connected via a SCSI-2 bus to one Challenge server; another group of five disk modules is bound by SP B, which is connected by a different SCSI-2 bus to the second Challenge server.
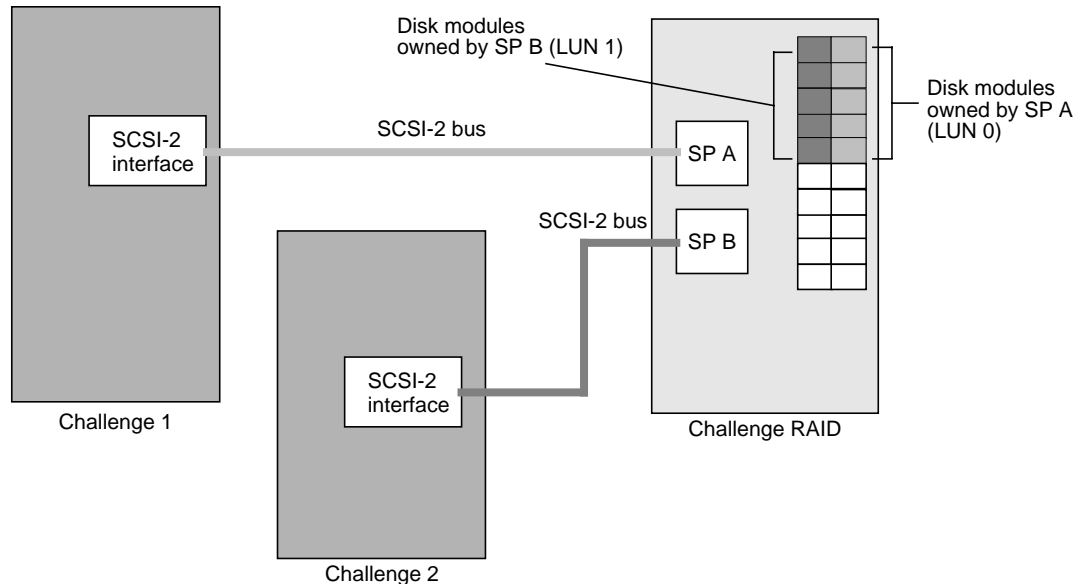
Figure 2-2   Split-Bus Configuration Example

Caution: This configuration does not afford failover capability.

If one SP fails or if the SCSI-2 connection from one host is broken, that host does not have access to the Challenge RAID storage system until the SP is replaced or the SCSI-2 connection is repaired. The host using the remaining SCSI-2 connection and remaining operational SP still has full access to its own data.

The storage-control processor that binds a disk module is the default owner of the disk module. The route through the SP that owns a disk module is the primary route to the disk module. The route through the other SP is the secondary route to the disk module.

In a dual-interface system, either Challenge server can use any of the disk modules in the storage system, but only one Challenge server at a time can use a disk module.

Dual-Bus/Dual-Initiator Configuration

The dual-bus/dual-initiator configuration provides the highest availability. Each host has two SCSI-2 adapters, each of which connects by a separate SCSI-2 bus to a separate SP in the storage system. Since this configuration protects against a SCSI-bus cable failure, it provides higher availability than the dual-initiator configuration. It is for enterprises requiring the highest level of availability, such as the Oracle Parallel Server and FailSafe products.

For better performance with this configuration, you can bind some physical disk units on one SP and the other physical disk units on the other SP. The SP that binds a physical disk unit is its default owner. The route through the SP that owns a physical disk unit is the primary route to the physical disk unit. The route through the other SP is the secondary route to the physical disk unit, and is available if a component in the primary route fails. Table 2-5 lists the error recovery features of the dual-bus/dual-initiator configuration.

Caution: Because both hosts can access the same disk modules simultaneously, the danger exists that one host can overwrite data stored by the other. This configuration requires specific hardware and software (such as a database lock manager) to protect the integrity of the stored data.

Table 2-5   Error Recovery: Dual-Bus/Dual-Initiator Configuration

Failing Component	Continue After Failure?	Recovery
Disk module	Yes	With RAID levels specified at any level other than 0, applications continue running. System operator replaces module.
Storage-control processor	Yes	I/O operations fail to the disk units owned by the failing SP. System operator can transfer control of the failed SP's disk units to the surviving SP, shut down the host, power the storage system off and on, and reboot both hosts. When convenient, Silicon Graphics SSE or other authorized service provider replaces the SP, and the system operation can transfer control of the disk units to the replacement SP.
Fan module	Yes	Applications continue running. System operator replaces module.

Table 2-5 (continued)   Error Recovery: Dual-Bus/Dual-Initiator Configuration

Failing Component	Continue After Failure?	Recovery
Power supply	Yes	If redundant power supply module is present, applications continue running; otherwise, storage system fails. Service provider replaces power supply.
SCSI-2 interface	Yes	The host with the failed adapter cannot access the disks owned by the SP connected to the failed adapter. System operator can transfer control of these disks to the SP connected to the working adapter, shut down both hosts, power off and on the storage system, and reboot both hosts. When convenient, the Silicon Graphics SSE or other authorized service provider can replace the interface, and the system operator can transfer control of disk units to replacement SP.
SCSI-2 cable	Yes	I/O operations fail to storage-system disk units owned by the SP connected to the failed cable. System operator can transfer control of these disk units to the other SP, shut down the host, power off and on the storage system, reboot the host, replace the cable, and transfer control of the disk units to the replacement SP.
Host	Yes	Operations continue on the surviving host.

In the example diagrammed in Figure 2-3, some modules are bound to one SP, which is their primary owner, and the other disk modules are bound to the other SP, which is their primary owner. Either host can use any of the physical disk units in the storage system, but only one host at a time can use a physical disk unit.
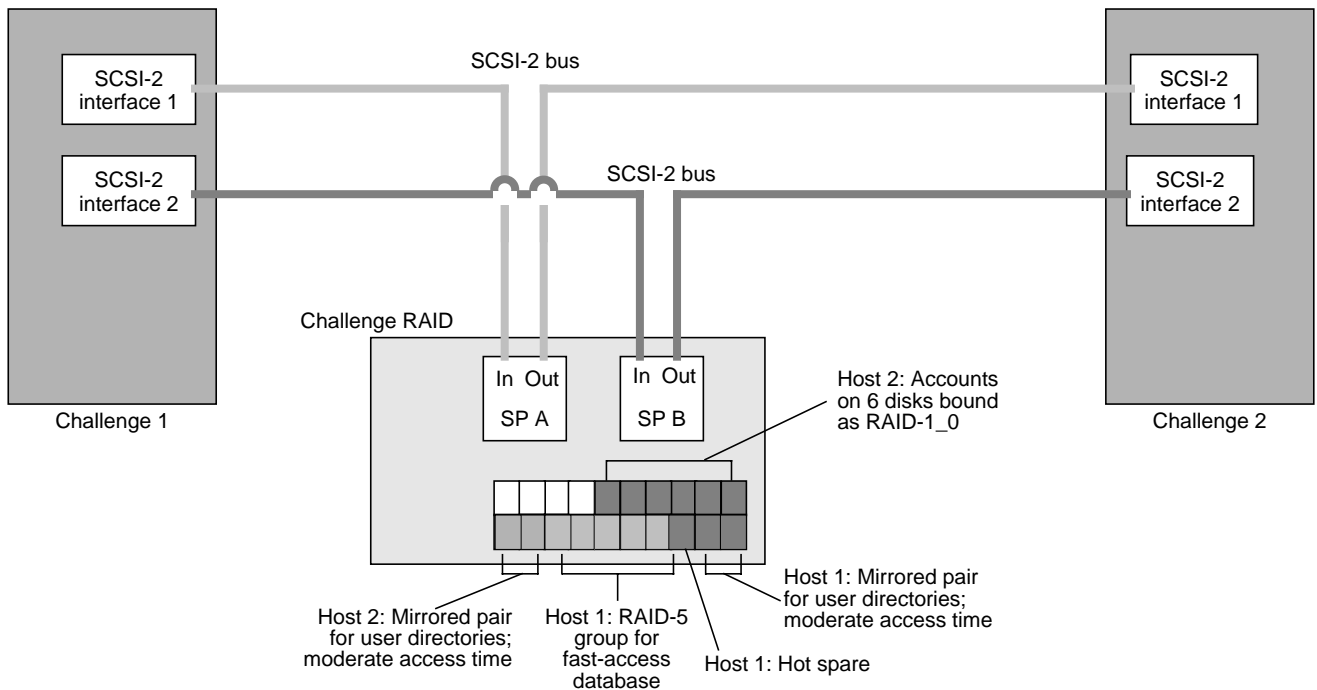
Figure 2-3   Dual-Bus/Dual-Initiator Configuration Example

Caution: Because both hosts have access to the all disk modules and their data in this configuration, it is possible for one host to overwrite the other's data unless appropriate filesystem configuration and failsafe software is installed.

Operating the Storage System

This chapter describes how to operate the Challenge RAID storage system after you have configured it. The chapter explains

- checking storage system status

- shutting down the Challenge RAID storage system

- restarting the Challenge RAID storage system

This chapter introduces the */usr/raid5/raidcli* command (command-line interface, or CLI). Use *raidcli* with its parameters in an IRIX shell on Challenge to get names of devices controlled by the storage-control processor (SP), display status information on disk modules, disk module groups (LUNs), SPs, and other system components, and display the storage processor log, in which error messages are stored.

Note: Although the directory and command are *raidcli*, the command is valid for all RAID levels.

Other chapters in this guide explain how to use the *raidcli* command to bind (group) physical disks into RAID units and unbind them, set up caching, and accomplish other tasks.

Checking Challenge RAID Storage System Status

To check storage system status, you may find it easiest to look at the storage system cabinet to see if the amber service light is lit.

Look for two lights at the right of the disk modules (deskside storage system) or above the disk modules (chassis in rack). The green light indicates that the unit is powered on; the amber light indicates a fault. See Figure 3-1.
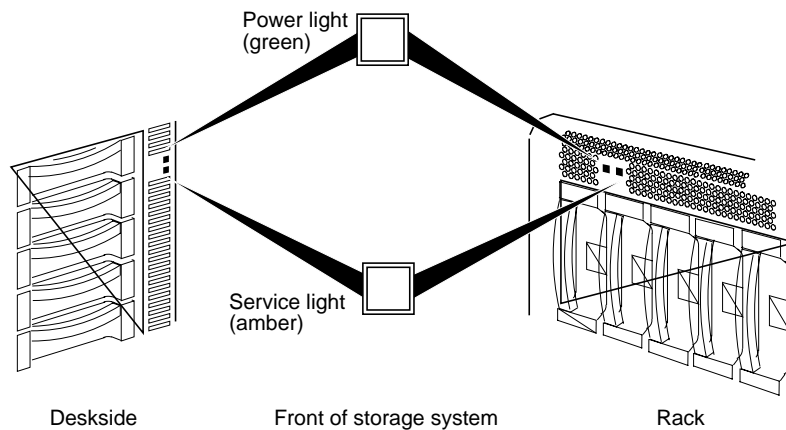
Figure 3-1   Challenge RAID Indicator Lights

The amber service light comes on when

- an SP is reseated

- the Challenge RAID is powered off and on

- the battery backup unit has not finished recharging (if a battery backup unit is present in the system)

If the service light is lit, look for a disk-module fault light that is lit. Then you can either explore status further using the *raidcli* command in an IRIX shell or by using RAIDGUI.

Checking System Status Using the Command-Line Interface

This section explains

- using the *raidcli* command
- getting the device name with *getagent*
- getting general system information
- getting information about disks
- getting information about other components
- displaying the Challenge RAID unsolicited event log

Using the raidcli Command

The *raidcli* command sends storage management and configuration requests to an application programming interface (API) on the Challenge server. For the *raidcli* command to function, the agent—an interpreter between the command-line interface and the Challenge RAID storage system—must be running.

The synopsis of the *raidcli* command is

```
raidcli [-vp] [-d device] parameter [optional_arguments]
```

In this syntax, the variables have the following meanings:

-v   Enables verbose return.

-p   Parses the *raidcli* command without calling the API. If the string does not parse correctly, an error message is printed to *stderr*; otherwise there is no output.

-d *device*   Target RAID device. Use *raidcli getagent* for a list of RAID devices. This switch must be present for all *raidcli* management and configuration commands unless the environment variable indicates otherwise. This switch overrides an environment variable.

Note: Appendix B is a complete alphabetical listing of *raidcli* parameters.

Getting Device Names With getagent

Use the *getagent* parameter with *raidcli* to display information on devices controlled by the API:

```
raidcli getagent
```

The following is a sample output for one device; normally, the output would give information on all devices:

```
Agent Rev: 1.10
Name: Disk Array
Desc: RAID5 Disk Array
Node: sc4d210
Signature:0xf3b51700
Peer Signature: 0x657e0a00
Revision: 7.12.4
SCSI ID: 1
Prom Rev: 0x0076100
SP Memory: 64
Serial No: 94-7240-808
```

Table 3-1 summarizes entries in the *raidcli getagent* output.

Table 3-1   Output of *raidcli getagent*

Entry	Meaning
Agent Rev	Revision number of RAID agent
Name	ASCII string found in the agent configuration file which assigns a name to the node being accessed (see Node description below)
Desc	ASCII string found in the agent configuration file which describes the node being accessed (see Node description below)
Node	The */dev/scsi* entry which the agent uses as a path to the actual SCSI device. This value must be entered by the user for every CLI command (except *getagent*)
Signature	Unique 32-bit identifier for the SP being accessed through Node
Peer Signature	Unique 32-bit identifier for the other SP in the chassis; 0 if no additional SP is present
Revision	Revision of firmware currently running on the SP

Table 3-1 (continued)   Output of *raidcli getagent*

Entry	Meaning
SCSI ID	SCSI ID number
Prom Rev	Current PROM revision on the SP
SP Memory	Amount of DRAM present on the SP
Serial No	12-digit ASCII string that uniquely identifies this subsystem

Getting General System Information

To get general system information, use

```
raidcli -d device getcontrol
```

A partial output of this command follows:

```
System Fault LED: OFF
Statistics Logging: ON
System Cache: ON
Max Requests: 23
Average Requests: 5
Hard errors: 0
Total Reads: 18345
Total Writes: 1304
Prct Busy: 25
Prct Idle: 75
System Date: 4/30/1995
Day of the week: Tuesday
System Time: 12:43:54
```

For more information on this command, see "getcontrol" in Appendix B.

Getting Information About Disk Modules

For information about all disk modules in the system, use this command in an IRIX shell:

```
/usr/raid5/raidcli -d device getdisk
```

For information on a particular disk module, use

```
/usr/raid5/raidcli -d device getdisk [diskposition]
```

In this command, `diskposition` has the format `bd`, where `b` is the bus the disk is located on (a through e; be sure to use lowercase) and `d` is the device number (0 through 3). Figure 3-2 diagrams disk module locations.
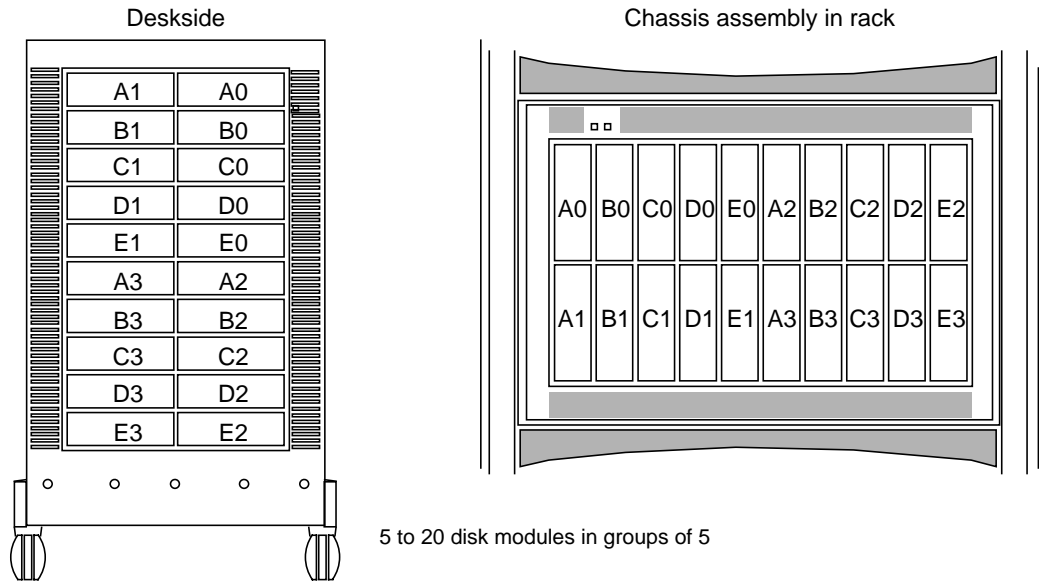
Figure 3-2   Disk Module Locations

For example, the following command gets information about disk module A2:

```
raidcli -d scsi4d210 getdisk a2
```

A sample output of this command follows:

```
A0 Vendor Id: <manufacturer>
A0 Product Id: <part number>
A0 Lun: 0
A0 State: Bound and Not Assigned
A0 Hot Spare: NO
A0 Prct Rebuilt: 100
A0 Prct Bound: 100
A0 Serial Number: 032306
A0 Capacity: 0x000f42a8
A0 Private: 0x00009000
A0 Bind Signature: 0x1c4eb2bc
```

```
A0 Hard Read Errors: 0
A0 Hard Write Errors: 0
A0 Soft Read Errors: 0
A0 Soft Write Errors: 0
A0 Read Retries: 0
A0 Write Retries: 0
A0 Remapped Sectors: 0
A0 Number of Reads: 1007602
A0 Number of Writes: 1152057
```

Table 3-2 interprets items in this output.

Table 3-2   Output of *raidcli getdisk*

Output	Meaning
Vendor Id	Manufacturer of disk drive
Product Id	Part number of disk
Lun	Logical unit number to which this disk is bound
State	Removed: disk is physically not present in the chassis or has been powered off
	Off: disk is physically present in the chassis but is not spinning
	Powering Up: disk is spinning and diagnostics are being run on it
	Unbound: disk is healthy but is not part of a LUN
	Bound and Not Assigned: disk is healthy, part of a LUN, but not being used by this SP
	Rebuilding: disk is being rebuilt
	Enabled: disk is healthy, bound, and being used by this SP
	Binding: disk is in the process of being bound to a LUN
	Formatting: disk is being formatted
Hot Spare	YES or NO
Prct Rebuilt	Percentage of disk that has been rebuilt
Prct Bound	Percentage of disk that has been bound
Serial Number	Serial number from disk inquiry command
Capacity	Actual disk capacity in blocks
Private	Amount of physical disk reserved for private space

Table 3-2 (continued)   Output of *raidcli getdisk*

Output	Meaning
Bind Signature	Unique value assigned to each disk in a logical unit at bind time
Hard Read Errors	Number of hard errors encountered on reads for this disk
Hard Write Errors	Number of hard errors encountered on writes for this disk
Soft Read Errors	Number of soft errors encountered on reads for this disk
Soft Write Errors	Number of soft errors encountered on writes for this disk
Read Retries	Number of retries occurring during reads
Write Retries	Number of retries occurring during writes
Remapped Sectors	Number of sectors that have been remapped
Number of Reads	Number of reads this disk has seen
Number of Writes	Number of writes this disk has seen

Getting Information About Other Components

For state information on other components—field-replaceable units—in the
Challenge RAID storage system besides disk modules, use

```
raidcli -d device getcrus
```

A sample output of this command follows:

```
FANA State: Present
FANB State: Present
VSCA State: Present
VSCB State: Present
VSCC State: Present
SPA State: Present
SPB State: Present
BBU State: Present
```

Note: In this output, information on the power supplies is shown under VSC
(voltage semi-regulated converter), SP information is shown under SP, and
battery backup unit information is shown under BBU.

Table 3-3 interprets items in this output.

Table 3-3   Output of *raidcli getcrus*

Output	Meaning
FANA, FANB	Fan banks A and B.
VSCA	Power supply (voltage semi-regulated converter).
VSCB	Optional third power supply.
SPA	Storage-control processor.
SPB	Optional second storage-control processor.
BBU	Battery backup unit, which has three states: Faulted (removed), Charging, and Present (fully charged or charging).
	If the battery backup unit takes longer than an hour to charge, it shuts itself off and transitions to the "Faulted" state.

Displaying the Challenge RAID Unsolicited Event Log

The storage-control processor maintains a log of event messages in processor memory. These events include hard errors, startups, and shutdowns involving disk modules, fans, SPs, power supplies, and the battery backup unit. Periodically, the SP writes this log to disk to maintain it when SP power is off. The log can hold over 2,000 event messages; it has a filter feature that lets you select events by device or error message.

The event messages are in chronological order, with the most recent messages at the end. To display the entire log, use

```
raidcli -d device getlog
```

To display the newest *n* entries in the log, starting with the oldest entry, use

```
raidcli -d device getlog +N
```

To display the oldest *n* entries in the log, starting with the oldest entry, use

```
raidcli -d device getlog -N
```

Output of the command *raidcli -d device getlog +5* might be

```
12/17/94 09:59;51 A3: (A07) Cru Powered Down [0x47]
12/17/94 09:59;51 A3: (608) Cru Ready [0x0]
12/17/94 09:59;51 A3: (603) Cru Rebuild Started [0x0]
12/17/94 09:59;51 A3: (604) Cru Rebuild Complete [0x0]
12/17/94 09:59;51 A3: (602) Cru Enabled [0x0]
```

These entries show that a field-replaceable unit (disk module, fan unit, battery backup unit, SP) has failed, been replaced, been rebuilt, and been enabled.

At the tail of each log entry is an error code in brackets (for example, [0x47]) that gives diagnostic information when it is available. See "getlog" in Appendix B for explanations of these codes.

To clear the event log, use

```
raidcli -d device clearlog
```

Note: You must be root to use the *clearlog* parameter.

Checking System Status Using RAIDGUI

Any user can run a RAIDGUI session from any system on which it is installed, and can use it to monitor the storage system. Only authorized users can use RAIDGUI to configure or reconfigure the storage system. A user is authorized if the server's agent configuration file contains an entry for the user.

Caution: The agent allows more than one RAIDGUI session to access the same storage system at the same time. You must make sure that two authorized users do not use RAIDGUI to configure or reconfigure the same storage-system chassis at the same time.

This section explains

- starting RAIDGUI
- getting general system information
- getting information about disk modules
- getting information about system components
- viewing SP status information
- viewing the SP event log
- enabling and disabling the statistics log
- using RAIDGUI automatic polling
- using alarm settings
- exiting RAIDGUI

Starting RAIDGUI

To start a RAIDGUI session on a client system, enter

```
xrdb -merge /usr/raid5/Xdefaults
cd /usr/raid5
./raidgui&
```

After introductory screens, the Select Hosts window appears, as shown in Figure 3-3.
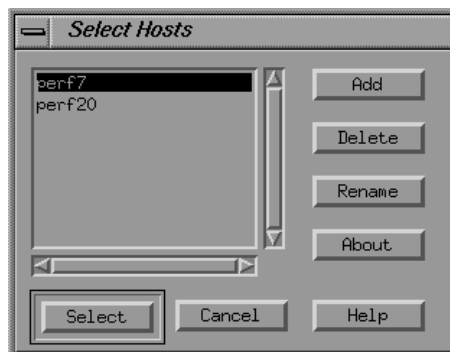
Figure 3-3   Select Hosts

Upon startup, RAIDGUI looks for a file named by the RAID_ARRAY_HOSTS environmental variable. If it can access the named file, the Select Hosts window appears, listing the hosts in the named file. If RAIDGUI cannot find or access this file, or if the RAID_ARRAY_HOSTS environment variable is not set, RAIDGUI looks for the *.hosts* file in your home directory. If it can access this file, the Select Hosts window appears, listing the hosts in the *.hosts* file.

If RAIDGUI cannot find or access the *.hosts* file, a popup appears noting that the host file could not be loaded. In this popup, click *OK*. In the Select Hosts window that appears, follow these steps to add hostnames to the empty selection box:

1. In the Select Hosts window, click *Add*.

2. In the text box that appears, enter the name or IP address of the host you want to add.

3. Click *OK*.

    The hostname you add appears in the selection box, and is added to the file named by the RAID_ARRAY_HOSTS environment variable. If this variable is not set, the hostname is added to the *.hosts* file in your home directory.

You can also use the Select hosts window to remove or rename hosts; click the *Delete* or *Rename* button, respectively. In either case, the information is added to the file named by the RAID_ARRAY_HOSTS environment variable or to the *.hosts* file in your home directory.

Getting General System Information

To get general system information, follow these steps:

1. In the Select Hosts window, select the name of the host with the storage system you want to manage.

2. Click *Select* at the bottom of the window. If RAIDGUI was able to communicate with the agent on the server, the Select Chassis window appears, as shown in Figure 3-4.
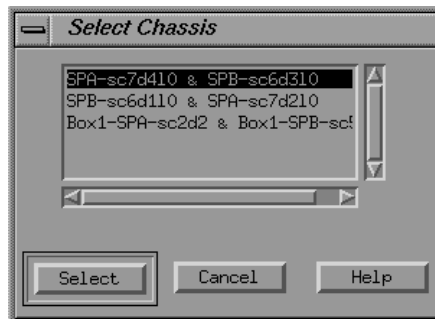
Figure 3-4   Select Chassis Window

3. Select the name of the chassis connection that you want to manage and click *Select*.

    The Select Chassis and Select Hosts windows close, and the Equipment View for the storage system appears. This window depicts the storage system; Figure 3-5 shows an example.
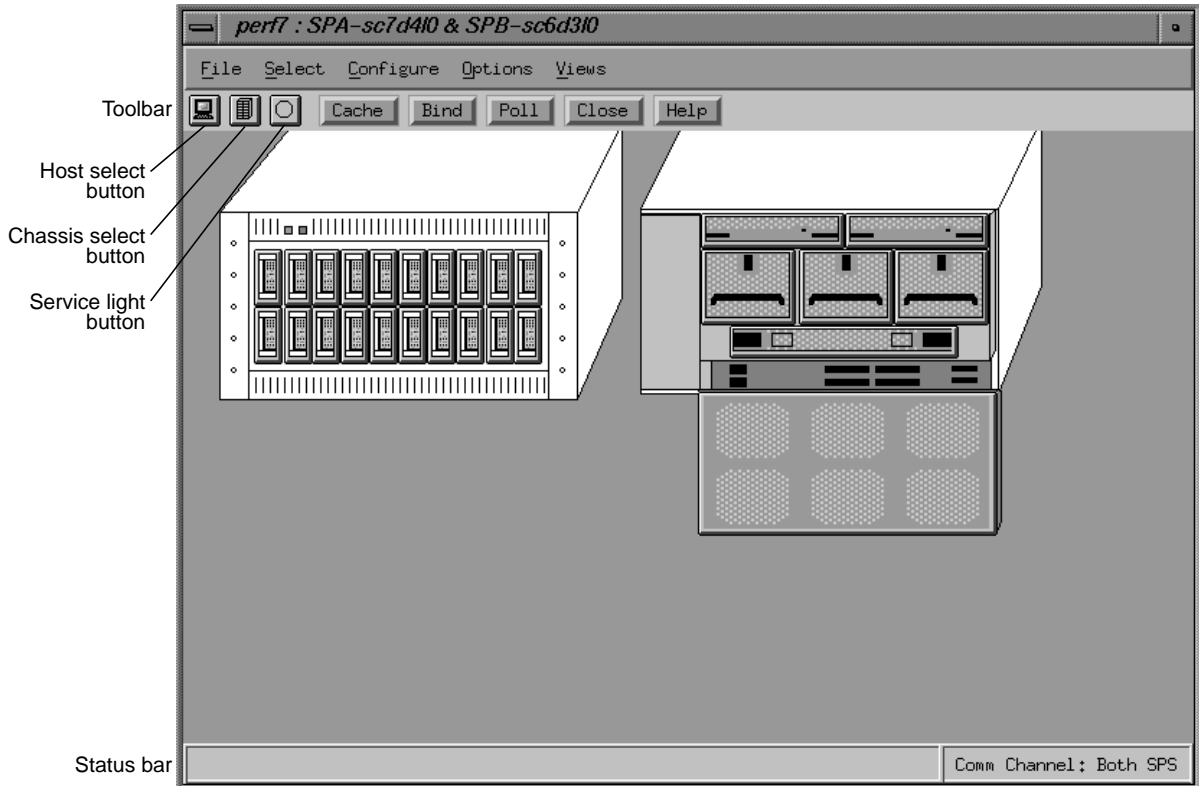
Figure 3-5   Equipment View

Notice the *service light* button in the toolbar. If the circle is amber, a storage-control processor has detected a hard or logical fault in the chassis. You can click this button to display the Chassis Status window, which is explained later in this section.

The color of the service light button usually reflects the state of the physical service light on the Challenge RAID chassis (see Figure 3-1). When the chassis service light is lit, the service light button is usually amber. When the chassis service light is not lit, the service light button is usually gray.

The host select and chassis select buttons display the Select Hosts and Select Chassis windows, respectively.

Note: If the agent or RAIDGUI was started (or restarted) after the chassis was powered up, the color of the service light button does not reflect that of the chassis service light.

The chassis drawing shows the disk modules; storage-control processors (SPs); fan module; voltage semiregulated converters (VSCs), which are power supplies; and a battery backup unit (BBU).

Each drawing of a component and an empty disk slot is a button whose color shows the health of that component:

- gray: operating normally

- blue: disk module is in a transition state, such as powering up or binding (becoming part of a LUN)

- amber: failed (disk module); failed or removed (other components)

- white: component was not present when the agent was started

In the Equipment View, each slot for a disk module has a disk ID by which you identify the disk module, such as A0, A1, A2, A3, B0, B1, B2, and so on.

To display disk IDs in the Equipment View, select "Use Disk IDs" from the Views menu. The window for monitoring LUN status is explained in "Getting Disk Group (LUN) Information" in Chapter 4.

Iconify the Equipment View by clicking the button in the upper right corner.

Getting Information About Disk Modules

Use the Summary View to determine disk module information. Figure 3-6
shows an example.
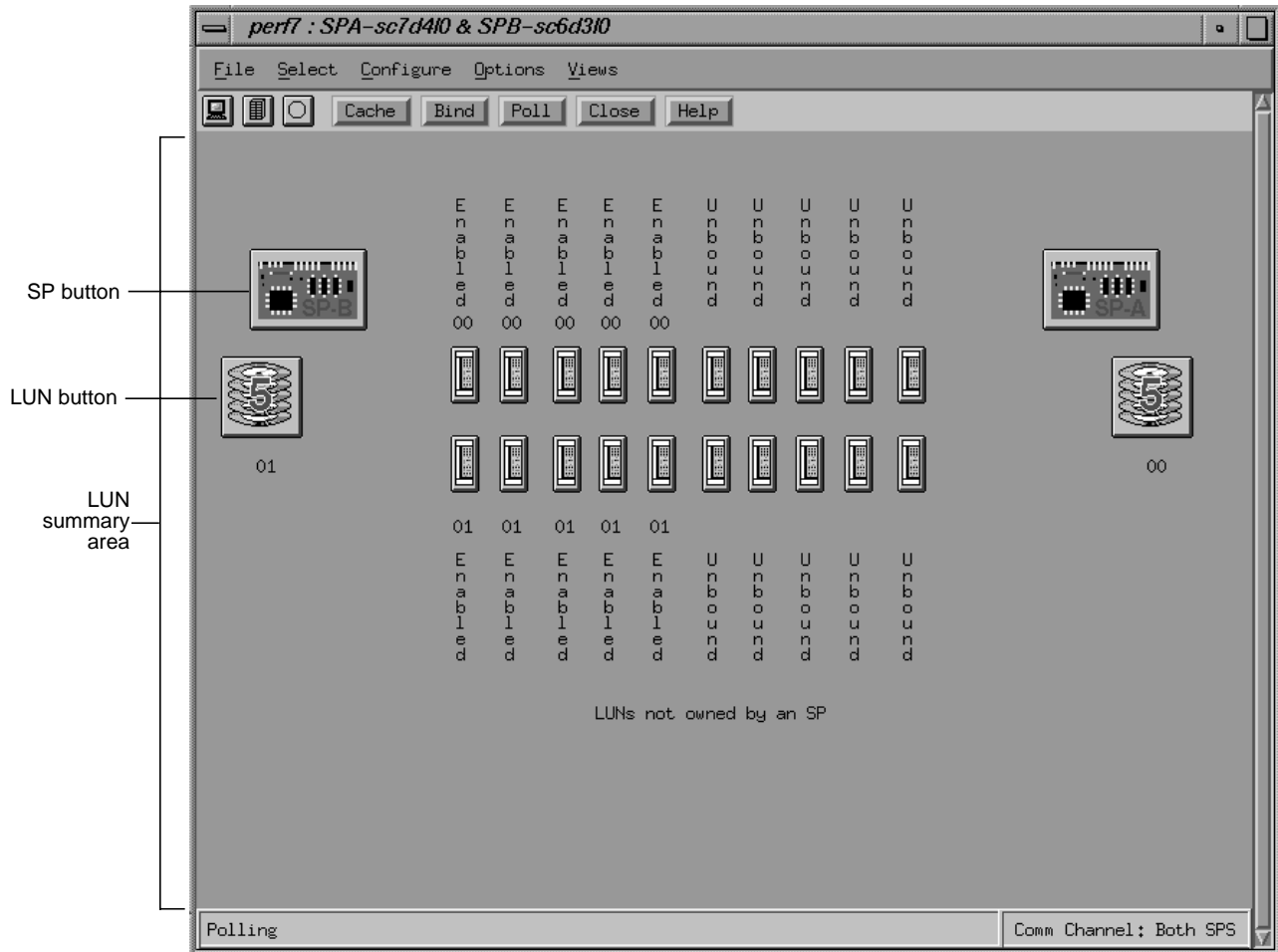
Figure 3-6   Example Summary View

Click an SP button to view information in the SP Summary window, as
explained in "Viewing SP Status Information," later in this chapter.

The LUNs belonging to each SP are shown below them in the Summary View. Disks not owned by an SP, such as hot spares, appear in the middle of the Summary View. "Using the LUN Information Window" in Chapter 4 contains more information about viewing LUN information.

Disk Module Buttons

The positions of the disk module buttons in the Summary View correspond to their positions in the Challenge RAID chassis's disk module slots.

Associated with each disk slot is an identifier based on its position in the chassis. To display disk IDs for the disk modules, select "Use Disk IDs" in the Views menu. Figure 3-7 shows the button for a disk module with disk ID enabled.

Figure 3-7   Disk Module Button With Disk ID

Table 3-4 summarizes the disk IDs.

Table 3-4   Disk IDs

Internal Challenge RAID SCSI Bus	Position	Disk IDs
A	0, 1, 2, 3	A0, A1, A2, A3
B	0, 1, 2, 3	B0, B1, B2, B3
C	0, 1, 2, 3	C0, C1, C2, C3
D	0, 1, 2, 3	D0, D1, D2, D3
E	0, 1, 2, 3	E0, E1, E2, E3

The operational state of the disk module appears beside the button. If the disk module is all or part of a LUN, that LUN's hexadecimal number also appears beside the button. Table 3-5 lists each disk module state, the color of

the disk module's button when the disk module is in that state, and a
description of the state.

Table 3-5   Disk Module States

State	Button Color	Meaning
Binding	Blue	Currently being bound into a LUN.
Empty	White	No disk module was in this slot when the agent was started. This slot can contain an unbound disk module that was removed while the agent was running, or a bound disk module that was removed when the agent was restarted.
Enabled	Gray	Either a hot spare on standby or part of a bound LUN that is assigned to (owned by) the SP you are using as the communication channel to the chassis. If the Challenge RAID storage system has another SP, this module's status is Ready when you use the other SP as the communication channel to the chassis.
Equalize	Blue	Data from a hot spare is being copied onto a replacement disk module.
Failed	Amber	Powered down or inaccessible.
Format	Blue	Being hardware-formatted. Generally, modules do not need hardware formatting.
Off	Amber	Powered off by the SP, which can happen if a wrong size module is inserted.
Power Up	Blue	Power is being applied to the disk module.
Ready	Gray	Module is part of a broken LUN or of a LUN that is bound and unassigned. The disk module might be part of a LUN that is not owned by the SP that you are using as the communication channel to the chassis. If the disk module is part of a LUN assigned to an SP other than the one you are using as the communication chassis, the module's status is either Enabled or Ready. It is Enabled when you use the other SP as the communication channel to the chassis.
Rebuild	Blue	Module replaced a failed module, and the data is being rebuilt on a hot spare or a replacement disk module.
Removed	Amber	Removed from the chassis; applies only to a disk module that is part of a LUN.
Unbound	Gray	Ready to be bound into a LUN.

Click on a disk module button to display the Disk Module Information window. Figure 3-8 shows an example.

Disk capacity ─────

RAID Type ─────

Figure 3-8   Example Disk Module Information Window

In the Disk Module Information window,

- the Description field gives the disk capacity

- the State field indicates the disk module state; see Table 3-5

- the LUN ID field contains the hexadecimal number identifying the LUN to which the disk module is bound; this number is the same as that specified in the Bind window (or with the *raidcli bind* command) when the LUN was bound

- the Owner Type field shows the RAID level of the LUN)

- the Sectors field gives the number of user-accessible sectors on the disk module

Click *Statistics* in the Disk Module Information window to view read/write statistics complied since the last time the statistics log for the SP that owns the LUN was turned on (see "Enabling and Disabling the Statistics Log" later in this chapter for more information). Figure 3-9 shows an example Disk Module Statistics window.

```
┌─────────────────────────────────────────────────────────┐
│ ▭ │ Disk Module Statistics: E0                            │
├─────────────────────────────────────────────────────────┤
│  Average Disk Request Service Time:        19 ms          │
│                                                           │
│  Number of Reads:                          238018         │
│                                                           │
│  Number of Writes:                         130052         │
│                                                           │
│  Number of Blocks Read:                    25975424       │
│                                                           │
│  Number of Blocks Written:                 16646148       │
│                                                           │
│  Number of Read Retries:                   1              │
│                                                           │
│  Number of Write Retries:                  4              │
│                                                           │
│    ┌──────────┐      ┌──────────┐      ┌──────────┐      │
│    │  Close   │      │ Top View │      │   Help   │      │
│    └──────────┘      └──────────┘      └──────────┘      │
└─────────────────────────────────────────────────────────┘
```
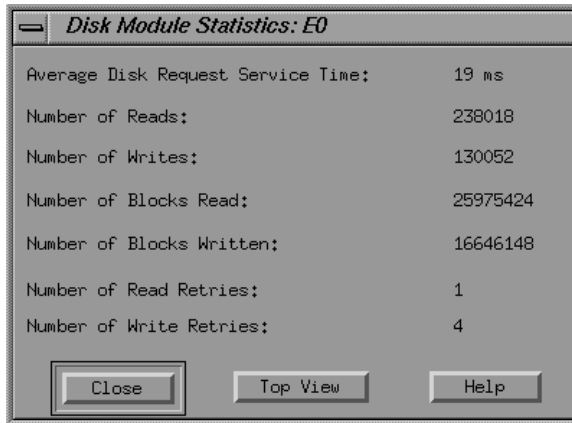
Figure 3-9   Example Disk Module Statistics Window

Table 3-6 explains the entries in this window.

Table 3-6   Disk Module Statistics Window Entries

Entry	Meaning
Average Disk Request Service Time	Average number of milliseconds that the disk module required to execute an I/O request after the request reached the top of the queue.
Number of Reads Number of Writes	Total read and write requests made to the disk module. (For read and write information for the entire LUN, see "Getting LUN Information Using RAIDGUI" in Chapter 4.)
Number of Blocks Read Number of Blocks Written	Total data blocks read from and written to the disk module.
Number of Read Retries Number of Write Retries	Total times read and write requests to the disk module were retried.

To view number of remapped sectors and hard and soft read/write errors, click Errors in the Disk Module Information window. Figure 3-10 shows an example Disk Module Errors window.
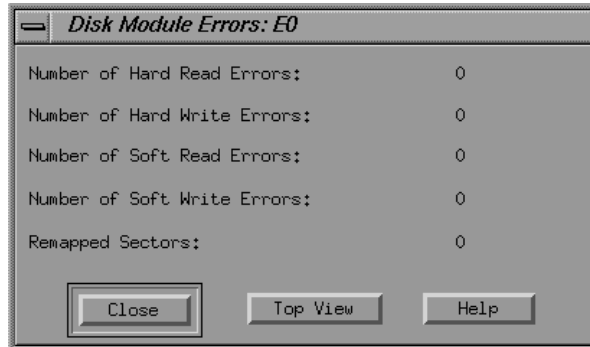
Figure 3-10   Disk Module Errors Window

Table 3-7 explains the entries in the window shown in Figure 3-10.

Table 3-7   Disk Module Errors Window Entries

Entry	Meaning
Number of Hard Read Errors Number of Hard Write Errors	Total read or write errors that persisted through all the retries. An increasing number of hard errors might mean that one or more of the LUN's disk modules is nearing the end of its useful life.
Number of Soft Read Errors Number of Soft Read Errors	Total number of read or write errors that disappeared before all the retries. An increasing number of soft errors might indicate that one of the LUN's disk modules is nearing the end of its useful life.
Remapped Sectors	Total disk sectors that were faulty when written to, and thus were remapped to a different part of the disk module.

Getting Information About System Components

To get details on a specific system component, follow these steps:

1. To display more details on system components, select "Summary View" from the Views menu of the Equipment View. The Summary view for the selected chassis appears, as shown in Figure 3-6.

   "Getting LUN Information Using RAIDGUI" in Chapter 4 explains features of this window in detail.

   • Iconify a Summary View by clicking on the second button from the upper right corner. Each icon bears a diagram of a chassis, as shown in Figure 3-11.
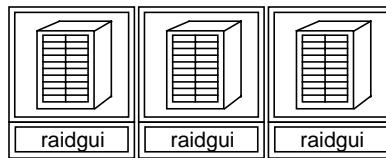
Figure 3-11   Chassis Icons

2. Use the chassis icons for checking component status. Figure 3-11 shows three chassis with normal operation. If the chassis icon is blinking and shows a split chassis, one or more components has failed or has been removed; Figure 3-12 shows such an icon.
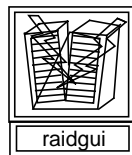
Figure 3-12   Chassis Icon Showing Failed Operation

   Note: To update the information in the icons, restore the view that the icon represents and click the *Poll* button in the toolbar. For more information on polling, see "Using RAIDGUI Automatic Polling," later in this chapter.

3. If the chassis icon is blinking and shows a split chassis, restore the icon and check the service light button color, as explained in "Getting General System Information," earlier in this chapter.

4.   Click the service light button in the toolbar to display the Chassis Status window. Figure 3-13 shows an example.
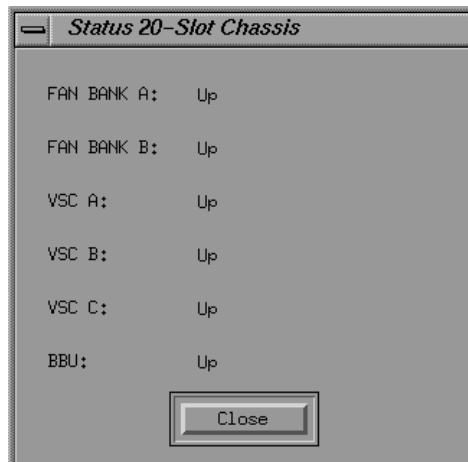
Figure 3-13   Chassis Status Window

Possibilities for each entry are Up, Down (component has failed or was removed after the agent started running), or Not Present (component failed or was removed before the agent started running).

Note: For information on how to identify a specific defective component, see Chapter 6, "Identifying Failed System Components."

Viewing SP Status Information

To view status information for an SP, use the SP Summary window. To
display this window, click the SP's button in the Equipment View or
Summary View. Figure 3-14 shows an example.
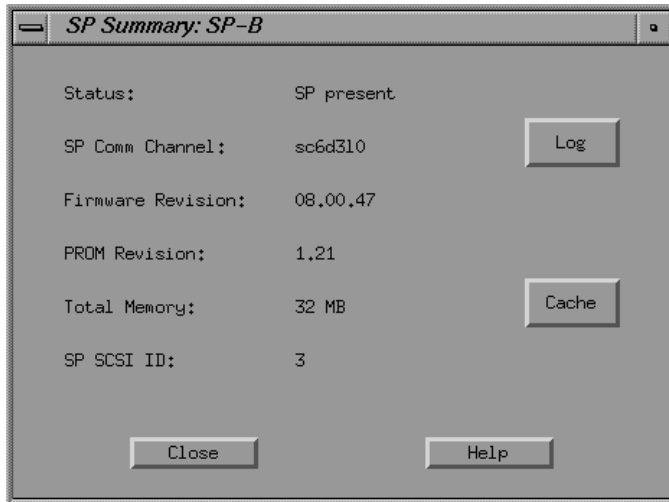
Figure 3-14   Example SP Summary Window

Note: To display cache information, click *Cache*; the window that appears is
explained in "Viewing Cache Statistics" in Chapter 7.

Table 3-8 lists the possibilities for the Status field.

Table 3-8   SP Summary Status Field

Status Entry	Meaning
SP Present	The SP is the communication channel you are using to communicate with the chassis.
SP Not Connected	Agent cannot talk to the SP because a communication channel specifying the SP is not in the agent's configuration file for the selected host. For example, the SP is connected to a host different from that of the SP in the communications channel for the chassis.
SP Not Present	SP that is in the communication channel to the selected chassis has failed or has been removed.
SP Removed	SP was not present when the agent was started.

The SP Comm Channel field contains a SCSI device address that is the same as the device entry in the agent configuration file for the host to which the Challenge RAID storage system is attached. If this communication channel is not through this SP, and the SPs in the Challenge RAID system are connected to different hosts, the word `Unknown` appears in this field, instead of a SCSI device address, and the values in the rest of the fields in this window are either `0` or `Unknown`. For more information on the SP communication channel, see "Shutting Down the Challenge RAID Storage System" later in this chapter.

The Firmware Revision field gives the revision number of the Licensed Internal Code (LIC) that the SP is running. All SPs in the system run the same LIC revision. The PROM Revision field indicates the revision number of the SP's PROM code; all SPs in the system run the same PROM revision. .

The Total Memory field lists the number of megabytes (8, 16, 32, or 64) in the SP's memory. To make full use of cache memory, each SP in the system should have the same amount of memory.

The SP SCSI ID field gives the SP's SCSI ID, which is determined by switch settings on the SP. The Silicon Graphics System Service Engineer sets these when the Challenge RAID storage system is installed.

Viewing the SP Event Log

Each SP maintains a log of event messages. These events include hard errors, startups, and shutdowns involving disk modules, fans, SPs, VSCs, and the BBU. Periodically, the SP writes this log to disk to maintain it when SP power is off. The log can hold over 2000 event messages; when that amount is reached, the oldest messages are deleted in chronological order as new messages come in.

To display the SP log, click *Log* in the SP Summary View. Figure 3-15 shows an example.
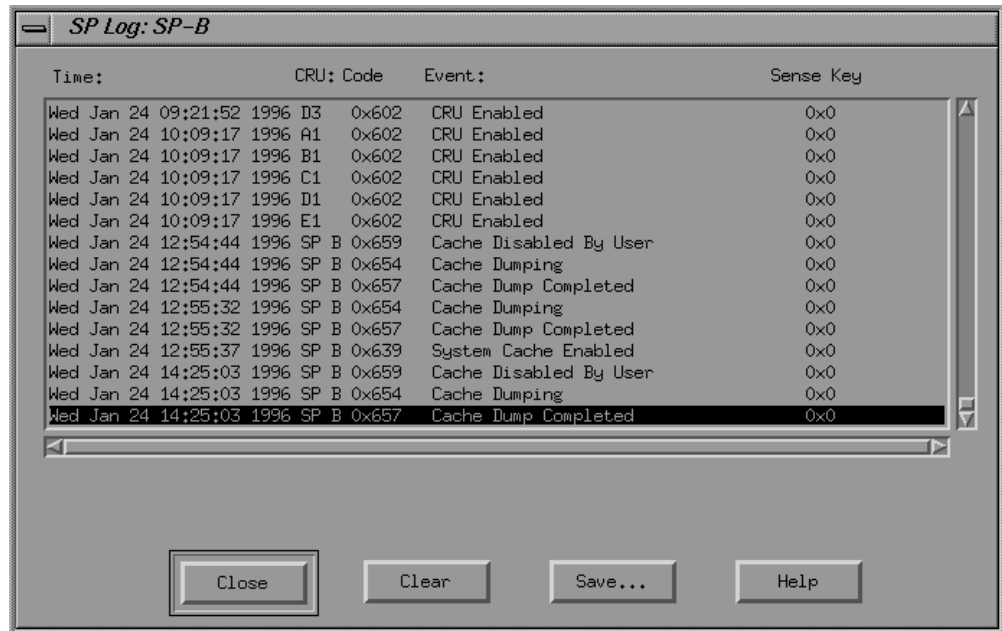
Figure 3-15   Example SP Log

Event messages are in chronological order, with the most recent ones at the end of the log. To display earlier messages, use the scroll bar to move backwards through the log.

Event codes and their corresponding messages are in Appendix C, "Storage-Control Processor Event-Log Error Codes."

Click *Save...* to save the log to a file.

Click *Clear* to clear the contents of the log and reset it.

Caution: Clearing the log can cause problems for other users who are viewing the same log.

To exit the log, click *Close*.

Enabling and Disabling the Statistics Log

The SP maintains a log of statistics for the LUNs, disk modules, and storage system caching, which you can turn on and off.

Because the log uses a 32-bit counter to maintain the statistics numbers, the statistics numbers start over at zero when the counter is full. Thus, you see a sudden decrease in a statistics number if you view it shortly before the counter is full and shortly after the counter restarts at zero. To keep the log turned on for more than two weeks, reset the log about every two weeks, so that you know when the numbers start at zero.

When the log is on, it affects storage system performance; you may want to disable it unless you have a reason to monitor performance. To turn the SP's statistics log on and off, select "Toggle Statistics Logging" in the Options menu of the Equipment View or Summary View. In the submenu, click on the item for the SP whose log you want to turn on or off. If the SP's log is on, a box appears in front of the item for the SP.

To reset (enable) the SP's statistics log, select "Reset Statistics Log" in the Options menu of the Equipment View or Summary View. In the submenu, click on the item for the SP whose log you want to reset.

Using RAIDGUI Automatic Polling

RAIDGUI automatically polls the agent on the selected host to get the
current status, statistics, and event log information for each selected chassis.
In the Poll Setting window, you can change the default polling interval or
turn automatic polling off or on for the current RAIDGUI session.

To poll an agent manually, click the *Poll* button on the Equipment View or
Summary View toolbar.

To change the automatic polling intervals, follow these steps:

1. In the Equipment or Summary View window, select "Poll Setting" from
    the Options menu. The Poll Settings window appears.
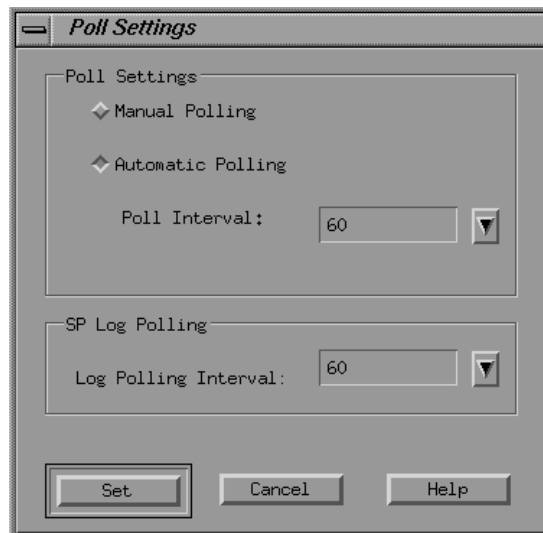
Figure 3-16   Poll Settings

2. To change the poll interval for status and statistics, click the list button
    beside the Poll Interval field, and select the desired number of seconds
    from the list box that appears.

    Available poll intervals are 5, 10, 15, 30, 60, 120, 300, 600, 1800, and 3600
    seconds; the default is 60 seconds.

3. Click *Set*. The new poll interval(s) take effect immediately.

To turn automatic polling off, follow these steps:

1. Select "Poll Setting" in the Options menu of the Equipment View or Summary View.

2. In the Poll Settings window that appears, click *Manual Polling*; click *Set*. Automatic polling is turned off immediately. To update status, statistics, and event log information, you must manually poll the agent.

Note: If automatic polling is turned off, manual polling is useless for updating the event log.

To turn manual polling on, click *Manual Polling* and click *Set* in the Poll Settings window.

Using Alarm Settings

The graphical user interface includes alarm settings that are useful for monitoring the status of system components. Alarm message can be sent to an e-mail address, appear in a window, or both. Figure 3-17 shows the alarm message window.
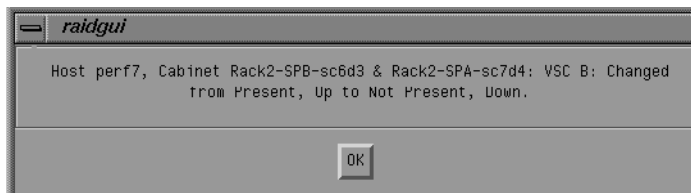
Figure 3-17   Alarm Message Window

To change alarm settings, select "Alarms" in the Options menu; the Alarm Settings window appears, as shown in Figure 3-18.
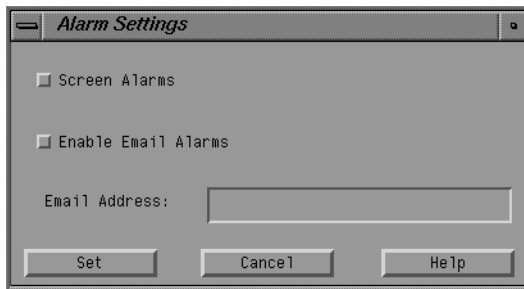
Figure 3-18   Alarm Settings

Use the Alarm Settings window as follows:

- To toggle display of alarm messages on or off, click *Screen Alarms.*

- To start or stop sending alarm messages to an e-mail address, click *Enable Email Alarms.* Click the Email Address field, and enter the address to which the alarms should be sent.

- To enable any of these settings, click *Set.*

Note: Alarm settings are not saved from one session to the next. If you exit RAIDGUI, you must reset alarm settings when you restart it.

Exiting RAIDGUI

To exit the Equipment View or Summary View, select "Exit" from the File menu and click *Exit Program* in the popup window.

Shutting Down the Challenge RAID Storage System

To shut down the Challenge RAID storage system, follow these steps:

1. If you are using storage system caching, make sure that it is disabled; use one of these methods:

    • Use *raidcli getcache* to check status.

    • In RAIDGUI, click an SP icon in the Equipment View; in the SP Summary window that appears, click *Cache*.

2. If necessary, disable it, as explained in Chapter 7.

3. Turn off the power switch on the back of the Challenge RAID storage system, as shown in Figure 3-19.

Note: You do not need to disable the power for the SP(s).
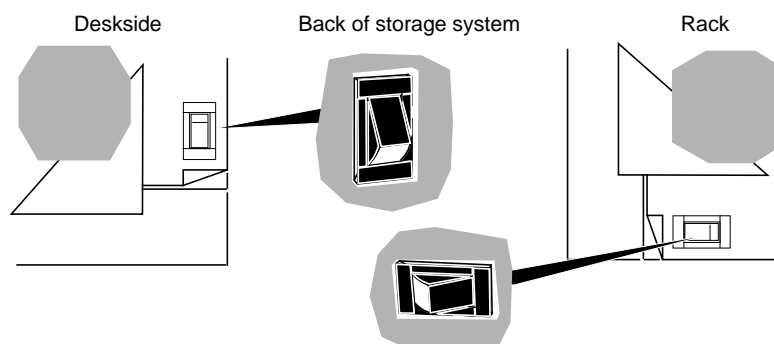
Figure 3-19   Power Button, Back of Challenge RAID Chassis

Restarting the Challenge RAID Storage System

To start the Challenge RAID storage system, follow these steps:

1. Turn on the storage system's power; see Figure 3-19.

    The green power light on the front of the storage system lights up (see Figure 3-20) and the fans rotate.
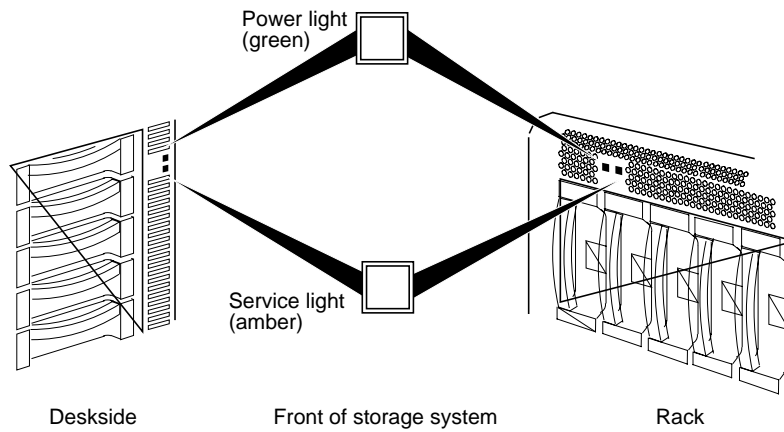
Figure 3-20   Challenge RAID Indicator Lights

2. If none of the busy lights on the drive modules light up, make sure that the power for each SP is enabled. Move the fan module's latch to the UNLOCK position, as indicated in Figure 3-21.

Deskside   Rack
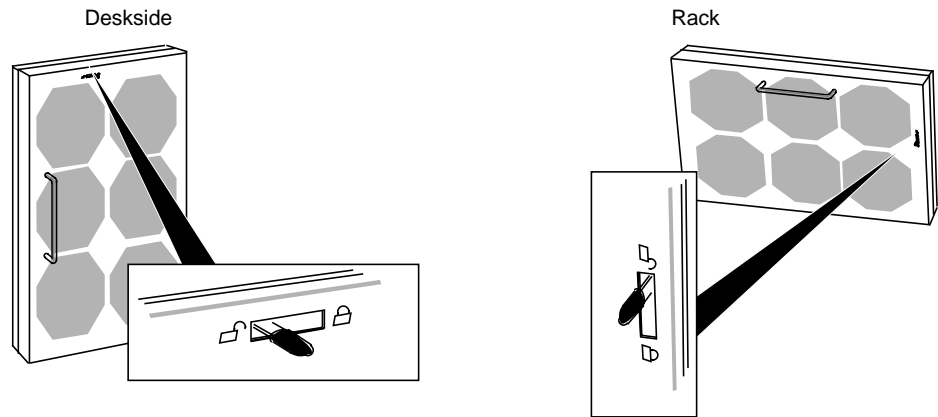
Figure 3-21   Unlocking the Fan Module

3. Swing open the fan module.

   Caution: To prevent thermal shutdown of the system, never leave the fan module open more than two minutes.

71

4. For the AMD®-based SP, move the SP's power switch to the enable position, as shown in Figure 3-22
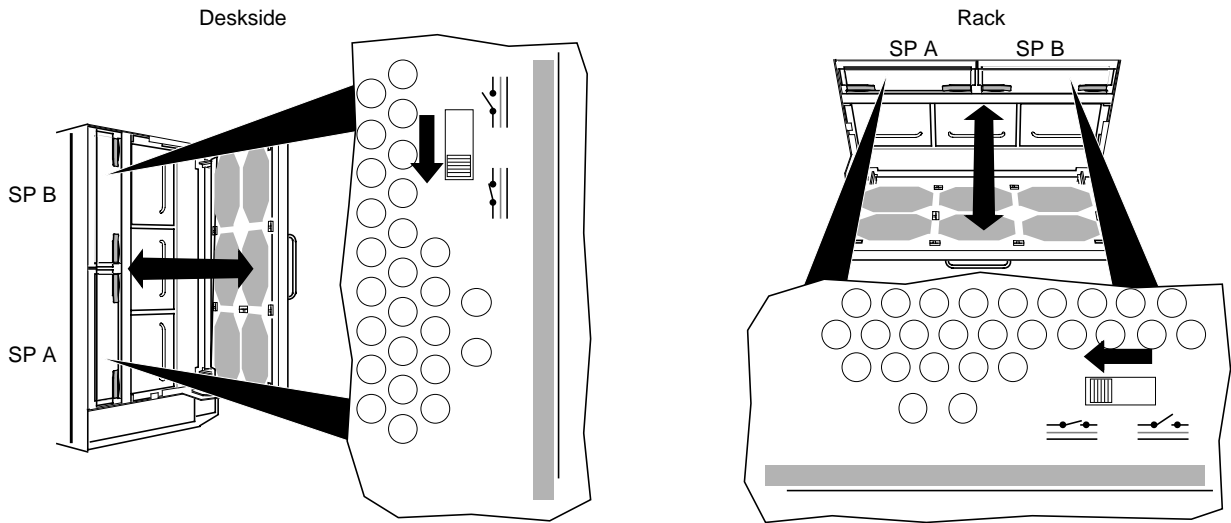
Figure 3-22   Enabling an SP's Power

Note: The PowerPC®-based SP does not have power switches.

5. Close the fan module and move the module's latch to the LOCK position.

Configuring Disks

This chapter describes how to bind disks into LUNs. It explains

- moving disk modules
- planning the bind
- binding disks into RAID units
- getting disk group (LUN) information

The chapter concludes with information on dual processors, load balancing, and device names.

Moving Disk Modules

Generally, disk modules should not be moved from one slot to another slot; but if moving one is absolutely necessary, the system operator or service person can move it with the following cautions:

- The disk module must be unbound. Moving a module that is part of a LUN to another slot may destroy all information on the entire LUN.

- The disk module must be removed and reinserted while the storage system is powered on.

- All the disk modules that you use for a LUN must have the same capacity to use the modules' disk space fully.

- For an individual unit or hot spare, use the next empty slot on an internal bus that is already used, provided it is not one of the slots for a database or a cache vault disk module (slots A0, B0, C0, D0, E0, or A3). For example, to add an individual unit when slots A0, B0, C0, D0, E0, A1, and B1 are used, use slot A2 or B2.

- Follow instructions in "Replacing a Disk Module" in Chapter 5.

73

Planning the Bind

This section discusses

- filesystem capacity considerations for RAID
- determining RAID levels

Filesystem Capacity Considerations for RAID

You can bind as many as 16 drives in a LUN; for 4 GB drives, this capacity totals 64 GB. However, filesystems have capacity limits that must be taken into account.

- For IRIS 5.3 and earlier, an EFS filesystem is a maximum of 8 GB.
- For IRIS 5.3 and earlier, an XFS filesystem can be over a terabyte; however, the device parameter portion of the volume header is limited. Because values cannot exceed the number of bits allocated for each of the fields in the volume header (*sec/trk*, *trks/cyl*, *#cyls*), decrease the *#cyls* to within the 16-bit value available and increase one of the other fields.

To determine allowed capacity limits for an IRIX release, view the header files *dvh.h* and *dksc.h* in */usr/include/sys*.

Determining RAID Levels

Before you create LUNs, plan the RAID levels you need. The LUNs are bound in a particular order, depending on the RAID level for each one:

- first: RAID-1_0
- next: RAID-0, RAID-3, or RAID-5
- next: RAID-1
- last: individual disk units and hot spares

Note the rules outlined in Table 4-1 for binding disk modules into LUNs.

Table 4-1   Binding LUNs

LUN to Bind	Restrictions and Recommendations
Any LUN	You can bind only unbound disk modules. All disk modules in a LUN must have the same capacity to use the modules' storage space fully.
	For higher availability with any redundant LUN, use slots on separate internal buses. For example, for a RAID-1 LUN, use slots A2 and B2 instead of A2 and A3.
RAID-0	You must bind a minimum of three disk modules, and no more than sixteen disk modules. If possible, use modules on different SCSI buses for highest availability. You can select the modules in any order.
RAID-1	You must bind exactly two disk modules, which you can select in any order.
RAID-1_0	You must bind an even number of modules: a minimum of four and a maximum of sixteen. RAIDGUI pairs modules into mirrored images in the order in which you select them. The first and second modules you select are a pair of mirrored images, the third and fourth are another pair of mirrored images, and so on. The first module you select in each pair is the primary image, and the second module is the secondary image. If possible, the modules you select for a pair should be on different SCSI buses for highest availability.
RAID-3	You must bind exactly five disk modules, selected in any order. If possible, use modules on separate SCSI buses for highest availability (for example, A2, B2, C2, D2, E2.
RAID-5	You can select the modules for the LUN in any order. You must bind a minimum of three disk modules and no more than sixteen disk modules. For most efficient use of space, bind five modules. If possible, use modules on different SCSI buses for highest availability.
Individual disk	None
Hot spare	You cannot bind the following disk modules as hot spares: A0 through E0 or A3.
	The capacity of a disk module bound as a hot spare must be at least as great as the capacity of the largest disk module that it might replace.

Binding Disks Into RAID Units

The physical disk unit number is also known as the *logical unit number*, or *LUN*. (The unit is a logical concept, but is recognized as a *physical disk unit* by the operating system; hence, the seemingly contradictory names.) The LUN is a hexadecimal number between 0 and 7.

Unlike standard disks, physical disk unit numbers (LUNs) lack a standard geometry. Disk capacity is not a fixed quantity between disk-array LUNs. The effective geometry of a disk-array LUN depends on the type of physical disks in the array and the number of physical disks in the LUN.

This section explains

- binding disks using the command-line interface
- binding disks using RAIDGUI

Binding Disks Using the Command-Line Interface

Storage systems typically ship from the factory with disk modules bound in groups of five as RAID-5 units. Binding disks into different LUNs consists of these steps:

- unbinding LUNs
- creating LUNs
- verifying the LUNs

Unbinding LUNs in a New Challenge RAID Storage System

When you change a physical disk configuration, you change the bound configuration of a physical disk unit. The physical disk unit configuration changes when you add or remove a disk module, or physically move one or more disk modules to different slots in the chassis.

Unbinding a LUN divides it into its constituent disk modules, allowing you to rebind the modules in any valid configuration you choose.

Unlike binding, unbinding takes only a few moments. You can unbind a LUN only if it is broken or bound, that is, its modules must be in the Enabled or Broken state. You can unbind a hot spare only if it is on standby.

To unbind a disk, use the *unbind* parameter with the *raidcli* command. Follow these steps:

1. In an IRIX window, use *raidcli getagent* to get the device name (node number):

   raidcli getagent

2. If necessary, use *raidcli getdisk* to verify the disk positions.

3. Use *raidcli unbind* to unbind the individual disk units:

   raidcli -d device unbind lun-number [-o]

   In this syntax, *device* is the device name as returned by *getagent* and *lun-number* is the number of the logical unit to unbind. The **-o** flag specifies that the user is not prompted for permission.

   The *unbind* parameter has no output.

The following example destroys LUN 3 without prompting for permission:

```
raidcli -d sc4d2l0 unbind 3 -o
```

For more information on the *unbind* parameter, see "unbind" in Appendix B.

Binding Disks

To group physical disks into RAID-0, RAID-1, RAID-1_0, RAID-3, or RAID-5 units or to create a hot spare, become superuser and use

```
raidcli -d device bind raid-type lun-number disk-names [optional-args]
```

77

Variables in this syntax are explained below.

-d *device*   Target RAID device, as returned by raidcli getagent; see "Getting Device Names With getagent," in Chapter 3.

raid-type   Choices are

- r0: RAID-0

- r1: RAID-1

- r1_0: RAID-1_0

- r3: RAID-3

- r5: RAID-5

- id: individual disk unit

- hs: hot spare

lun-number   Logical unit number to assign the unit (a hexadecimal number between 0 and 7).

disk-names   Indicates which physical disks to bind, in the format *bd*, where *b* is the physical bus name (a through e; be sure to use lowercase) and *d* is the device number on the bus (0 through 3). For example, a0 represents the device 0 on bus A, and e2 represents device 2 on bus E.

A RAID-0 bind requires a minimum of three and a maximum of 16 disks.

A RAID-1 bind requires two disks on separate buses.

A RAID-1_0 bind requires a minimum of four disks per bus. The maximum is 16 disks, grouped in sets of four on different buses. A RAID-1_0 bind requires separate buses for each member of the image pair. Select the disks in this order: first disk on first bus, first disk on second bus, second disk on first bus, second disk on second bus, third disk on first bus, third disk on second bus, and so on.

A RAID-5 bind also requires separate buses, each with five disk modules. Legal RAID-5 bind configurations are

- a0 b0 c0 d0 e0

- a1 b1 c1 d1 e1

- a2 b2 c2 d2 e2

- a3 b3 c3 d3 e3

A hot spare bind requires one disk, which cannot be A0, B0, C0, D0, E0, or A3. The capacity of the hot spare must be at least as great as the capacity of the largest disk module it might replace.

All disk modules in a bind should have the same capacity to use the modules' disk space fully.

The optional arguments are as follows:

Note: In the command-line interface, auto assign is enabled by default. For most uses (including high availability), disable it by using RAIDGUI, as explained in step 8 in "Creating LUNs," later in this chapter.

-r *rebuild-time*   Maximum time in hours to rebuild a replacement disk. Default is four hours; legal values are any number greater than or equal to 0.

A rebuild time of two hours rebuilds the disk more quickly but degrades response time slightly. A rebuild time of zero hours rebuilds as quickly as possible but degrades performance significantly.

If your site requires fast response time and you want to minimize degradation to normal I/O activity, you can extend the rebuilding process over a longer period of time, such as 24 hours. You can change the rebuild time later without damaging the information stored on the physical disk unit.

Caution: The LUN is not fault-tolerant during the rebuild; selecting a short rebuild time minimizes the chance of data loss.

79

-s *stripe-size*   Number of blocks per physical disk in a RAID stripe. Default is 128; legal values are any number greater than zero.

                            The smaller the stripe element size, the more efficient the distribution of data read or written. However, if the stripe size is too small for a single host I/O operation, the operation requires accessing two stripes, thus causing the hardware to read and/or write from two disk modules instead of one. Generally, it is best to use the smallest stripe element size that will rarely force access to another stripe. The default stripe element size is 128 sectors. The size should be an even multiple of 16 sectors (8 KB).

                            Note: This value cannot be used with RAID-3 binds; also, it is ignored for hot spares.

-c *cache-flags*   Values are

- none: no caching

- read: read caching only

- write: write caching only

- rw: read and write caching

                            The default is none.

                            Note: For more information on caching, see Chapter 7, "Caching."

Note: Although *bind* returns immediate status for a RAID device, the bind itself does not complete for 15 to 60 minutes, depending on system traffic. Use *raidcli -d device getlun 0 | grep Bound* to monitor the progress of the bind. When the bind is complete, each disk is noted as "Bound But Not Assigned" in the *getlun* output.

The following example binds disks A1, B1, C1, D1, and E1 into a RAID-5 logical unit with a logical unit number of 3, a four-hour maximum rebuild time, and a 128-block stripe size per physical disk, with read cache enabled:

```
raidcli -d sc4d2l0 bind r5 3 a1 b1 c1 d1 e1 -r 4 -s 128 -c
read
```

The following example binds A2 and B2 into a RAID-1 logical unit with a LUN number of 2 and a four-hour maximum rebuild time, with read cache enabled:

```
raidcli -d sc4d2l0 bind r1 2 a2 b2 -r 4 -c read
```

The following example binds A1, B1, C1, D1, and E1 into a RAID-3 logical unit with a LUN number of 3.

```
raidcli -d sc4d2l0 bind r3 3 a1 b1 c1 d1 e1 -r 4 -c read
```

The following example binds disks A1, B1, C1, and D1 into a RAID-1_0 logical unit with a LUN number of 1, a four-hour maximum rebuild time, and a 128-block stripe size per physical disk, with read cache enabled:

```
raidcli -d sc4d2l0 bind r1_0 1 a1 b1 c1 d1 -r 4 -s 128 -c read
```

The following example binds A2, B2, C2, D2, and E2 into a RAID-0 logical unit with a LUN number of 3, and a 128-block stripe size per physical disk, with read cache enabled:

```
raidcli -d sc4d2l0 bind r0 3 a2 b2 c2 d2 e2 -s 128 -c read
```

The following example binds disk E3 as a hot spare with a LUN number of 7:

```
raidcli -d sc4d2l0 bind hs 7 e3
```

There is no output for *raidcli* with the *bind* parameter. Errors are printed to *stderr*.

Note: For complete messages, use the -*v* option.

Verifying LUNs

Verify that the LUNs are the way you want them; see "Getting Disk Group (LUN) Information" later in this chapter.

Binding Disks Using RAIDGUI

Binding disks consists of these steps:

- unbinding LUNs as necessary
- creating LUNs

Unbinding LUNs in a New Challenge RAID Storage System

Storage systems typically ship from the factory with disks bound in groups of five as RAID-5 units. To change this configuration, you must unbind LUNs to create the LUNs you want.

Note: Creating a LUN usually takes 15 to 60 minutes, depending on system traffic.

To use RAIDGUI to unbind disks, follow these steps:

1. Open the Summary View window for the Challenge RAID storage system. Figure 4-1 shows an example.
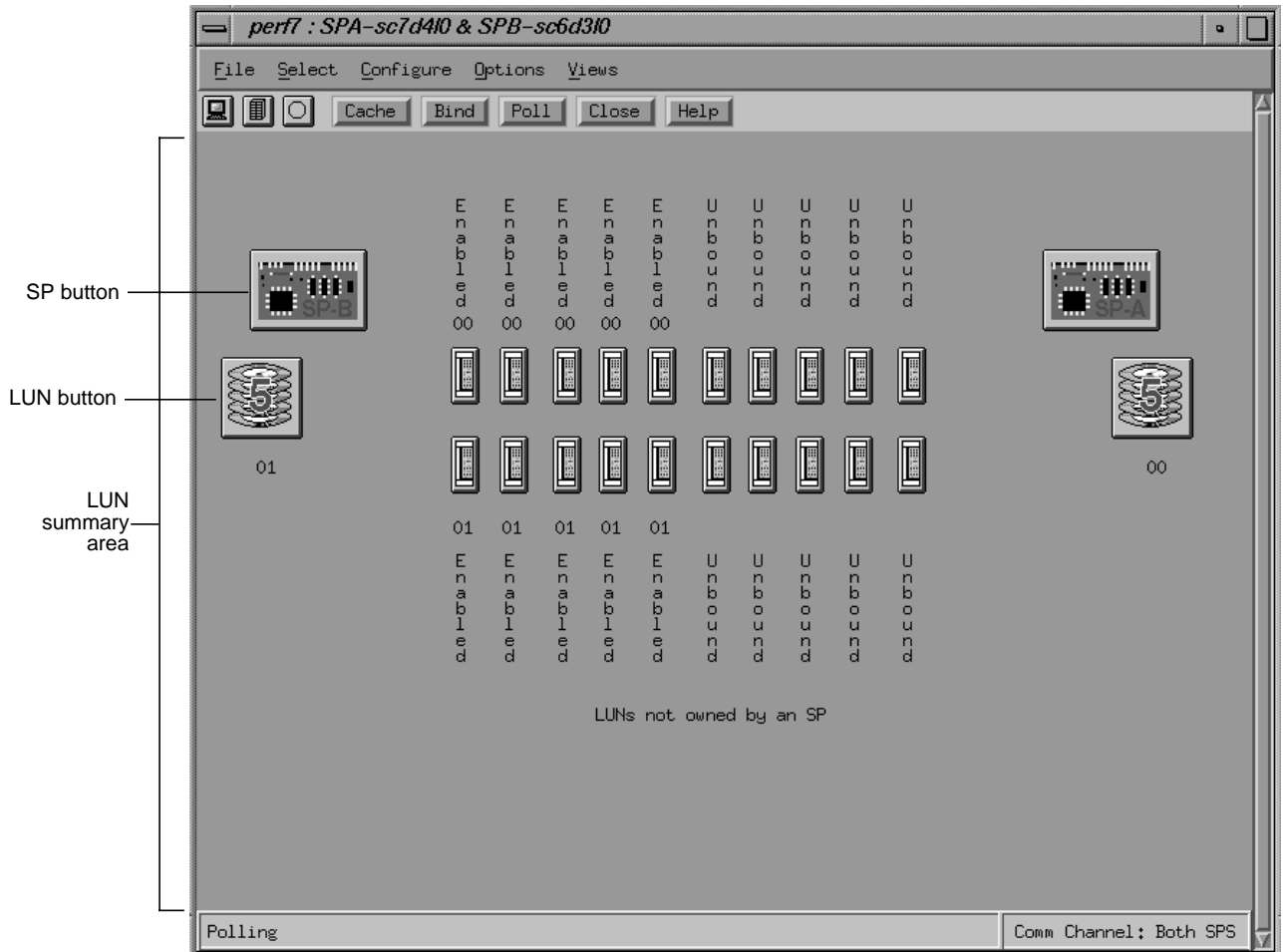
Figure 4-1   Summary View

2. Click the button of a LUN you want to unbind. The LUN information window appears, as shown in Figure 4-2.
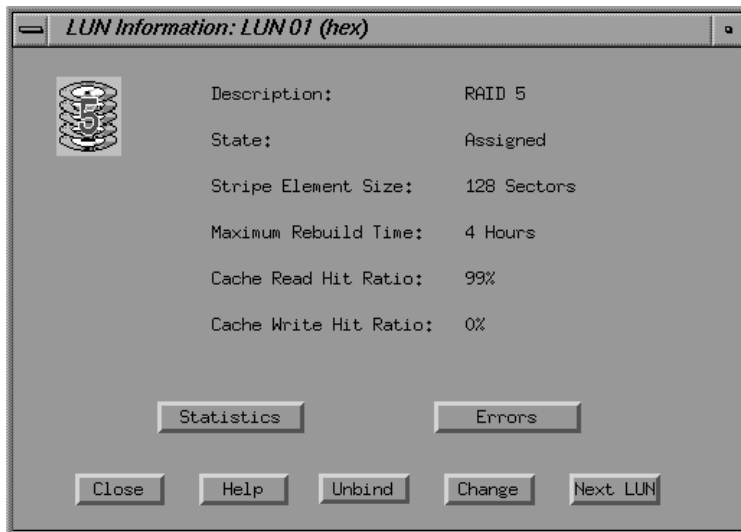
Figure 4-2   LUN Information Window

3. Click *Unbind*. Verification dialog boxes appear; answer them appropriately to unbind the individual disk unit.

    The SP unbinds the LUN. In the Bind window, the black outline disappears from around the disk module buttons, and select buttons appear beside the disk module buttons for the newly unbound disk modules. In the Summary View window, the state of the newly unbound disk modules changes from Enabled to Unbound, the LUN number disappears from beside the disk module buttons, and the button for the LUN that was unbound disappears.

Creating LUNs

To use RAIDGUI to create LUNs, follow these steps:

1. To speed up the binding operation, turn off automatic polling; see "Using RAIDGUI Automatic Polling" in Chapter 3. If you want status information during binding, poll for it manually by clicking *Poll* in the Equipment or Summary View toolbar.

2. Click *Bind* in the Equipment View or Summary View. The Bind window for that chassis appears; Figure 4-3 shows an example.
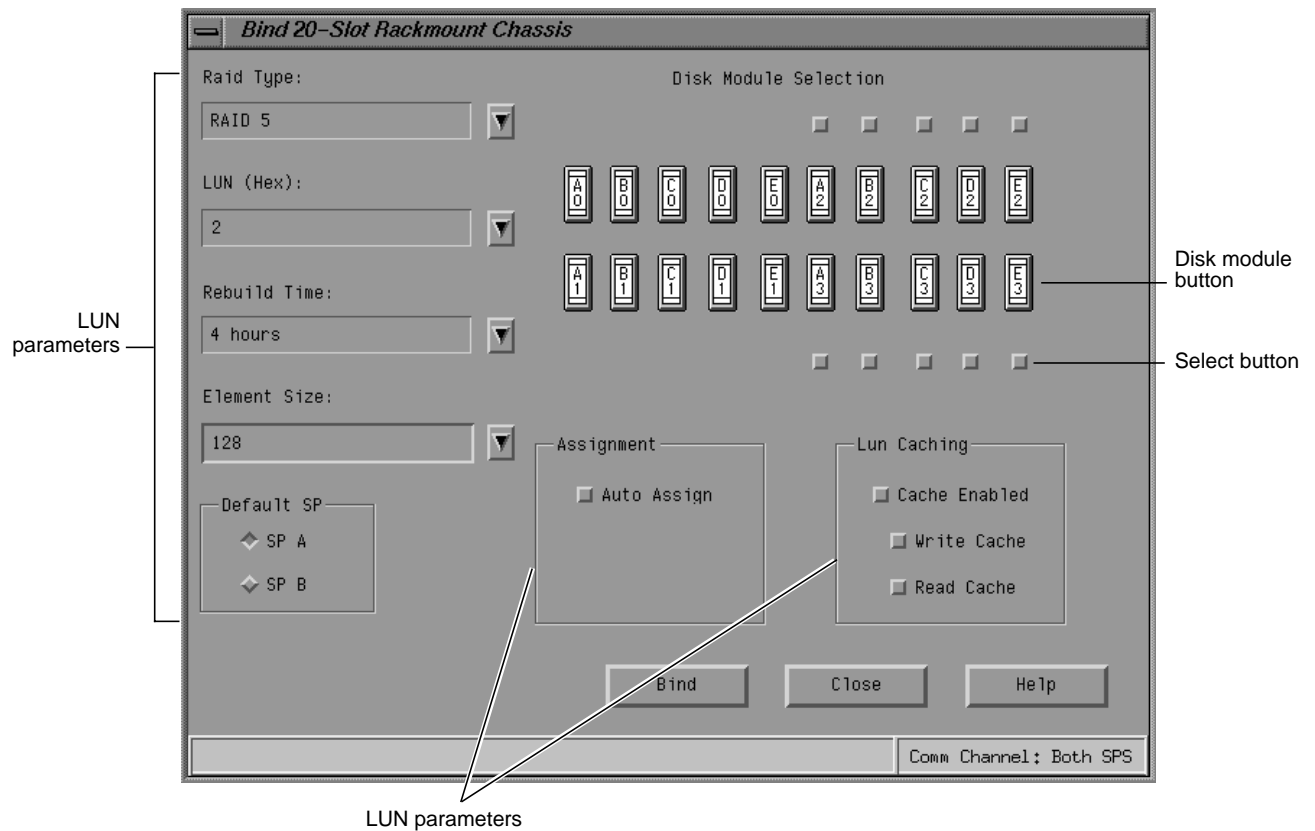
Figure 4-3   Bind Window

You can also open this window by selecting "Bind" in the Equipment View or Summary View's Configure menu.

The disk module buttons in the Bind window resemble those in the Summary View window. You can click on a disk module button to display information on the disk module.

A disk module button for an unbound disk has a select button next to it, as shown in Figure 4-3.

The parameters for the LUN that you can select in the Bind window vary, depending on the RAID level. Table 4-2 summarizes the LUN parameters that are available to the various RAID levels.

Table 4-2   RAID Levels and LUN Parameters

RAID Level	LUN	Rebuild Time	Element Size	Default SP	Auto Assign	Write Caching	Read Caching
RAID-0	x		x	x	x	Acceptable	Recommended
RAID-1	x	x		x	x	Acceptable	Recommended
RAID-1_0	x	x	x	x	x	Acceptable	Recommended
RAID-3	x	x		x	x	Highly recommended	Recommended
RAID-5	x	x	x	x	x	Highly recommended	Recommended
Individual disk	x			x	x	Acceptable	Recommended
Hot spare	x					N/A	N/A

3. In the RAID Type text window, select the RAID level for the first LUN you want to create. If you plan to use different RAID levels for the LUNs you create, follow the order required in "Planning the Bind," earlier in this chapter.

   Note: Fast RAID-3 is not available.

4. In the LUN (Hex) text window, select the LUN number you want for the first LUN.

   The default number is 0 for the first LUN that you bind, regardless of the number of SPs or hosts attached to the Challenge RAID storage system. The default number for the second LUN you bind is 1; for the third LUN, it is 2; for the fourth LUN, it is 3, and so on. You can select nondefault numbers if desired.

The Challenge RAID storage system allows a maximum of eight LUNs (0-7).

Note: For hot spares, assign LUN numbers starting with the highest number available and continue downward.

All disk modules in a bind should have the same capacity to use the modules' disk space fully.

5. If the LUN you are creating is RAID-1, -1_0, -3, or -5, use the Rebuild text window to select a different rebuild time from the default 4 hours, if desired.

   The rebuild time is the amount of time that the storage system allots to reconstruct the data on either a hot spare or a new disk module that replaces a failed disk module in a LUN. It applies to all RAID LUNs except RAID-0. The time you specify determines the amount of resource the SP devotes to rebuilding instead of to normal I/O activity.

   The default time of 4 hours is adequate for most situations. A rebuild time of 2 hours rebuilds the disk more quickly, but degrades response time slightly. A rebuild time of 0 hours rebuilds the disk module as quickly as possible, but degrades response time significantly. If your site requires fast response time and you want to minimize degradation to normal I/O activity, you can extend the rebuilding process over a longer period of time, such as 24 hours.

   Note: The actual rebuild time can differ significantly from the time you specify, especially for a RAID-1_0 LUN, or a LUN containing 4 GB disk modules. Since a RAID-1_0 with *n* disk modules can continue functioning with up to as many as *n/2* failed drive modules and only one drive at a time is rebuilt, the actual rebuild time for such a LUN is the time you specify multiplied by the number of failed drives. For 4 GB disk modules, the minimum rebuild time should be 2 hours, because a rebuild time of less than 2 hours actually takes about 2 to 3 hours. If you specify any other time, the actual rebuild time for the disk module is about twice as long.

6. If the LUN you are creating is RAID-0, 1_0, or -5, use the Element text window to select a different number of sectors for the stripe element for the default (128 sectors).

   The stripe element size is the number of disk sectors that the storage system can read or write to a single disk module without requiring access to another disk module (assuming that the transfer starts at the

first sector in the stripe). The stripe element size can affect the performance of a RAID-5 or RAID-1_0 LUN. A RAID-3 LUN has a fixed stripe element size of one sector, a value that you cannot change.

The smaller the stripe element size, the more efficient the distribution of data read or written. However, if the stripe size is too small for a single I/O operation, the operation requires access to two stripes, which causes the hardware to read or write from two disk modules instead of one. Generally, you use the smallest stripe element size that rarely forces access to another stripe.

7. For the LUN to have a different default SP owner than that shown in the Default SP text window, select the diamond-shaped button for the other SP.

The default SP assumes ownership of the LUN after the storage system's power is cycled. If the storage system has two SPs, you can choose to bind some LUNs using one SP, and the rest using the other SP. Splitting the LUNs can balance the load across the SPs and establish the primary server for the LUN in dual-server configuration. The primary route to a LUN is the route through the default SP, and the secondary route is through the other SP.

8. Disable auto assign by unselecting the Auto Assign select button in the Assignment field. Disabling auto assign is recommended for most uses, particularly if failover software is present.

Auto assign controls the ownership of the LUN when an SP fails in a storage system with two SPs. With auto assign enabled, if the SP that owns a LUN fails and the server tries to access that LUN through the second SP, the second SP assumes ownership of the LUN so the access can occur. The second SP continues to own the LUN until the SP's power is cycled (turned off and on again). When the power is cycled, ownership of each LUN returns to its default SP. If auto assign is disabled in the previous situation, the other SP does not assume ownership of the LUN, so the access to the LUN does not occur.

For a LUN to be used in a high-availability configuration, auto assign must be disabled. By default, auto assign is disabled in RAIDGUI.

9. For any LUN except a hot spare, click the button at Cache Enable in the LUN Caching field to enable or disable caching. If the caching is enabled, click the button at Write Caching or Read Caching in the LUN Caching field to enable or disable write or read caching.

88

Whether or not you should use caching for a specific LUN depends on the LUN's RAID type. See Table 4-2 for recommendations for using caching for the different RAID levels.

Note: "Setting Cache Parameters" in Chapter 7 explains how to set other caching parameters.

10. Once you have selected all the parameters for the LUN, select each disk module for the LUN by clicking on the select button that appears beside each disk module button. Follow guidelines in Table 4-1 in "Planning the Bind," earlier in this chapter.

Clicking on this select button selects the disk module; a heavy black border appears around the disk module button to show that it is selected. To select more than one disk module, hold down a <Shift> key and click the disk module's select button. To deselect a disk module button, click the disk module's select button again.

Note: If you are creating a RAID-1_0 LUN, the order in which you select the modules is important. The first module you select is the first part of the mirror's primary image, the second module is the first part of the mirror's secondary image, the third module is the second part of the primary image, the fourth module is the second part of the secondary image, and so on, for any other modules in the LUN. In other words, the first and second modules selected are a pair of peer image parts, the third and fourth modules are another pair of peer image parts, and so on. Figure 4-4 diagrams this scheme.

Primary image	Secondary image
Peer image 1. First module selected	Peer image 1. Second module selected
Peer image 2. Third module selected	Peer image 2. Fourth module selected
Peer image 3. Fifth module selected	Peer image 3. Sixth module selected
Peer image 4. Seventh module selected	Peer image 4. Eighth module selected
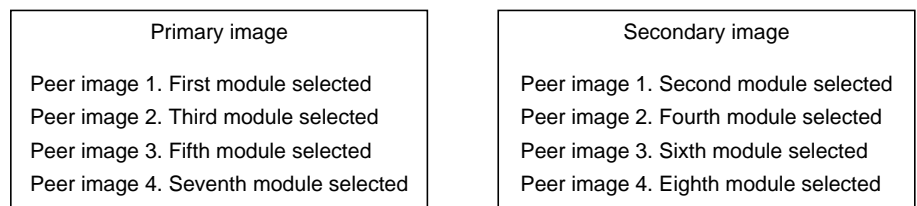
Figure 4-4   Disk Module Selection for RAID-1_0 Mirror Image

The RAIDGUI Bind window displays a different color border for each pair of disk modules you select. The first and second modules you select have a red border, the third and fourth a blue border, and so on.

89

Deselecting one disk module in a pair causes the other module in the pair to be deselected, so that both are no longer paired and are available.

For high availability, the modules in a pair must be on different SCSI buses. For highest availability and performance in a RAID-1_0 LUN, select modules on consecutive SCSI buses. For example, for a six-module LUN, select these modules in order: A0, B0, C0, D0, E0, and A1. Modules A0 and B0 are peers, C0 and D0 are peers, and E0 and A1 are peers.

11. When you have selected all disk modules for the LUN and the LUN parameters specify the values you want, click *Bind*.

    If you selected less than the minimum number of disk modules required for the LUN's RAID type, a popup appears stating the minimum number of modules required. Click *OK* in the popup and continue to select the additional modules needed. See Table 4-1 in "Planning the Bind," earlier in this chapter.

12. If you selected at least the minimum number of disk modules required for the RAID level you are binding, verification dialog boxes appear; answer them appropriately to bind the disk modules into the LUN.

    The bind operation begins, and the borders around the buttons of the disk modules in the LUN disappear, along with their select buttons. In the Summary View window, the state of each disk module in the LUN changes from Unbound to Binding. Generally, binding takes about 15 to 60 minutes for 2 GB disk modules, depending on system traffic. When the bind operation is completed, the state of each disk module in the LUN changes to Ready.

13. Repeat these steps for the remaining LUNs you want to create.

14. On the server connected to the Challenge RAID storage system assign the LUNs you just bound to an SP, perform an I/O request.

    When a LUN is assigned to an SP, its state becomes Assigned, and the state of its disk modules becomes Enabled when you use the SP that owns the LUN as the communications path to the chassis.

    Note: For more information on caching, see Chapter 7, "Caching."

15. Verify that the LUNs are the way you want them; see "Getting Disk Group (LUN) Information" later in this chapter.

Getting Disk Group (LUN) Information

This section explains

- getting LUN information using the command-line interface
- getting LUN information using RAIDGUI

Getting LUN Information Using the Command-Line Interface

To display information on a logical unit and the components in it, use the *getlun* parameter:

```
raidcli getlun lun-number
```

The following example displays information about LUN 3:

```
raidcli -d sc4d2l0 getlun 3
```

For a sample output, see "getlun" in Appendix B.

Note: Information on individual disks is not displayed unless statistics logging is enabled with *raidcli getcontrol*. See "Getting Information About Other Components" in Chapter 3 of this guide.

Getting LUN Information Using RAIDGUI

This section explains

- using the Summary View for LUN information
- using the LUN information window
- getting LUN read/write statistics

Using the Summary View for LUN Information

Use the Summary View to determine LUN information and verify that
LUNs you create are bound as you want them. Figure 4-5 shows an example
Summary View with LUNs.
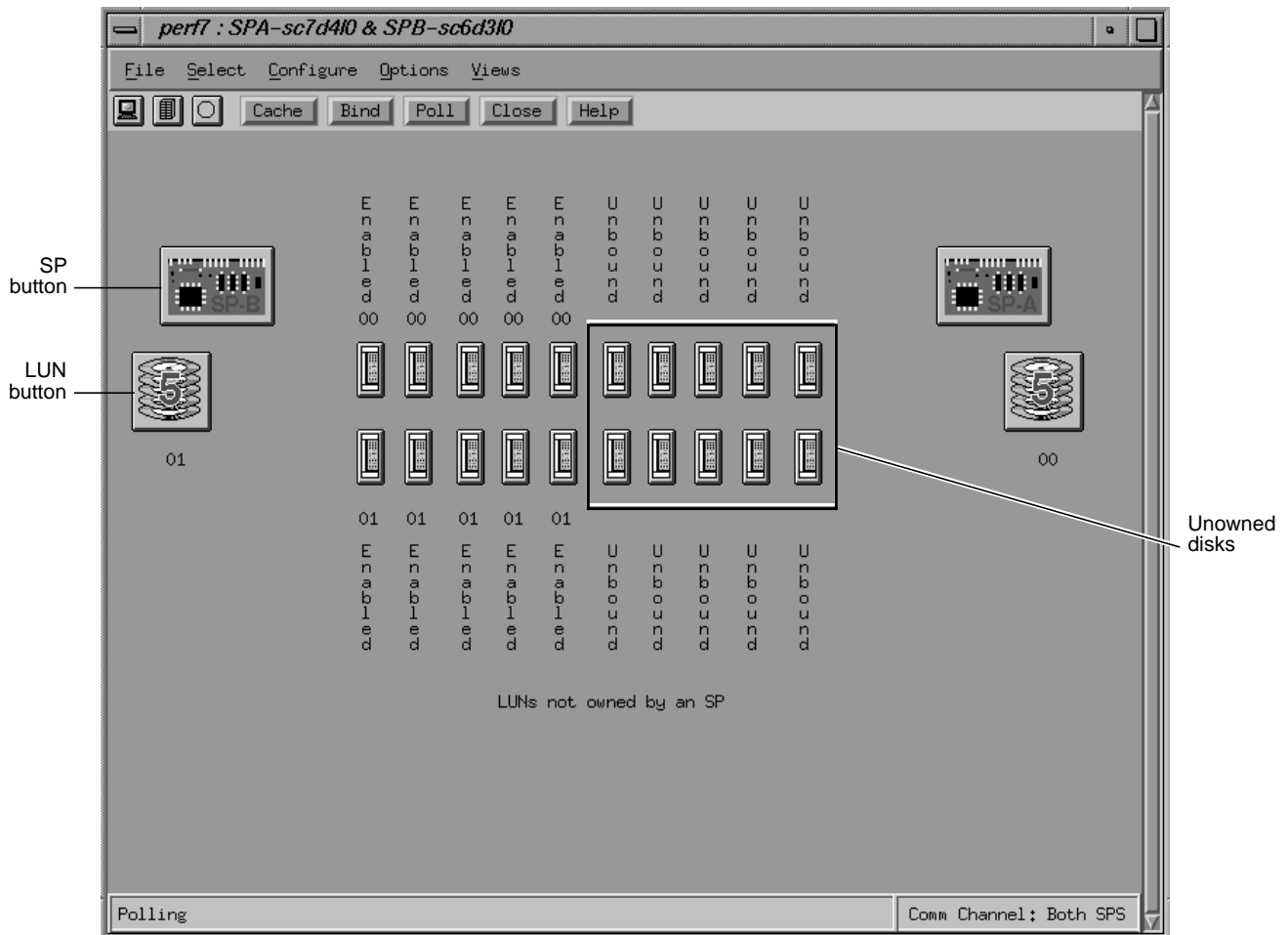
Figure 4-5   Example Summary View

A LUN button tells you the following information:

•   hexadecimal number

•   RAID level, as shown in Figure 4-6

RAID-0   RAID-1   RAID-1_0   RAID-3   RAID-5   Individual disk   Hot spare

Figure 4-6   RAID Levels on LUN Buttons

•   status, as indicated by the color of the button:

    –   gray: operating normally

    –   blue: binding, rebuilding, or, in the case of a hot spare, being used
        to replace a failed disk module

    –   amber: failed

    –   white: disk module is not present

Make sure that the disk module buttons surrounded by black borders
represent the disk modules that you want in the LUN. If any of these disk
modules is a module that you do not want in the LUN or you want
additional disk modules in the LUN, unbind the LUN (see "Unbinding
LUNs in a New Challenge RAID Storage System," earlier in this chapter)
and then rebind it with the correct disk modules.

Using the LUN Information Window

Clicking a LUN button displays the LUN Information window; Figure 4-7 shows an example.
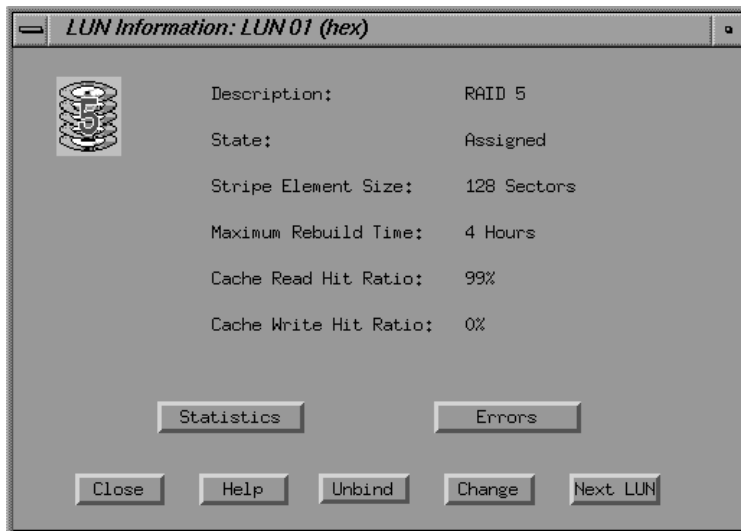
Figure 4-7   Example LUN Information Window

Verify that the fields in the LUN Information window are as you want them.

In this window, the Description field indicates the RAID type (level), as shown in Table 4-3.

Table 4-3   RAID Types in the LUN Information Window

RAID Type	Explanation
RAID-0	Nonredundant individual access array
RAID-1	Mirrored pair
RAID-1_0	Mirrored RAID-0 group
RAID-3	Parallel access array
RAID-5	Individual access array
DISK	Individual disk unit
HOT SPARE	Hot spare

The State field indicates the LUN's operational state. Table 4-4 lists possibilities.

Table 4-4   LUN States

LUN State	Meaning
Assigned	Owned by an SP and all its disk modules are enabled
Binding	Being bound
Bound Not Assigned	Bound but not assigned to either SP
Failed	Shut down by SP because a component is broken
Rebuilding	One or more of its disk modules is rebuilding or equalizing
Replacing XY	Hot spare replacing failed disk module with disk ID XY in another LUN
Standby	Hot spare available for replacing a failed disk module in another LUN

Other fields are as follows:

- The Stripe Element Size field indicates the number of sectors that the Challenge RAID storage system can read or write to a single disk module without requiring access to another disk module in the LUN. The default size is 128 sectors. This size was specified when the LUN was bound. Stripe element size does not apply to a RAID-1 LUN, individual unit, or hot spare.

- The Maximum Rebuild Time field shows the maximum number of hours the storage system can take to integrate a hot spare or new disk module. This time was specified when the LUN was bound.

- The Cache Read Hit Ratio and Cache Write Hit Ratio fields show the percentage of cache read hits and cache write hits for the LUN. For more information on these parameters, see "Viewing Cache Statistics" in Chapter 7.

To change the Description (RAID type) or Element Size parameters, click *Unbind* to unbind the LUN (see instructions in "Unbinding LUNs in a New Challenge RAID Storage System," earlier in this chapter) and then rebind it with correct parameters.

To change caching parameters or the maximum rebuild time, click *Change* in the LUN Information window, as explained in "Changing LUN Parameters" in Chapter 8.

Getting LUN Read/Write Statistics

Click *Statistics* in the LUN Information window to view the read and write statistics for the LUN compiled since the last time the statistics log for the SP that owns the LUN was turned on (see "Enabling and Disabling the Statistics Log" in Chapter 3 for more information). Figure 4-8 shows an example LUN Statistics window.
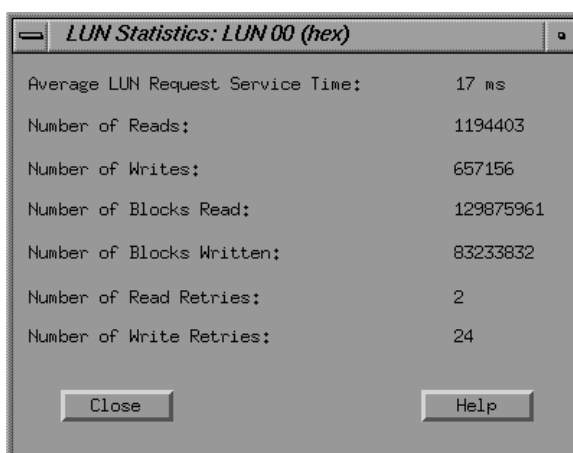
Figure 4-8   Example LUN Statistics Window

Table 4-5 explains the entries in this window.

Table 4-5   LUN Statistics Window Entries

Entry	Meaning
Average LUN Request Service Time	Average number of milliseconds that all the disk modules in the LUN required to execute an I/O request after the request reached the top of the queue
Number of Reads Number of Writes	Total read and write requests made to all the disk modules in the LUN
Number of Blocks Read Number of Blocks Written	Total data blocks read from and written to all the disk modules in the LUN
Number of Read Retries Number of Write Retries	Total times read and write requests to all the disk modules in the LUN were retried

Click *Errors* in the LUN Information window to view the number of remapped sectors and the number of hard and soft read and write errors. Figure 4-9 shows an example.
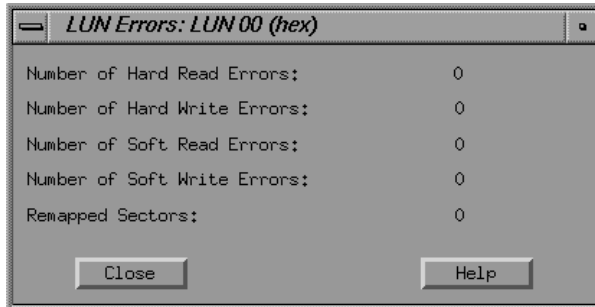
Figure 4-9   Example LUN Errors Window

Table 4-6 explains the entries in this window.

Table 4-6   LUN Errors Window Entries

Entry	Meaning
Number of Hard Read Errors Number of Hard Write Errors	Total read or write errors for all the disk modules in the LUN that persisted through all the retries. An increasing number of hard errors might mean that one or more of the LUN's disk modules is nearing the end of its useful life.
Number of Soft Read Errors Number of Soft Read Errors	Total number of read or write errors for all the disk modules in the LUN that disappeared before all the retries. An increasing number of soft errors might indicate that one of the LUN's disk modules is nearing the end of its useful life.
Remapped Sectors	Total disk sectors on all the disk modules in the LUN that were faulty when written to, and thus were remapped to different parts of the disk modules.

You can use other buttons in the LUN Information window to unbind the LUN, change the LUN's parameters, or display the information for the next LUN.

Dual Interfaces, Load Balancing, and Device Names

If your storage system has two SPs (split-bus, dual-bus/dual-initiator, or dual-interface/dual-processor), you can choose which disks to bind on each SP. This flexibility lets you balance the load on your SCSI-2 interfaces and SPs.

The SP on which you bind a physical disk unit is its default owner. The route through the SP that owns a physical disk unit is the primary route to the disk unit, and determines the device name of the disk unit. The route through the other SP is the secondary route to the disk unit.

When the storage system is running, you change the primary route to a physical disk unit by transferring ownership of the physical disk unit from one SP to another. The change in ownership and the new route take effect at the next power-on. See "Transferring Control of a LUN" in Chapter 8 for information.

99

Maintaining Disk Modules

This chapter describes

- identifying and verifying a failed disk module
- setting up the workplace
- replacing a disk module
- installing an add-on disk module array

Identifying and Verifying a Failed Disk Module

If you have determined that a module has failed by examining the cabinet fault light or by using the *raidcli getdisk* or *raidcli getcrus* command, as explained in Chapter 3 in this guide, you can replace the defective module and rebuild your data without powering off the Challenge RAID storage system or interrupting user applications.

Caution: Removing the wrong drive module can introduce an additional fault that shuts down the physical disk containing the failed module. Before removing a disk module, verify that the suspected module has actually failed.

The fault indicator on a disk module does not necessarily mean that the drive itself has failed. Failure of a SCSI bus, for example, lights the fault indicator on each disk module on that bus.

To verify a suspected disk-module failure, follow these steps:

1. Look for the module with its amber fault light on. Figure 5-1 shows the fault indicator light and other lights on a disk module.
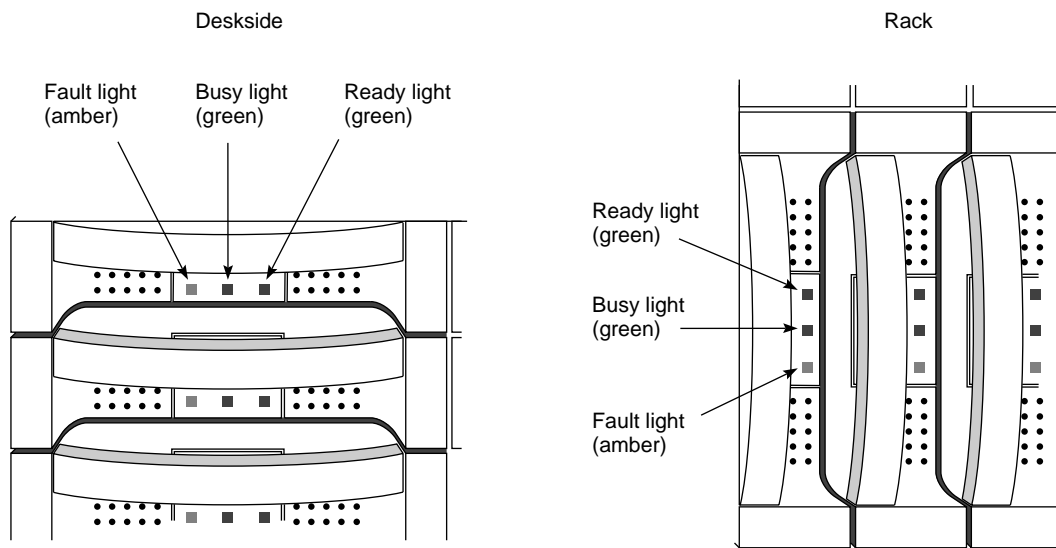
Figure 5-1   Disk Module Status Lights

2. Determine the failed module's ID; use Figure 5-2.

Caution: Use only Challenge RAID disk modules to replace failed disk modules. Order them from the Silicon Graphics hotline: 1-800-800-4SGI (1-800-800-4744). Challenge RAID disk modules contain proprietary firmware that the storage system requires for correct functioning. Using any other disks, including those from other Silicon Graphics systems, can cause failure of the storage system.

Deskside

A1	A0
B1	B0
C1	C0
D1	D0
E1	E0
A3	A2
B3	B2
C3	C2
D3	D2
E3	E2

Chassis assembly in rack

| A0 | B0 | C0 | D0 | E0 | A2 | B2 | C2 | D2 | E2 |
| A1 | B1 | C1 | D1 | E1 | A3 | B3 | C3 | D3 | E3 |

5 to 20 disk modules in groups of 5
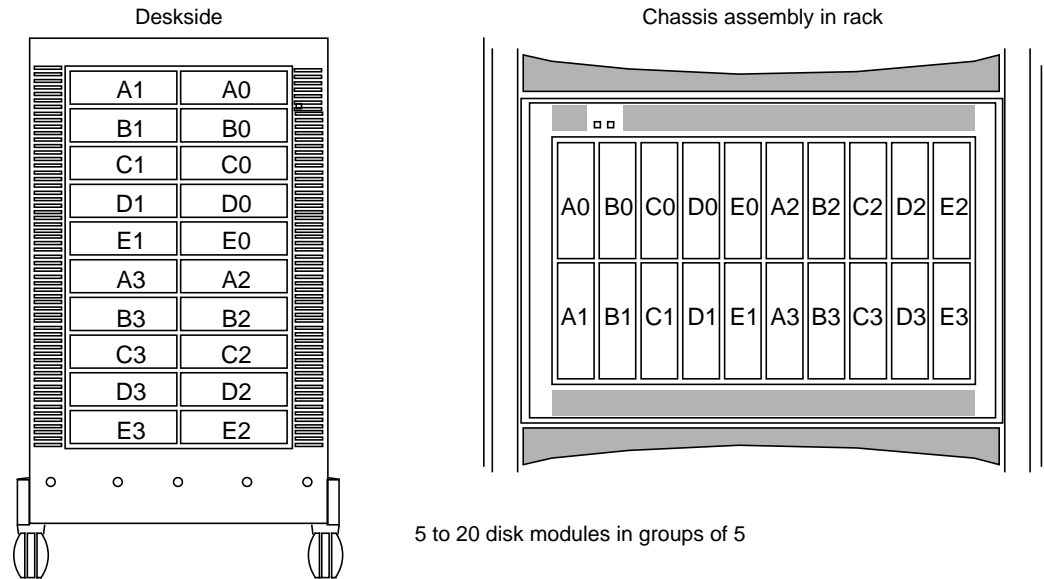
Figure 5-2　　Disk Module Locations

3.　If you have not already checked the module status with *raidcli getdisk*, do so now; see "Getting Information About Disk Modules" in Chapter 3.

4.　If you have not already checked the unsolicited error log for a message about the disk module, as explained in "Displaying the Challenge RAID Unsolicited Event Log" in Chapter 3, do so now.

A message about the disk module contains its module ID (such as A0 or B3). Check for any other messages that indicate a related failure, such as failure of a SCSI bus or a general shutdown of a chassis, that might mean the disk module itself has not failed.

Note: If you are using storage system caching, the system uses modules A0, B0, C0, D0, and E0 for its cache vault. If one of these modules fails, the storage system dumps its cache image to the remaining modules in the vault; then it writes all dirty (modified) pages to disk and disables caching. The cache status changes, as indicated in the output of the *raidcli getcache* command. Caching remains disabled until you insert a replacement module and the storage system rebuilds the module into the physical disk unit. For information on caching, see Chapter 7.

Caution: Although you can remove a disk module without damaging the disk data, do this only when the disk module has actually failed. Never remove a disk module unless absolutely necessary, and only when you have its replacement available. Never replace more than one disk module at a time; use only correct disk modules available from Silicon Graphics, Inc.

Setting Up the Workplace for Replacing or Installing Disk Modules

Electrostatic discharge can build up on your body, and can accidentally discharge through the circuits of disk modules that you handle. To avoid this situation, follow these guidelines for setting up the work area before you replace disk modules or install additional arrays:

- If the air in the work area is very dry, running a humidifier in the work area will help decrease the risk of electrostatic charge damage (ESD).

- Provide enough room to work on the equipment. Clear the work site of any unnecessary materials; remove materials that naturally build up electrostatic charge, such as foam packaging, foam cups, cellophane wrappers, and similar materials.

- The disk module is extremely sensitive to shock and vibration. Even a slight jar can severely damage it.

- Do not remove disk modules from their antistatic packaging until the exact moment that you are ready to install them.

- Before removing a disk module from its antistatic bag, place one hand firmly on an unpainted metal surface of the chassis, and at the same time, pick up the disk module while it is still sealed in its antistatic bag. Once you have done this, do not move around the room or contact other furnishings, personnel, or surfaces until you have installed and secured the disk module in the equipment.

- After you remove a disk module or filler module, avoid moving away from the work site; otherwise, you may build up an electrostatic charge.

- If you must move around the room or touch other surfaces before securing the disk module in the storage system, first place the disk module back in the antistatic bag. When you are ready again to install the disk module, repeat these procedures.

104

Replacing a Disk Module

This section explains

- ordering replacement disk modules

- unbinding the disk

- removing the failed disk module

- installing the replacement disk module

Ordering Replacement Disk Modules

Order replacement disk modules only from the Silicon Graphics, Inc. hotline:

1-800-800-4SGI (1-800-800-4744)

Use Table 5-1 as a guide to ordering replacement disk modules.

Table 5-1   Ordering Replacement Disk Modules

Unit	Silicon Graphics Marketing Code
Replacement 2.1 GB drive	P-S-RAID-1X2
Replacement 4.2 GB drive	P-S-RAID-1X4

Caution: Use only Challenge RAID disk modules as replacements; only they contain the correct device firmware. Other disk modules, even those from other Silicon Graphics equipment, will not work. Do not mix disk modules of different capacities within one array.

Unbinding the Disk

When you change a physical disk configuration, you change the bound configuration of a physical disk unit. Physical disk unit configuration changes when you add or remove a disk module, or physically move one or more disk modules to different slots in the chassis.

Caution: Unbinding destroys all the data on a physical disk unit. Before unbinding any physical disk unit, make a backup copy of any data on the unit that you want to retain.

You cannot use any unbound disk module until you bind it into a LUN again.

See "Unbinding LUNs in a New Challenge RAID Storage System" in Chapter 4 for detailed instructions. Use *raidcli bind* or RAIDGUI to configure disks, as explained in Chapter 4.

Removing a Failed Disk Module

You can replace a failed disk module while the storage system is powered on. If necessary, you can also replace a disk module that has not failed, such as a module that has reported many "soft" errors. When replacing a module that *has not* failed, you must do so while the storage system is powered up so that the SP knows the module is being replaced.

Caution: To maintain proper cooling in the storage system, never remove a disk module until you are ready to install a replacement. Never remove more than one disk module at a time.

To remove a disk module, follow these steps:

1. Verify that the suspected module has actually failed.

    Caution: If you remove the wrong disk module, you introduce an additional fault that shuts down the physical disk containing the failed module. In this situation, the operating system software cannot access the physical disk until you initialize it again.

2. Read "Setting Up the Workplace for Replacing or Installing Disk Modules," earlier in this chapter.

3. Locate the disk module that you want to remove; see Figure 5-2 if necessary.

4. Position the new disk module in its antistatic packaging within reach of the storage system.

5. If you are using an ESD wrist strap, attach its clip to the ESD bracket at the bottom of the storage system. Figure 5-3 shows where to attach the clip on a deskside storage system.
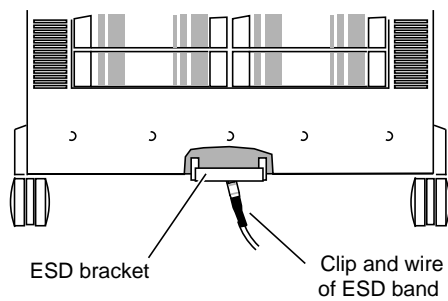
Figure 5-3 Attaching the ESD Clip to the ESD Bracket on a Deskside Storage System

Figure 5-4 shows where to attach the clip on a rack storage system.
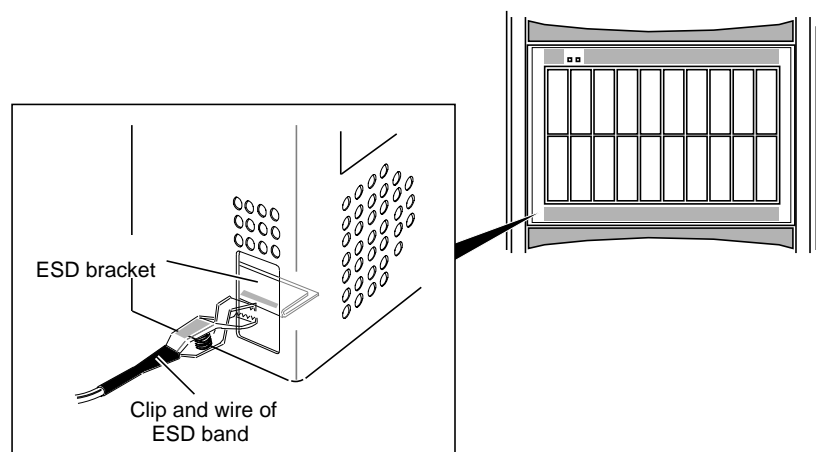
Figure 5-4 Attaching the ESD Clip to the ESD Bracket on a Rack Storage System

6. Put the wrist band around your wrist with the metal button against your skin.

7. Make sure the disk has stopped spinning and the heads have unloaded.

8. Grasp the disk module by its handle and pull it part of the way out of the cabinet, as shown in Figure 5-5.
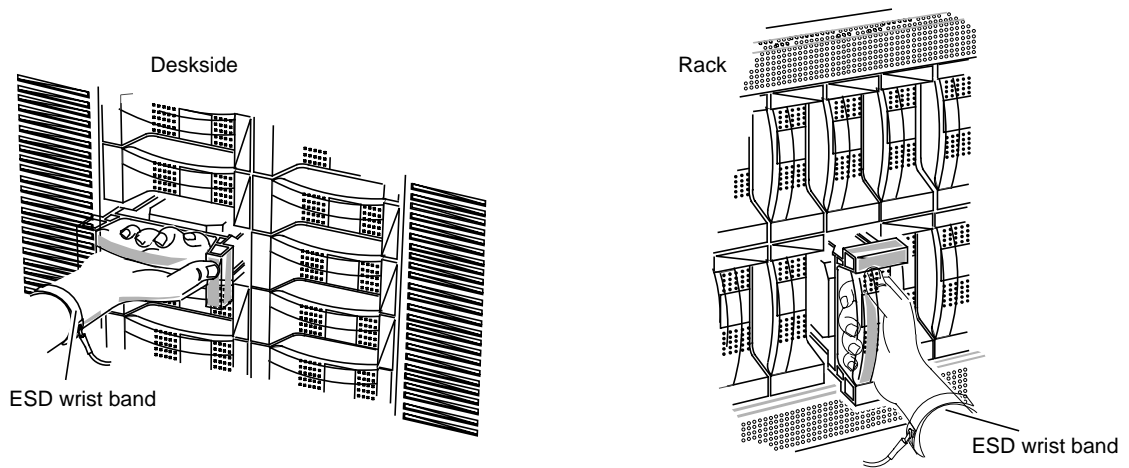
Figure 5-5   Pulling Out a Disk Module

Caution: Never remove more than one disk module at a time.

Warning: **When removing a disk module from an upper chassis assembly in a Challenge RAID rack system, make sure that you adequately balance the weight of the disk module.**

9. Supporting the disk module with your free hand, pull it all the way out of the cabinet, as shown in Figure 5-6.
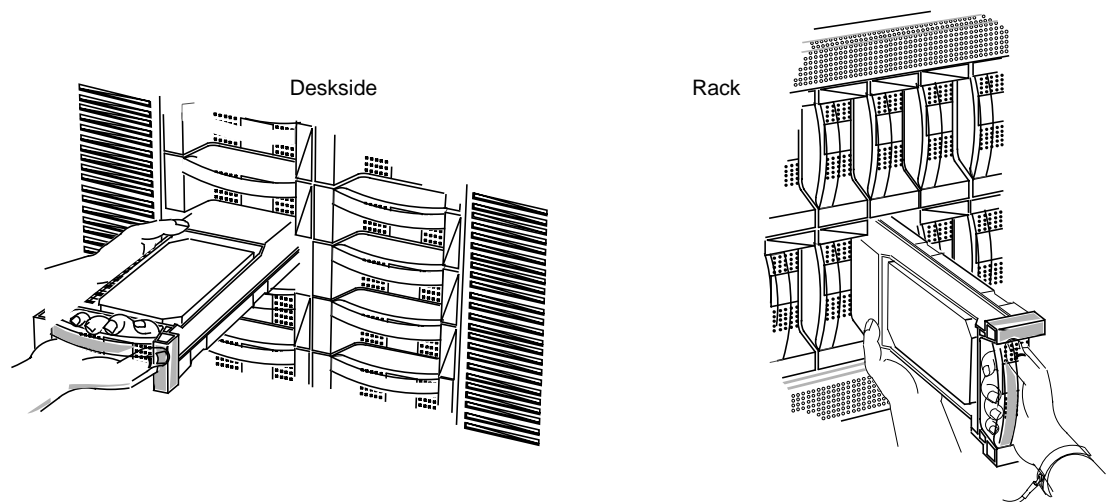
Figure 5-6   Removing a Disk Module

Caution: When removed from the chassis, the disk modules are extremely sensitive to shock and vibration. Even a slight jar can severely damage them.

10. If the label on the side of the disk module does not show the ID number for the compartment from which you removed the drive, write it on the label; for example, A1.

For the compartment ID numbers, refer to Figure 5-2 or the slot matrix attached to the storage system when it was installed.

11. Put the failed disk module in an antistatic bag and store it in a place where it will not be damaged.

Caution: Before installing a replacement module, wait at least 15 seconds after removing the failed module to allow the SP time to recognize that the module has been removed. If you insert the replacement module too soon, the SP may report the replacement module as defective.

Installing a Replacement Disk Module

To install the replacement disk module, follow these steps:

1. Touch the new disk module's antistatic packaging to discharge it and the drive module. Remove the new disk module from its packaging.

   Caution: The disk module is extremely sensitive to shock and vibration. Even a slight jar can severely damage it.

2. On the label on the side of the disk module, write the ID number for the compartment into which the drive is going; for example, A3.

3. Engage the disk module's rail in the chassis rail slot, as shown in Figure 5-7.

Figure 5-7   Engaging the Disk Module Rail

4. Engage the disk module's guide in the chassis guide slot, as shown in Figure 5-8.

Figure 5-8   Engaging the Disk Module Guide

5. Insert the disk module, as shown in Figure 5-9. Make sure it is completely seated in the slot.
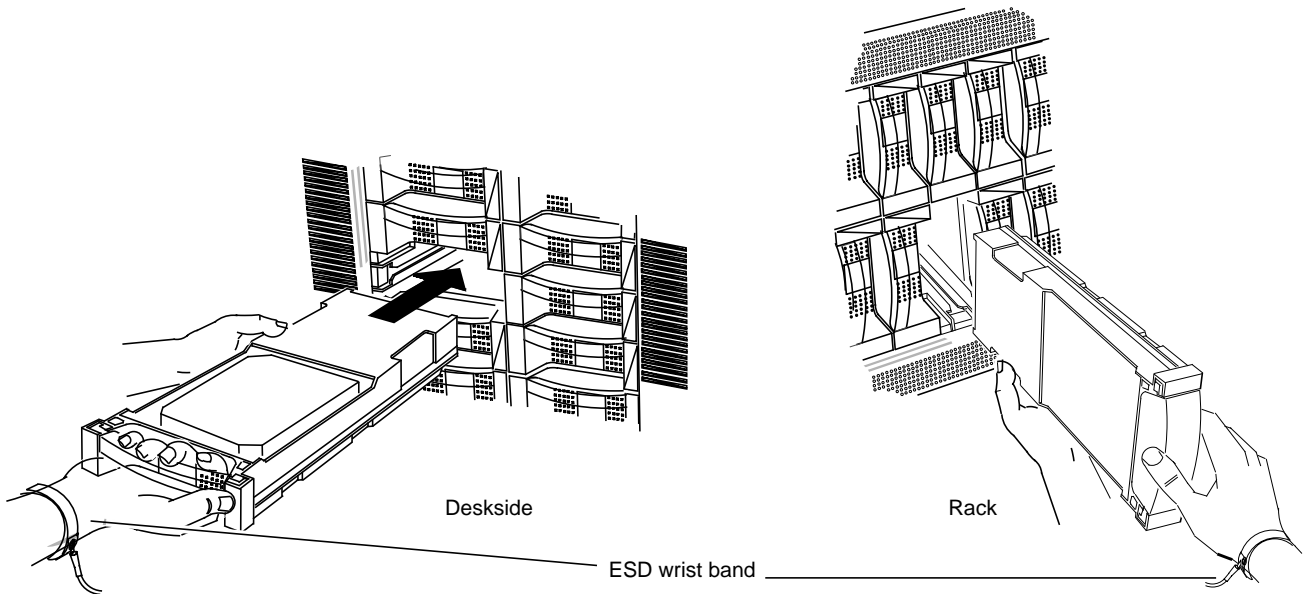
Figure 5-9   Inserting the Replacement Disk Module

6.   Remove and store the ESD wrist band, if you are using one.

The SP formats and checks the new module, and then begins to reconstruct
the data. While rebuilding occurs, you have uninterrupted access to
information on the physical disk unit.

Note: If you have replaced an unbound (database) disk module—A0, B0,
C0, or A3—unmount any filesystems or partitions on the storage system and
reboot the system so that the remaining unbound modules update the
firmware on the replacement unbound module.

The default rebuild period is four hours.

Note: During the rebuild period, performance might degrade slightly,
depending on the rebuild time specified and on I/O bus activity.

For more information on changing the default rebuild period, see "Binding
Disks Into RAID Units" in Chapter 4.

Installing an Add-On Disk Module

As your organization's needs change, you may need to add to or change the configurations of your storage system's physical disk units. For example, you might want to add disk modules to unused slots or change the ownership of a physical disk unit from one SP to another.

If the storage system has unused disk module slots containing only filler modules, you can increase the available storage capacity by installing additional disk modules. Normal processing can continue while you install disk modules.

This section explains

- ordering add-on disk modules
- inserting the new disk module
- creating device nodes and binding the disks

Ordering Add-On Disk Module Arrays

Call the Silicon Graphics, Inc., hotline to order add-on disk modules:

1-800-800-4SGI (1-800-800-4744)

Use Table 5-2 as a guide to ordering add-on disk module arrays.

Table 5-2   Ordering Add-On Disk Module Sets

Unit	Marketing Code
Add-on five 2.1-GB drives	P-S-RAID-5X2
Add-on five 4.2-GB drives	P-S-RAID-5X4
Base array with five 2.1-GB drives	P-S-RAID-B5X2
Base array with five 4.2-GB drives	P-S-RAID-B5X4
Replacement 2.1-GB drive	P-S-RAID-1X2
Replacement 4.2-GB drive	P-S-RAID-1X4

Caution: Use only Challenge RAID disk modules as replacements; only they contain the correct device firmware. Other disk modules, even those from other Silicon Graphics equipment, will not work. Do not mix disk modules of different capacities within one array. Do not remove disk modules from bus 0 (slots A0, B0, C0, D0, and E0) for use in other disk module positions.

Installing New Disk Modules

To install new disk modules, follow these steps:

Caution: Leave the filler modules installed until the add-on replacement modules are available. Never remove more than one disk module or disk filler module at a time.

1. Read "Setting Up the Workplace for Replacing or Installing Disk Modules," earlier in this chapter.

2. Position the new disk modules in their antistatic packaging within reach of the storage system.

3. If you are using a wrist band, attach its clip to the ESD bracket on the bottom of the storage system, as shown in Figure 5-3. Put the wrist band around your wrist with the metal button against your skin.

4. Locate the slots where you will install the disk modules; see Figure 5-10.

Deskside

A1	A0
B1	B0
C1	C0
D1	D0
E1	E0
A3	A2
B3	B2
C3	C2
D3	D2
E3	E2

Chassis assembly in rack

A0	B0	C0	D0	E0	A2	B2	C2	D2	E2
A1	B1	C1	D1	E1	A3	B3	C3	D3	E3

5 to 20 disk modules in groups of 5
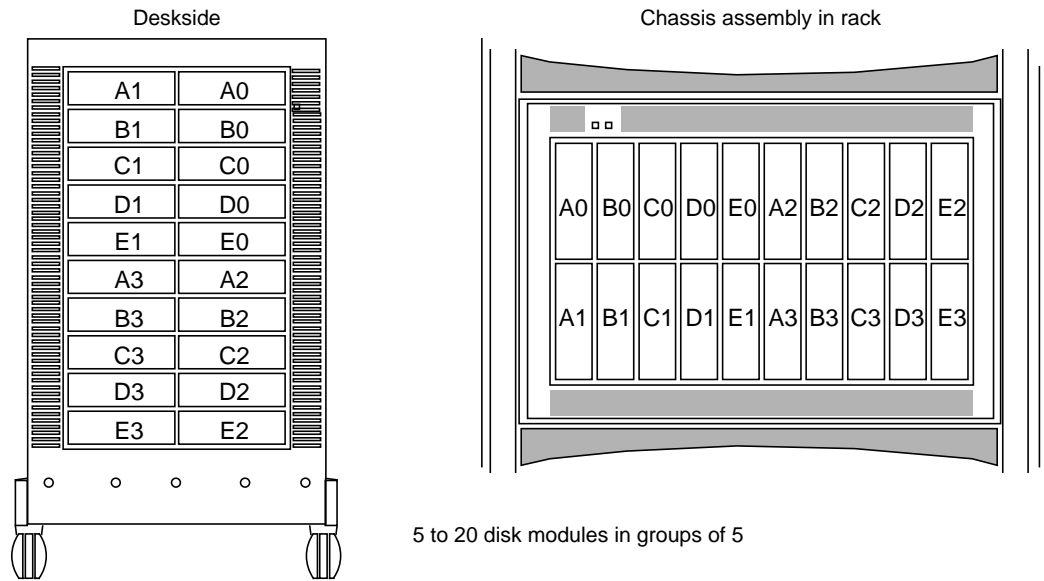
Figure 5-10   Disk Drive Locations

Warning: **For a Challenge RAID rack, although you need not complete each chassis assembly that is partially filled before installing more disk modules in the next chassis assembly, avoid making the rack top-heavy.**

Fill the slots in this order:

- first, in this order: A0, B0, C0, D0, E0

- next, in this order: A1, B1, C1, D1, E1

- next, in this order: A2, B2, C2, D2, E2

- finally, in this order: A3, B3, C3, D3, E3

5. Grasp the filler module for the first slot and pull it out of the cabinet; set it aside for possible future use. If you cannot grasp the module, use a medium-size flat-blade screwdriver to pry it out gently.

6. Touch the new disk module's antistatic packaging to discharge it and the drive module. Remove the new disk module from its packaging.

115

7. On the label on the side of the disk module, write the ID number for the compartment into which the drive is going. You can either write the slot position on the label in the corresponding place on the matrix or make a check mark in the position to indicate the slot that the disk module occupies. Figure 5-11 shows these two ways of labeling disk module B0.
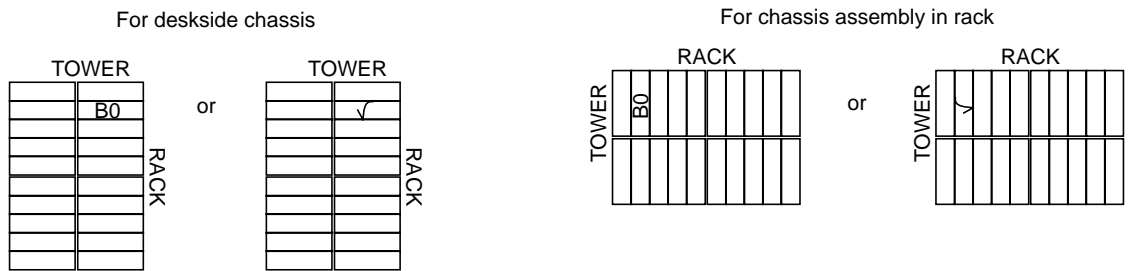
Figure 5-11   Marking the Label for Disk Module B0

8. Engage the disk module's rail in the chassis rail slot, as shown in Figure 5-12.

    Caution: Disk modules are extremely sensitive to shock and vibration. Even a slight jar can severely damage them.
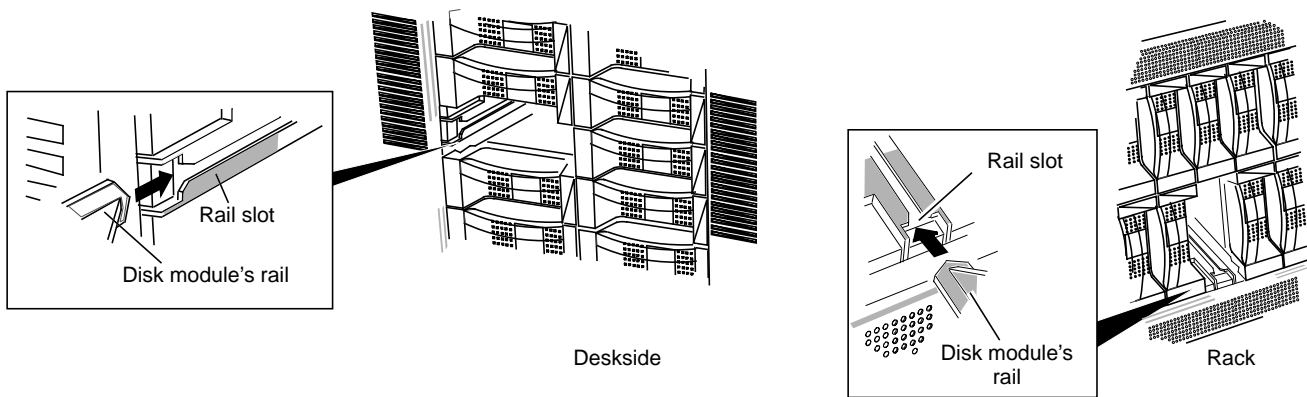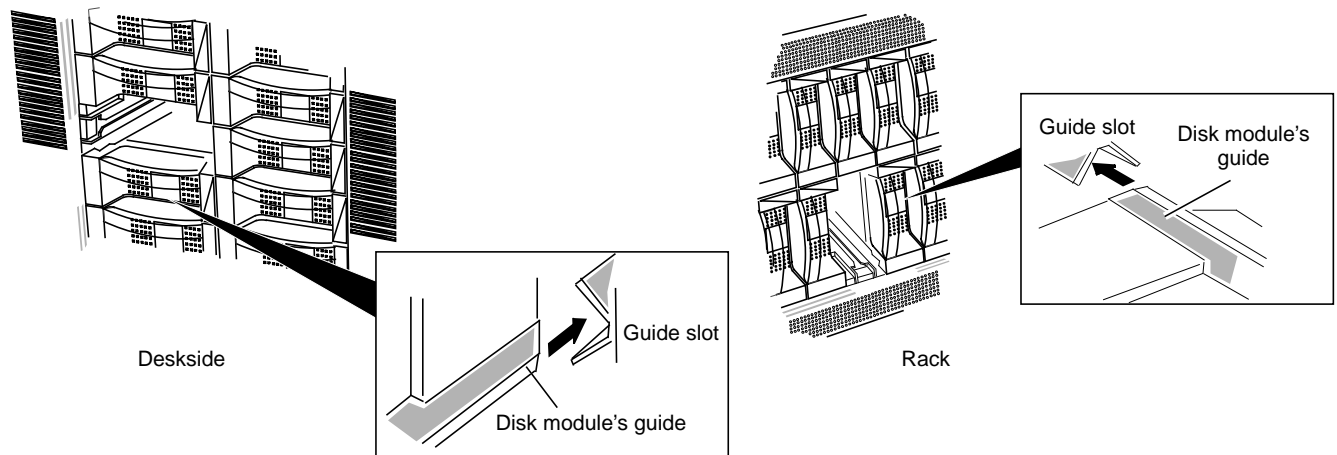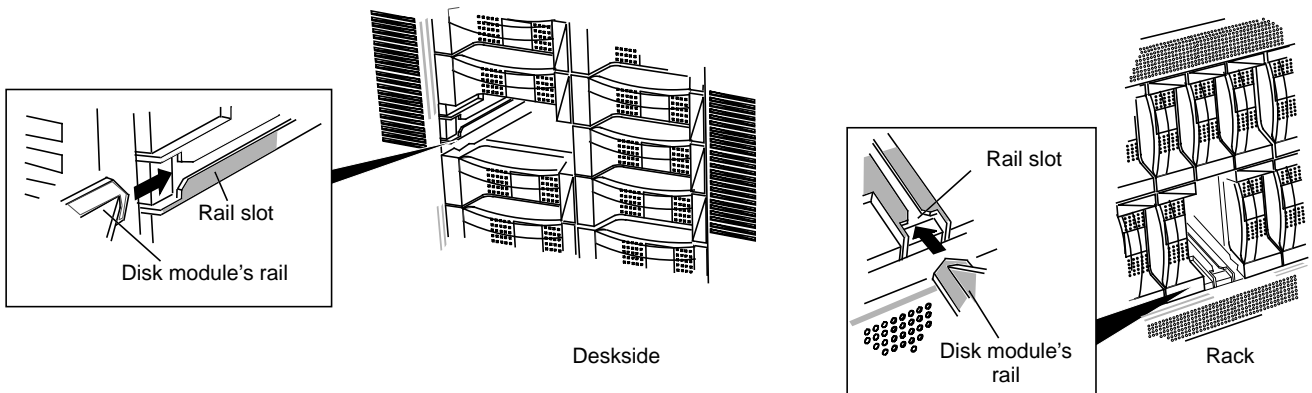

Figure 5-12   Engaging the Disk Module Rail

9. Engage the disk module's guide in the chassis guide slot, as shown in Figure 5-13.
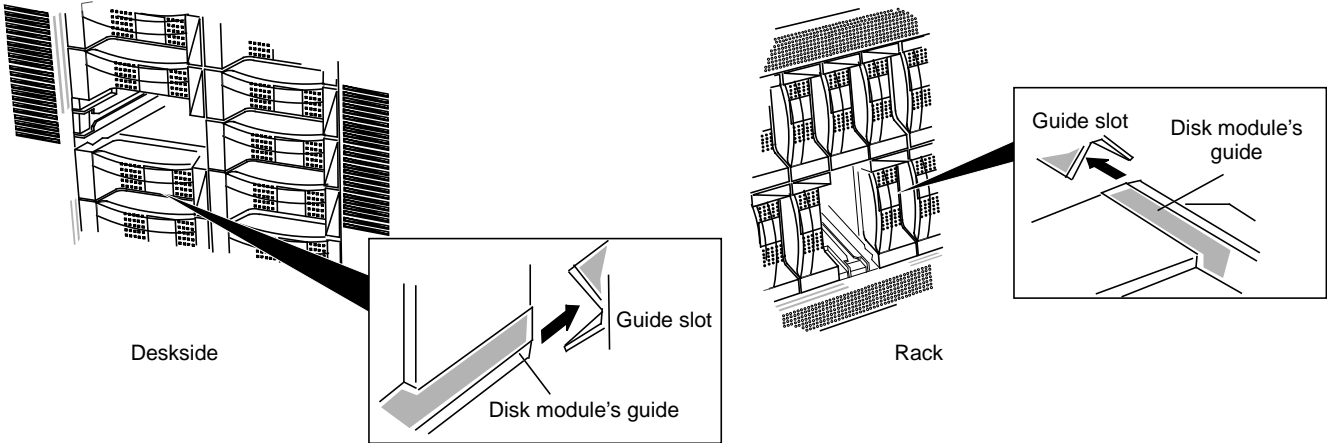
Figure 5-13   Engaging the Disk Module Guide

10. Insert the disk module, as shown in Figure 5-14. Make sure it is completely seated in the slot.
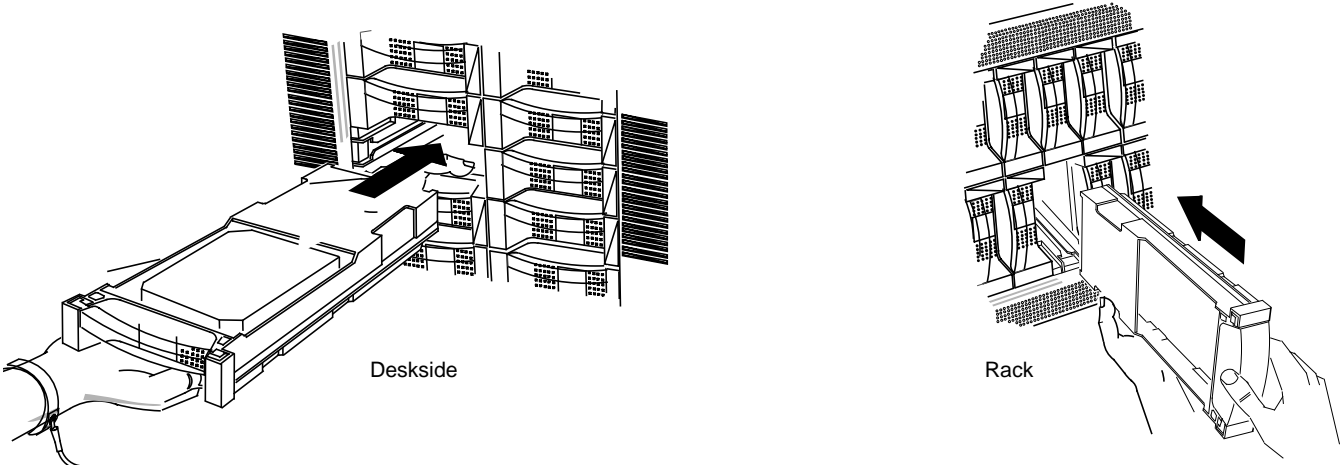
Figure 5-14   Inserting a Disk Module

11. Repeat steps 4 through 10 until all add-on modules are installed.

12. When you are finished installing add-on modules, remove and store the ESD wrist band, if you are using one.

Creating Device Nodes and Binding the Disks

If you are adding disk modules to a storage system that already has at least one LUN configured, the SPs must be made aware of the new disks. This section explains how to accomplish this task for IRIX 5.3 systems without rebooting. Also, in a system with two storage-control SPs that are used for primary and secondary paths, both SPs must be made aware of the new disks. Finally, the new disks must be bound into LUNs.

Follow these steps:

1. Change to the */dev* directory:

   ```
   cd /dev
   ```

2. Type

   ```
   ./MAKE_VLUNS <controller-number> <target-number>
   ```

   This command creates the device nodes for the new disks.

3. Bind the newly installed modules into one or more physical disk units, as described in "Binding Disks Into RAID Units," in Chapter 4 in this guide.

Identifying Failed System Components

This chapter describes how the Challenge RAID storage system operates in the case of failure of a system component (field-replaceable unit) other than a disk module: a fan module, power supply, optional battery backup unit, or storage-control processor board.

Read this chapter after you have determined that one of these components has failed as indicated by the cabinet fault light, in the output of the *raidcli getcrus* command, or in the graphical user interface Equipment view, as explained in "Getting Information About Other Components," in Chapter 3.

Note: These components can be replaced only by qualified Silicon Graphics System Service Engineers or other qualified service providers. Only disk modules are owner replaceable or end-user replaceable; Chapter 5 provides instructions.

Call the Silicon Graphics hotline to order a replacement module:

1-800-800-4SGI (1-800-800-4744)

Power Supply

The Challenge RAID storage system has two or three redundant power supplies, or VSCs (voltage semi-regulated converters): VSC A, VSC B, and, optionally, VSC C. If the storage system has three power supplies, it can recover from power supply component failures and provide uninterrupted service while the defective component is replaced.

In a storage system with two power supplies, a failed power supply unit can fail or be removed without affecting the disk modules. If a second power supply fails, the entire chassis shuts down immediately. Any defective power supply must be replaced immediately by a Silicon Graphics System Service Engineer.

Failure of the AC distribution system (line cord, utility power, and so on) also shuts down the entire chassis immediately.

Note: When the storage system shuts down, the operating system loses contact with the LUNs. When the storage system starts up automatically, you may need to reboot it to let the operating system access the LUNs.

In RAIDGUI, an amber power supply button in the Equipment View indicates that the power supply this button represents is in one of the following states:

- Down: the power supply failed or was removed after the agent started running

- Not Present: the power supply failed or was removed before the agent started running

These states appear in */var/adm/SYSLOG*; if alarms are enabled in RAIDGUI, they appear in e-mail messages, on-screen alarm messages, or both, depending on the setting.

Fan Module

Each Challenge RAID storage system or rack chassis assembly has one fan module, containing six fans wired in two groups (FAN A and FAN B) of three fans each. If any fan fails, the fan fault light on the back of the fan module turns on, and all other fans speed up to compensate. The storage system can run after one fan fails; however, if another fan failure occurs and temperature rises, the storage system shuts down after two minutes.

If the fault light comes on, if you see a fan fault in the *raidcli getcrus* command output, or if a fan button is amber in the RAIDGUI Summary View, have the entire fan unit replaced as soon as possible by a Silicon Graphics System Service Engineer.

An amber fan button in the Equipment View of RAIDGUI indicates that the fan group this button represents is down (the fan group failed or the fan module was removed after the agent started running) or not present (the fan group failed or the fan module was removed before the agent started running). (These states appear in */var/adm/SYSLOG*; if alarms are enabled in RAIDGUI, they appear in e-mail messages, on-screen alarm messages, or both, depending on the setting.)

Note: Leaving the rear door open for more than two minutes can cause the storage system to overheat. If the temperature within a Challenge RAID chassis reaches an unsafe level, the power system shuts down the storage system immediately.

Battery Backup Unit

When the optional battery backup unit fails, the following events occur:

- The battery backup unit's service light turns on.

- If the storage system is using the cache, the storage system's performance may become sluggish while it writes any modified pages from the cache to disk.

- Caching is disabled; the cache state as shown in the *raidcli getcache* command or in the SP Cache Summary window is "Disabled." Caching is not re-enabled until the battery backup unit is replaced with a fully charged one.

The *raidcli getcrus* output has three possible states for the battery backup unit: Faulted (removed), Charging, and Present (fully charged or charging). If the battery backup unit takes longer than an hour to charge, it shuts itself off and transitions to the "Faulted" state.

If the fault light comes on, if the battery backup unit state is shown as "Faulted" in the *raidcli getcrus* command output, or if a BBU button is amber in the RAIDGUI Summary View, have the battery backup unit replaced as soon as possible by a Silicon Graphics System Service Engineer.

After a power outage, a BBU takes 15 minutes to recharge. From total depletion, recharging takes an hour or less.

Each week, the SP initiates BBU self-test to ensure that the BBU's monitoring circuitry is working. While the test runs, storage-system caching is disabled, but communication with the server continues. I/O performance may decrease during the test. When the test is finished, storage-system caching is re-enabled automatically. The default time for the BBU test to start is 1:00 a.m. on Sunday.

Storage-Control Processor

When an SP fails, the storage system's performance may become sluggish if the storage system is using caching. You may find that one or more physical disk units are inaccessible. The failed SP's service light turns on, and the service light on the front of the storage system turns on.

You can determine SP failure these ways:

- View the unsolicited event log, which contains the message "Peer Controller: Removed." View this log with *raidcli getlog* or by clicking *Log* in the RAIDGUI Summary View, as explained in Chapter 3.

- Use *raidcli getcrus* to verify the failed SP, as explained in Chapter 3. The *getcrus* output indicates the failed SP, for example, "SPA State: Not Present."

- In RAIDGUI, click on the SP's button in the Equipment View or Summary View to display the SP Summary window, as explained in Chapter 3.

If the Challenge RAID storage system has a second SP and a number of LUNs are inaccessible, you might want to transfer control of the LUNs to a surviving SP. See "Transferring Control of a LUN" in Chapter 8.

Caching

Storage-system caching requires

- fully charged storage-system backup battery

- cache enabling using the *raidcli setcache* command, as explained in this chapter

- disk modules in slots A0, B0, C0, D0, and E0 as a fast repository for cached data

- for mirrored caching: two SPs, each with at least 8 MB of memory; both SPs must have the same amount of memory

- for nonmirrored caching: one SPs with at least 8 MB of memory

Caching cannot occur unless all these conditions are met.

This chapter explains

- setting cache parameters

- viewing cache statistics

- upgrading Challenge RAID to support caching

- changing cache unit parameters

Setting Cache Parameters

This section explains

- setting cache parameters using the command-line interface
- setting cache parameters using RAIDGUI

Setting Cache Parameters Using the Command-Line Interface

The cache parameters you specify for the entire storage system are the cache size of 8 or 64 MB, depending on the amount of memory the SPs have, and the cache page size, as 2, 4, 8, or 16 KB.

To set up caching, use the *raidcli setcache* command:

raidcli –d *device* setcache *enable* | *disable* [–u *usable*] [–p *page*] [–l *low*] [–h *high*]

In this syntax, the variables and their meanings are as follows:

enable | disable Enter 1 to enable caching or 0 to disable caching.

-u *usable* Size in megabytes to use for caching, not greater than the SP memory size as displayed by *raidcli getagent* (see "Getting Device Names With getagent," in Chapter 3). Valid values are 0 through 64 in increments of 1 MB.

The command-line interface does not let you specify more memory than you have. If you specify less than you have, the remaining memory is unused.

Note: For mirrored caching, both SPs must have the same amount of cache memory in order for caching to be preserved in the event of shutdown or other power loss.

-p *page* Size in KB of pages into which to partition the cache. Valid sizes are 2, 4, 8, and 16. The default is 2, regardless of whether caching is enabled or disabled.

Generally, set cache page size to 8 KB for IRIX fileserver applications and 2 KB or 4 KB for database applications.

124

-l *low*   Percentage of cache full that discontinues flushing. Valid values are 0 through 100; the default is 50, regardless of whether caching is enabled or disabled.

-h *high*   Percentage of cache full that initiates flushing. Valid values are 0 through 100; the setting must be greater than the low watermark. The default is 75, regardless of whether caching is enabled or disabled.

-sma *system-memory-SPA*, -smb *system-memory-SPB*
     Set the system memory on SP A or B, respectively.

-sta *read-cache-state-SPA*, stb *read-cache-state-SPB*
     Set the read cache state for SP A or B, respectively: 0 to disable, 1 to enable.

-rca *read-cache-size-SPA*, -rcb *read-cache-size-SPB*
     Set the read cache size for SP A and B, respectively.

     Note: You can change the write cache size only if the write cache is disabled and no unassigned dirty pages exist.

Note: The sum of an SP's write cache size, system memory size, and read cache size must be less than or equal to the SP's physical memory size.

This command has no output.

You can change the cache size, the cache page size values, or the type of caching for any physical disk unit without affecting the information stored on it. Follow these steps:

1. Disable the cache:

   raidcli -d *device* **setcache 0**

2. Wait for the cache memory to be written to disk, which may take several minutes.

3. Reassign cache size and re-enable caching, for example, with a cache size of 64 MB:

   raidcli -d *device* **setcache 1 -u 64 [parameters]**

Note: Before changing a cache parameter, you must always disable cache.

Setting Cache Parameters Using RAIDGUI

To set up storage-system caching, you specify the read/write cache parameters and the size of the memory partitions. This section describes

- when to disable or enable read and write caching

- disabling and enabling read and write caching

- selecting read and write cache parameters

- specifying memory partitions

When to Disable or Enable Read and Write Caching

When the Challenge RAID storage system is powered on, it enables read and write caching if the required hardware is working and the read and write cache size parameters have values other than zero.

You can disable or enable storage-system caching without affecting the information stored on the LUNs. You must disable certain caches before changing the memory partition parameters and most LUN caching parameters. You should also disable the storage-system write cache before replacing an SP.

In a storage system with two SPs, you can disable or enable an SP's read cache only if the server's agent configuration file contains a device entry with a communication channel to the SP (the device must exist in the */usr/raid5/raid5_agent.config* file). If this entry is not present, the SP's Read Cache Enabled parameter is dimmed in the SP Cache Settings window.

Disabling and Enabling Read and Write Caching

To disable or enable storage-system caching, follow these steps:

1. Click *Cache* in the Equipment View or Summary View toolbar, or select "Caching" in the Equipment View or Summary View Options menu. The SP Cache Settings window appears, as shown in Figure 7-1.
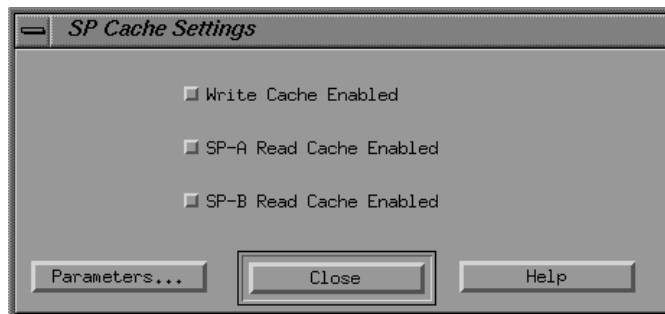
Figure 7-1   SP Cache Settings Window

2. To change the write cache status, click *Write Cache Enabled.* Answer the confirmation popup accordingly.

3. To change an SP's read cache status, click the *Read Cache Enabled* button for that SP. Answer the confirmation popup accordingly.

Disabling a cache may take several minutes. The parameter in the SP Cache Settings window for any cache you disabled remains dimmed until the cache is actually disabled.

If a popup appears saying that caching cannot be enabled, make sure that the hardware required for caching is operating. The SP(s) must be powered up, as indicated by the gray color of the SP buttons in the Equipment and Summary Views.

If you cannot enable write caching, make sure that

- the BBU is fully charged

    Note: The only sure indication that the BBU is not fully charged is the appearance of event log messages that the BBU is recharging.

- there are disk modules in slots A0, B0, C0, D0, andE0

- all disk modules are functioning normally; none are currently being rebuilt

- if an SP was just replaced, the replacement SP must have the same amount of memory as the remaining SP

- the write cache size is not zero

If you cannot enable read caching, make sure that

- read cache size is not zero

- if an SP was just replaced, the replacement SP has at least as much memory as the SP it replaced

Even though storage-system caching is enabled, the storage system does not use caching with a LUN unless you have enabled the read and/or write caching for the LUN itself. You can enable caching for a LUN when you bind it; you can also set caching in the Change LUN Parameters window, as explained in "Upgrading for Caching Using RAIDGUI," later in this chapter.

Selecting Read and Write Cache Parameters

Table 7-1 gives the choices and default values for read/write cache parameters.

Table 7-1   Default Read/Write Cache Parameters

Parameter	Choices	Default
Mirror write cache	Selected, Deselected	Selected
Cache page size	2, 4, 8, or 16 KB	2 KB
Write cache high watermark	0 through 100%	75
Write cache low watermark	0 through 100%	50

To set cache parameters for the SPs in the Challenge RAID storage system, follow these steps:

1. Click *Cache* in the Equipment View or Summary View toolbar, or select "Caching" in the Equipment View or Summary View Options menu. The SP Cache Settings window appears, as shown in Figure 7-1.

2. In the SP Cache Settings window, disable caching depending on the parameter you are changing. If you are changing

    • mirror write cache: disable Write Cache

    • cache page size: disable all caches

3. Click *Parameters*; the Cache Parameters window appears. Figure 7-2 shows an example.

129

Memory parameters

Mirror
write cache
button

Memory
map

SP Cache Parameters

Read/Write Cache Parameters

☐ Mirror Write Cache   Write Cache High Watermark   75

Cache Page Size   2   ▼   Write Cache Low Watermark   50

Memory Partitions

SP-A Memory Map   SP-B Memory Map

Write: 8 MB   Write   8   ▼   Write: 8 MB   Write   8   ▼

                   Read   0   ▼   Read   0   ▼

Flare(P): 4 MB   RAID3   0   ▼   Flare(P): 4 MB   RAID3   0   ▼

                   Flare(E) 0   ▼   Flare(E) 0   ▼

                   Flare(P) 4 MB   Flare(P) 4 MB

Free: 20 MB   Free   20 MB   Free: 20 MB   Free   20 MB

                   Total   32 MB   Total   32 MB

Set   Revert   Close   Help

Figure 7-2   Example Cache Parameters Window

4. To change the cache page size, click the button beside the Cache Page
    Size scroll box, and select the size.

    Generally, set cache page size to 8 KB for IRIX fileserver applications
    and 2 KB or 4 KB for database applications.

    Note: You can change the write cache size only if the write cache is
    disabled and no unassigned dirty pages exist.

5. To change the percentage for the write cache high or low watermark, enter it in the Write Cache High Watermark field or Write Cache Low Watermark field, respectively.

   The high watermark is the percentage of dirty pages; when it is reached, SP begins flushing the cache. The default is 75%. If you specify a lower value, the SP(s) start flushing the cache sooner.

   The lower watermark is the percentage of cache dirty pages that determines when cache flushing stops. When the write cache low watermark is reached during a flush operation, the SP stops flushing the cache.

   Note: The high watermark cannot be less than the low watermark. To turn off watermark processing, set both watermark values to 100.

6. To set memory partitions, proceed to the next section. Otherwise, click *Set* when you are finished setting cache parameters.

   Note: To change prefetch (read-ahead caching) parameters for LUNs in a new Challenge RAID storage system, see "Changing LUN Parameters Using RAIDGUI" in Chapter 8.

   To restore settings to the values they had when the Cache Parameters window was opened, click *Revert.*

Specifying Memory Partitions

An SP's memory provides space for system buffers in addition to cache memory. An SP running Licensed Internal Code (LIC) Revision 8 requires a system buffer space of 2 MB on an AMD-based SP and 4 MP on a PowerPC-based SP. You can allocate part or all of the remaining memory to the read cache, the write cache, or divide it between both caches. If the storage system has two SPs, the write caches on both SPs are allocated the same amount of memory automatically because they mirror each other. For this reason, you can specify one write cache size only.

Table 7-2 summarizes the memory partition fields in the Cache Parameters window.

Table 7-2   Memory Partition Settings

Setting	Meaning	Default
Write	Specifies the size of the write cache, which is the same on both SPs in a storage system with two SPs.	0
Read	Specifies the size of the SP's read cache.	0
RAID3	Leave set at 0; reserved for future revisions of the Licensed Internal Code.	0
FLARE(E)	Leave set at 0; do not change this value. (Changing the value reboots the SPs.)	0
FLARE(P)	Specifies the memory required by the SP for system buffers; you cannot change this value.	4 MB
Free	Specifies the memory available for allocation. The free memory is equal to the total SP memory (Total) minus the memory allocated to the write cache (Write), read cache (Read), extra system buffer space (FLARE(E)), and the required system buffer space (FLARE(P)).	Total SP memory minus the sum of the other partition sizes (2 MB)
Total	The full amount of the SP's memory: 8, 16, 32, or 64 MB.	N/A

The memory map in the Cache Parameters window shows the allocation of the SP's memory into different partitions. If the SP is present in the system, this map shows all partitions with nonzero values proportionately. If the SP is not present in the system, the partitions are all the same size and no values are shown.

In the Write and Read fields, you can specify 0, 1, 2, 3 MB, through the maximum size, where the maximum size is the Free partition's size plus the current size of the Write or Read partition. For example, if the Free partition's size is 8 MB and the Write partition's current size is 4 MB, the Write partition can have a maximum size of 12 MB. You cannot change the size of the Flare(P) partitions, and you should not change the size of the Flare(E) partition.

Table 7-3 gives suggested sizes for Read and Write partitions for AMD-based SPs.

Table 7-3   Read and Write Partition Sizes in MB

Total SP Memory	FLARE(P)	Available Cache Memory	Suggested Read Partition	Suggested Write Partition
8	2	6	3	3
16	2	14	8	6
32	2	30	18	12
64	2	62	40	22

Table 7-4 gives suggested sizes for Read and Write partitions for PowerPC-based SPs.

Table 7-4   Read and Write Partition Sizes in MB

Total SP Memory	FLARE(P)	Available Cache Memory	Suggested Read Partition	Suggested Write Partition
8	4	4	2	2
16	4	12	8	4
32	4	28	18	10
64	4	60	40	20

To specify memory partition sizes, follow these steps:

1. Click *Cache* in the Equipment View or Summary View toolbar, or select "Caching" from the Equipment View or Summary View Options menu. The SP Cache Settings window appears, as shown in Figure 7-1.

2. In the SP Cache Settings window, disable caching depending on the partition size you are changing. If you are changing

    •   write partition: disable Write Cache

    •   read partition: disable the SP's read cache

133

3. Click *Parameters*; the Cache Parameters window appears, as shown in Figure 7-2 earlier in this chapter.

4. To change write or read cache size, click the scroll button beside the SP's Write or Read field, as shown in Figure 7-3; select the desired size.
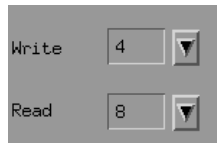
Figure 7-3   Memory Partition

Note: If you set the SP's write or read cache size to 0, write or read caching is disabled until you change the size to a nonzero value.

5. When you are finished changing memory partition settings, click *Set*; answer the confirmation popup appropriately.

Note: You can change the write cache size only if the write cache is disabled and no unassigned dirty pages exist.

Viewing Cache Statistics

If you use storage-system caching, you can use the *raidcli getcache* command to get information on cache activity. The information in this command's output, particularly the percentage of cache hits, may help you decide on the most efficient cache page size and whether a physical disk unit really benefits from caching.

Note: If a disk module in location A0, B0, C0, D0, or E0 fails, write caching is disabled until the failed module is replaced. If an SP fails, write caching is disabled until the failed SP is replaced. If the BBU fails, write caching is disabled until the BBU is replaced.

This section explains

• viewing cache statistics using the command-line interface

• viewing cache statistics using RAIDGUI

Viewing Cache Statistics Using the Command-Line Interface

To display cache information, use

```
raidcli -d device getcache
```

For a sample output and an explanation thereof, see "getcache" in Appendix B.

Viewing Cache Statistics Using RAIDGUI

To use RAIDGUI to view cache statistics, use the SP Cache Summary window. Follow the steps in this section.

1. In the Equipment View or Summary View, click the button for the SP for which you want information.

2. In the SP Summary window that appears, click *Cache.* The SP Cache Summary window appears; Figure 7-4 shows an example.
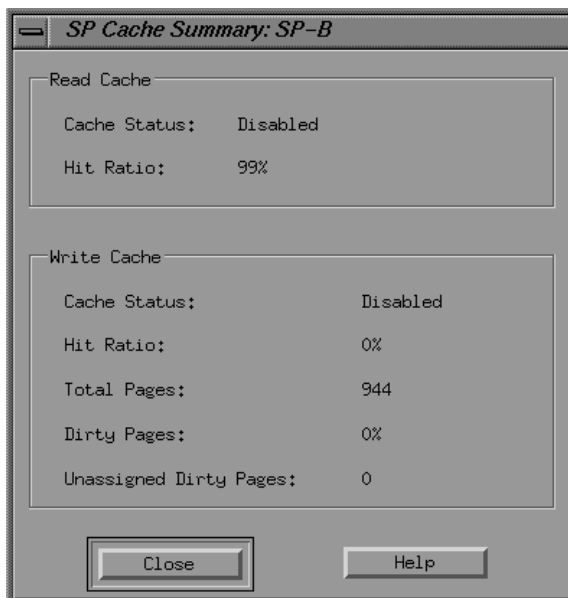
Figure 7-4   SP Cache Summary Window

The statistics in the SP Cache Summary window show the state of caches when you open the window. To update the statistics, close the window and reopen it. Table 7-5 explains read cache entries in this window.

Table 7-5   SP Cache Summary Window Read Cache Entries

Entry	Meaning
Cache Status	Current state of the SP's read cache. The possible states are Enabled, Disabling, and Disabled. An SP's read cache is automatically enabled at power-on if the cache size is a valid number and the SP has at least 8 MB of memory.
Hit Ratio	The percentage of read cache hits for the SP. A read cache hit occurs when the SP finds a sought page in read cache memory, and thus does not need to read the page from disk. The ratio is meaningful only if the SP's read cache is enabled. For more information about setting up the SP's read cache, including how to disable and enable it, see "Setting Cache Parameters," earlier in this chapter.
	A read hit occurs when the SP finds a sought page in cache memory, and thus does not need to read the page from disk. High hit ratios are desirable because each hit indicates at least one disk access that was not needed. You can compare the read and write hit ratios for the LUN with the read and write hit ratio for the entire Challenge RAID storage system in the SP Cache Summary window (Figure 7-4). For a LUN to have the best performance, the hit ratios should be higher than those for the storage system. A low read or write hit rate for a busy LUN might mean that caching is not helping the LUN's performance.

Table 7-6 explains write cache entries in this window.

Table 7-6   SP Cache Summary Window Write Cache Entries

Entry	Meaning
Cache Status	Current state of the SP's write cache. The possible states are Enabled or Disabled, and several transition states: Initializing, Enabling, Disabling, Dumping, and Frozen. An SP's read cache is automatically enabled at power-on if the cache size is a valid number and the Challenge RAID storage system has the requirements for write caching given at the beginning of this chapter.
Hit Ratio	The percentage of write cache hits for the SP's write cache. A write cache hit occurs when the SP finds and modifies data in the write cache memory, which usually saves a write operation. For example, a hit that occurs when a page is sought in a RAID 5 LUN eliminates the need to read, modify, and write the data. High ratios are desirable because each hit indicates at least one disk access that was not needed.
	A write hit occurs when the SP finds and modifies data in cache memory, which usually saves a write operation. For example, with a RAID 5 LUN, a write hit eliminates the need to read, modify, and write the data. High hit ratios are desirable because each hit indicates at least one disk access that was not needed. You can compare the read and write hit ratios for the LUN with the read and write hit ratio for the entire Challenge RAID storage system in the SP Cache Summary window (Figure 7-4). For a LUN to have the best performance, the hit ratios should be higher than those for the storage system. A low read or write hit rate for a busy LUN might mean that caching is not helping the LUN's performance.
Total Pages	Total number of pages in the SP; each page has the cache page size selected when storage-system caching was set up. This number equals the cache size divided by the cache page size, minus space for checksum tables. If the storage system has two working SPs, they divide the total number of pages between them. If an SP is idle for a long period or fails, the active SP may increase its share of pages.

Table 7-6 (continued)	SP Cache Summary Window Write Cache Entries
Entry	**Meaning**
Dirty Pages	Percentage of pages that have been modified in the SP's write cache, but that have not yet been written to disk. A high percentage of dirty pages means the cache is handling many write requests.
Unassigned Dirty Pages	Percentage of unassigned dirty pages in both SPs' write caches. Unassigned dirty pages are dirty pages belonging to a LUN that is not enabled, that is, not accessible from either SP. For example, unassigned pages result if an SP fails and its write cache contains dirty pages; these pages become unassigned pages.
	If the LUNs owned by the failed SP are transferred to the working SP, any unassigned pages for those LUNs transfer automatically to the working SP. The working SP writes these unassigned pages to the LUNs and the Unassigned Dirty Pages value returns to 0%. If the LUN that owns the dirty pages has an irrecoverable failure, you can clear the unassigned pages by unbinding the LUN as described in "Unbinding LUNs in a New Challenge RAID Storage System" in Chapter 4.

Upgrading Challenge RAID to Support Caching

Please note these points before you enable caching:

- Because disk modules are required in slots A0, B0, C0, D0, and E0 for caching, you might need to add disk modules to your storage system.

- If you add or move disk modules, you need to bind them into new physical disks or change the existing physical disk units to include them; see "Binding Disks Into RAID Units," in Chapter 4.

- Do not move a disk module to another slot unless it is absolutely necessary to do so. Never move disk modules from slots A0, A3, B0, C0, D0, and E0.

- Set up caching on one SP at a time.

Once the necessary hardware components have been installed, use either interface to set up caching.

Upgrading for Caching Using the Command-Line Interface

To upgrade caching using *raidcli*, follow these steps:

1. Enable the read and write caches for the physical disk units that will use caching:

    `/usr/raid5/raidcli -d` *device* **`chglun -l`** *lun-number* **`read`**

    or

    `/usr/raid5/raidcli -d` *device* **`chglun -l`** *lun-number* **`write`**

    or

    `/usr/raid5/raidcli -d` *device* **`chglun -l`** *lun-number* **`rw`**

2. Specify the cache page size and cache size parameters, as explained earlier in this chapter.

3. Enable storage system caching with *raidcli -d device setcache*, as explained in "Setting Cache Parameters," earlier in this chapter.

Upgrading for Caching Using RAIDGUI

To upgrade the cache, follow these steps:

1. In the Summary View, click the LUN for which you want to enable
    caching. The LUN Information window appears, as shown in
    Figure 7-5.
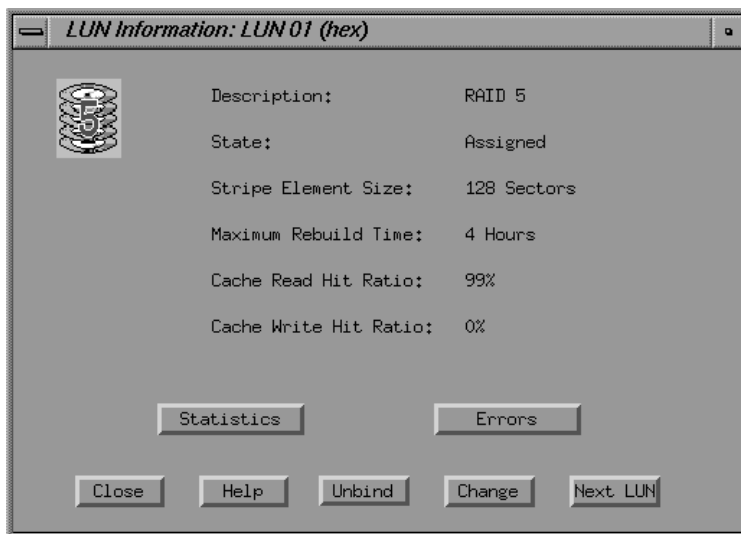
Figure 7-5   LUN Information Window

2. In the LUN Information window, click *Change*. The Change LUN
    Parameters window appears, as shown in Figure 7-6.
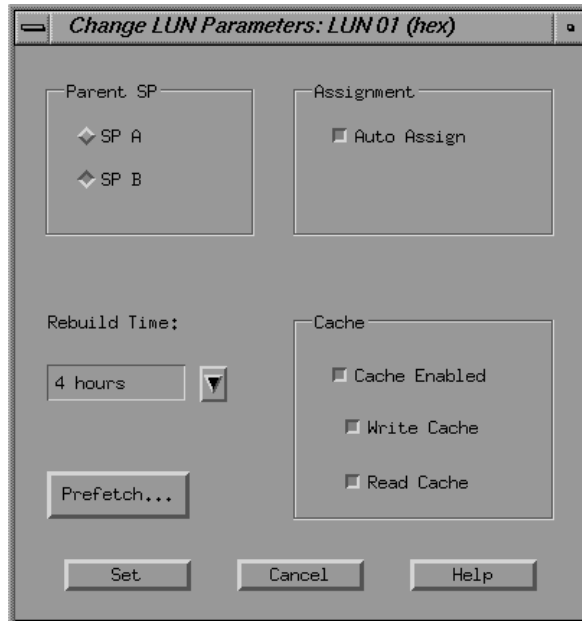
Figure 7-6   Change LUN Parameters Window

3. In the Change LUN Parameters window, click LUN Caching
    checkboxes appropriately to enable caching for the LUN. Click *Set* and
    close the window.

4. Set cache parameters as explained in "Setting Cache Parameters Using
    RAIDGUI," earlier in this chapter.

5. Enable storage-system caching: in the Equipment or Summary View, click *Cache*, or select "Caching..." from the Options menu. The SP Cache Settings window appears, as shown in Figure 7-7.
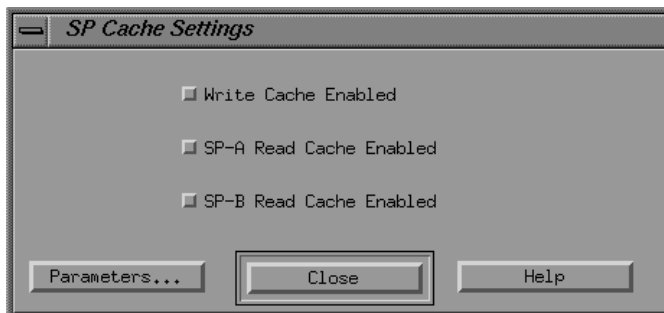
Figure 7-7   SP Cache Settings Window

6. Enable caching by clicking the checkboxes appropriately; exit the SP Cache Settings window.

Changing Unit Caching Parameters

You can use the command-line interface or RAIDGUI to change the caching parameter for a physical disk unit.

Note: If caching is enabled, you must disable it before you can change any parameter. Disabling caching can affect storage system performance; do it only when system activity is relatively low.

Changing Caching Parameters Using the Command-Line Interface

To change the caching parameter for a physical disk unit, follow these steps:

1. Run *raidcli getcache* to determine if caching is enabled. If it is, make sure both SPs are powered on; run *raidcli getcrus* to find out.

2. If caching is enabled, disable it with

    `raidcli -d device setcache 0`

3. Wait for the cache memory to be written to disk, which may take several minutes. Use *raidcli getcache* to check cache state ("Disabled").

4. Follow instructions in "Setting Cache Parameters Using the Command-Line Interface," earlier in this chapter.

Changing Caching Parameters Using RAIDGUI

To change the caching parameter for a physical disk unit, follow these steps:

1. To determine if caching is enabled, click the button for the SP for which you want information in the Equipment View or Summary View.

2. In the SP Summary window that appears, click *Cache.* The SP Cache Summary window appears, as shown in Figure 7-4. Do not close this window.

3. If caching is enabled, check the Equipment View to make sure both SPs are powered on.

4. If caching is enabled, disable it in the SP Cache Settings Window; see Figure 7-7 earlier in this chapter.

5. Wait for the cache memory to be written to disk; check status in the SP Cache Summary window.

6. Follow instructions in "Setting Cache Parameters Using RAIDGUI," earlier in this chapter.

Reconfiguring

This chapter explains

- changing LUN configuration
- transferring control of a LUN
- changing LUN parameters

Note: For information on changing caching parameters, see Chapter 7, "Caching."

Changing LUN Configuration

When you change a LUN configuration, you change the bound configuration of a LUN; that is, you add or subtract a disk module, or physically move one or more disk modules to different slots in the chassis.

For information on

- adding disk modules, see "Installing New Disk Modules" in Chapter 5
- binding the new modules into LUNs, see "Binding Disks Into RAID Units" in Chapter 4
- unbinding LUNs, see "Unbinding the Disk" in Chapter 5

For the new configuration, follow the guidelines given in "Planning the Bind" in Chapter 4.

Transferring Control of a LUN

This section explains how to transfer control of a LUN from a primary to a secondary route. It applies only to Challenge RAID storage systems with two SPs. Note that transferring control is also known as changing SP ownership of a LUN.

Use the procedure in this section when the system operator or service person installed a second SP and you want to assign some of the LUNs to the new SP, or when you want to balance your LUNs between two SPs.

Note: Transferring control of a LUN from one SP to the other can affect how the operating system accesses the LUNs. However, any change you make in ownership does not take effect until the Challenge RAID storage system is powered-cycled.

Depending on the type of server, you may want to use the procedure if any of the following failure situations occurs:

- A failed SP has been replaced, and you want to restore control to the SP you have replaced.

- One of the two SCSI-2 adapters or the SCSI bus cable connected to the Challenge RAID storage system fails; you want the working SCSI-2 adapter to access the failed SCSI-2 adapter's LUNs. If a SCSI adapter fails, you see an error message from the operating system specifying a disk storage-system controller (SCSI adapter):

  ```
  device does not respond to selection
  timeout resetting scsi bus
  .....
  ```

- A LUN becomes unavailable to the operating system, indicating a SCSI-2 adapter or cable failure.

- One host in a dual-host configuration has failed, and auto assign is not enabled and you want the working host to access the failed host's LUNs. To enable auto assign, see "Creating LUNs" in Chapter 4.

  Note: The Silicon Graphics IRIS FailSafe™ product includes special failover software that automatically handles storage reassignment in failure situations.

146

This section explains

- using the command-line interface to change LUN ownership
- using RAIDGUI to change LUN ownership

Using the Command-Line Interface to Change LUN Ownership

To change LUN ownership, follow these steps:

1. If necessary, use *raidcli getagent* to determine the target device.
2. Use *raidcli getlun* to determine the LUN whose ownership you want to change.
3. Become root.
4. To transfer control of a LUN to the other SP, use *raidcli chglun*:

    raidcli -d *device* **chglun -l** *lunnumber* **-d1**

    The -d1 parameter changes ownership of the LUN specified with -l.
5. Take precautions for system users, and then power-cycle the Challenge RAID storage system to put the ownership change into effect.

For the complete syntax of *raidcli chglun*, see "chglun" in Appendix B.

The LUN number is required, along with at least one other switch. For complete information on this command, see "chglun," in Chapter 8, "Reconfiguring."

The following example changes the ownership of LUN 3:

```
raidcli -d sc4d2l0 chglun -l 3 -d1
```

Using RAIDGUI to Change LUN Ownership

To transfer control manually of a LUN to the other SP when auto assign is disabled (the recommended setting), follow these steps:

Note: Changing LUN ownership requires some time.

1. In the Summary View window, click the button of the LUN whose ownership you want to change.

    All the modules in the LUN are outlined in black, and the LUN Information window appears.

2. In the LUN Information window, click *Change*. The Change LUN Parameters window appears, as shown in Figure 8-1.
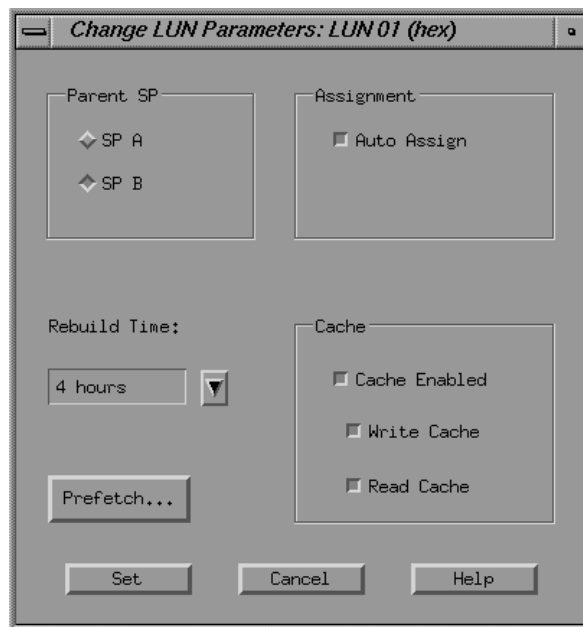
Figure 8-1   Example Change LUN Parameters Window

3. In the Parent SP field, click the button for the SP that you want as the owner of the LUN.

4. Use the fields and buttons in this window to change other parameters as desired. When all parameters are as you want them, click *Set*.

148

5. Verification dialog boxes appear; answer them appropriately to set the ownership change.

6. Take precautions for system users, and then power-cycle the Challenge RAID storage system to put the ownership change into effect.

Changing LUN Parameters

You can change certain LUN parameters for any LUN except a hot spare, provided the LUN is available to the operating system. You can

- change LUN ownership (see "Transferring Control of a LUN" earlier in this chapter)

- reset the rebuild time for RAID-0, -1, -1_0, -3, or -5 LUNs

- enable or disable auto assign for a LUN

- enable or disable read/write caching for a LUN

- read-ahead (prefetch) parameters for any LUN

When you plan changes to LUN parameters, note the following points:

- You cannot change the rebuild time while a LUN is binding or rebuilding.

- You cannot change the write caching enable parameter for a LUN unless storage-system write caching is disabled.

- You cannot change the read caching enable or the prefetch parameters for a LUN unless the read cache for the SP that owns the LUN is disabled.

This section explains how to use the command-line interface and RAIDGUI to change auto assign, caching parameters, and rebuild time.

Note: For information on changing caching parameters, see Chapter 7, "Caching."

Changing LUN Parameters Using the Command-Line Interface

To change parameters for a logical unit, use *raidcli chglun* (as root). To change caching parameters, use *raidcli setcache*. For complete information on both these commands, see Appendix B. Chapter 7, "Caching," has complete information on caching.

The following example disables write caching for the Challenge RAID storage system and disables write caching for LUN 2:

```
raidcli -d sc4d2l0 setcache 0
raidcli -d sc4d2l0 chglun -l 2 -c read
```

The following example changes the rebuild time for LUN 2 to 3 hours:

```
raidcli -d sc4d2l0 chglun -l 2 -r 3
```

Changing LUN Parameters Using RAIDGUI

Follow these steps:

1. In the Summary View, click the button of the LUN whose parameters you want to change.

2. In the LUN Information window that appears, click *Change*. The Change LUN Parameters window appears; see Figure 8-1.

3. To change the write caching or read caching, first disable caching by clicking the checkbox at Cache Enabled so that it is blank. Click the checkboxes at Write Cache and Read Cache as desired and enable caching by clicking the checkbox at Cache Enabled so that it is filled.

    For complete information on setting caching, see "Binding Disks Into RAID Units" in Chapter 4.

4. If desired, change the rebuild time by selecting a different number in the scroll box. For complete information on rebuild time, see "Binding Disks Into RAID Units" in Chapter 4.

    Note: You cannot change the rebuild time while a LUN is binding or rebuilding.

150

5. To change prefetch parameters, click *Prefetch* in the Change LUN
    Parameters window. The LUN Prefetch Parameters window opens, as
    shown in Figure 8-2.
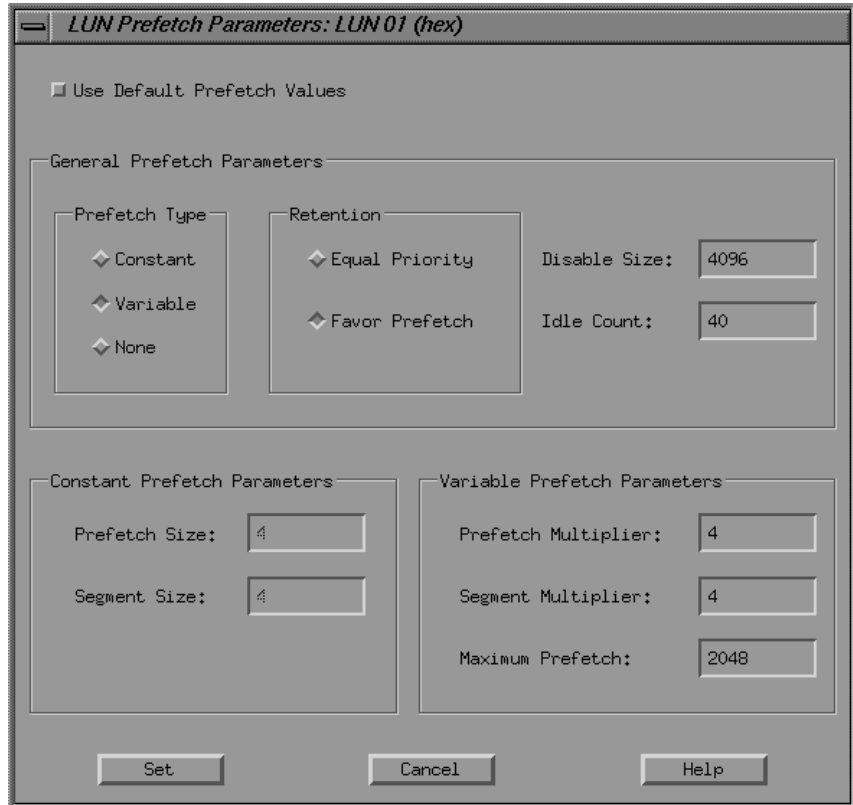
Figure 8-2   LUN Prefetch Parameters Window

Read-ahead caching (prefetching) lets the SP anticipate the data that an
application will request so it can read the data into its read cache before
the data is needed. The SP monitors I/O requests to each LUN for
sequential reads; if it finds any, it prefetches data automatically from
the LUN. Prefetching works best when the host issues many large
sequential read requests to a physical unit.

You can define a specific type of prefetch operation (None, Variable, or Constant) by specifying the values of the following prefetch parameters:

- Prefetch type determines whether to prefetch data of a variable or constant length or to disable prefetching. The default is variable-length prefetching.

- Retention determines whether or not prefetched data has equal or favored priority over host requested data when the read cache becomes full. The default is Favor prefetch, which is suitable for most applications.

- Prefetch Size and Prefetch Multiplier determine the amount of data prefetched for one host read request. For constant-length prefetching, the prefetch size is the number of blocks to prefetch.

   For variable-length prefetching, the prefetch multiplier is the amount of data to prefetch relative to the amount of data requested. For example, if the prefetch size is 8, the amount of data to prefetch is 8 times the amount of data requested.

   The range for the prefetch size is 1-2048 blocks. The range for the prefetch multiplier is 1-32; the default is 4.

- Segment Size and Segment Multiplier determine the size of the segments that make up a prefetch operation. An SP reads one segment at a time from the LUN because smaller prefetch requests interfere less with other host requests.

   The amount of data retrieved during a single read operation (segment size) equals the number of blocks of data requested multiplied by the segment multiplication factor. The amount of data prefetched for each read request equals the number of blocks of data requested multiplied by the prefetch multiplier.

   Note: The segment multiplier value cannot larger than the prefetch multiplier.

   You can specify a segment multiplication factor that is equal to or less than the prefetch multiplier. If you specify a segment multiplication factor that is equal to the prefetch multiplier, prefetch operations are not divided into segments.

For constant-length prefetching, the segment size is the number of blocks of data to prefetch in one read operation from the LUN. For variable-length prefetching, the segment multiplier determines the amount of data to prefetch in one read operation relative to the amount of requested data.

For example, if the segment multiplier is 4 (the default), the segment size is 4 times the amount of data requested.

The range for the segment size is 1 to 254 blocks. The range for the prefetch multiplier is 1 to 32; the default is 4.

- Maximum Prefetch is the maximum number of blocks to prefetch for variable-length prefetching. The range is 1 to 2048;the default is 512 sectors.

6. To use default values, click *Use Default Prefetch Values* if it is not highlighted.

   If you are sure that the applications that access this LUN benefit from different prefetch values, click the buttons and enter values in fields as desired.

7. When you are finished, click *Set* and confirm your choice in the configuration popup that appears.

   Note: The LUN cannot use read caching with the prefetch parameters you specify unless read caching is enabled for the SP that owns it.

Technical Specifications

This appendix lists the technical specifications for the Challenge RAID deskside disk-array storage systems. Table A-1 summarizes specifications for the deskside storage system.

Table A-1   Challenge RAID Deskside Chassis Specifications

Classification	Specification	Value
AC power requirements	Voltage	90 VAC to 264 VAC, single-phase autoranging 47 Hz to 63 Hz
	Current draw	9.0 A max at 100 VAC input
	Power consumption	Apparent power: 900 VAC max True power: 880 W max
Connector	Type	L6-15R/L6-15P
Operating limits	Ambient temperature	10 degrees C to 38 degrees C (50 degrees F to 100 degrees F)
	Relative humidity	20 to 80% noncondensing
	Elevation	2439 m (8000 ft)
	Host dissipation	3168 x 103 J/hr (3000 Btu/hr) max
	Shock	3 g @ 11 ms
	Vibration	0.25 g peak @ 5 Hz to 500 Hz
Nonoperating limits (shipping and storing)	Ambient temperature	-40 degrees C to 65 degrees C (-40 degrees F to 149 degrees F)
	Gradient, maximum	24 C/hr (43.2 degrees F/hr)
	Relative humidity	10 to 90% noncondensing
	Elevation	7625 m (25,010 ft)

Table A-1 (continued)   Challenge RAID Deskside Chassis Specifications

Classification	Specification	Value
Physical	Dimensions	Height: 62.9 cm (24.75 in.) Width: 35.6 cm (14.0 in.) Depth: 76.2 cm (30.0 in.)
	Minimum chassis weight	Chassis with 5 disk modules, 1 SP, 2 VSCs, without packaging: 51.6 kg (114.0 lb)
	Maximum chassis weight	Chassis with 20 disk modules, 2 SPs, 3 VSCs, BBU, without packaging: 83.9 kg (184.5 lb)
	Disk drive module	1.6 kg (3.5 lb)
	Third power supply module	2.4 kg (5.3 lb)
	Second storage-control processor	1.2 kg (2.6 lb)
	Battery backup unit	5.4 kg (12 lb)
Service clearance	Front	81.3 cm (32.0 in.)
	Rear	81.3 cm (32.0 in.)
Buses	External host bus	Differential fast and wide SCSI-2 (synchronous)
	Internal storage system buses	Five single-ended SCSI buses

Table A-2 summarizes specifications for the rack storage system.

Table A-2   Challenge RAID Rack Specifications

Classification	Specification	Value
AC power requirements	Cabinet voltage	200 VAC to 240 VAC, single-phase 47 Hz to 63 Hz
	Current draw per chassis assembly	5.0 A max at 200 VAC input (20.0 A max for 4 chassis assemblies)
	Power consumption per chassis	Apparent power: 1000 VAC max True power: 980 W max
Connector	Type	Domestic: L6-30P/L6-30R International: IEC Plug 309 2-P

Table A-2 (continued)		Challenge RAID Rack Specifications
Classification	**Specification**	**Value**
Operating limits	Ambient temperature	10 degrees C to 38 degrees C (50 degrees F to 100 degrees F)
	Relative humidity	20 to 80% noncondensing
	Elevation	2439 m (8000 ft)
	Heat dissipation per chassis	3168 x 103 J/hr (3000 Btu/hr) max per chassis assembly
	Shock	3 g @ 11 ms
	Vibration	0.25 g peak @ 5 Hz to 500 Hz
Nonoperating limits (shipping and storing)	Ambient temperature	-40 degrees C to 65 degrees C (-40 degrees F to 149 degrees F)
	Gradient, maximum	24 C/hr (43.2 degrees F/hr)
	Relative humidity	10 to 90% noncondensing
	Elevation	7625 m (25,010 ft)
Physical	Cabinet dimensions	Height: 180.34 cm (71.0 in.) Width: 58.42 cm (23.0 in.) Depth: 78.74 cm (31.0 in.)
	Maximum cabinet weight	407.96 kg (899.4 lb): 4 chassis assemblies, each with 20 disk modules, 2 SPs, 3 power supplies, battery backup unit, without packaging
	See Table A-1 for weights of FRUs.	
Service clearance	Front access	81.3 cm (32.0 in.)
	Rear and side access	76.2 cm (30.0 in.)
Buses	External host bus	Differential fast and wide SCSI-2 (synchronous)
	Internal storage system buses	Five single-ended SCSI buses

The raidcli Command-Line Interface

This appendix is a complete guide to the */usr/raid5/raidcli* command (command-line interface, or CLI). It explains the *raidcli* parameters, which are summarized in Table B-1.

Table B-1   *raidcli* Parameters

Parameter	Use
bind	Bind (group) physical disks
chglun	Change parameters for a currently bound group
clearlog	Destroy the contents of the Challenge RAID storage controller's error log
clearstats	Reset statistics logging on the Challenge RAID storage-control processor (SP), setting all log counters to 0
getagent	Get names and descriptions of devices controlled by the SP
getcache	Get information about the storage-system caching environment
getcontrol	Get general system information
getcrus	Display status information on all system components, such as the fan module, battery backup unit, and so on
getdisk	Display status information on one disk module or all disk modules in the system
getlog	Display the SP log, in which errors are stored, or display the head and tail of the log
getlun	Display information about a disk group
setcache	Change information about the Challenge RAID controller's caching environment
unbind	Deconfigure physical disks from their current logical configuration, destroying all data on the logical unit (group)

The *raidcli* command sends Challenge RAID storage management and configuration requests to an application programming interface (API) on the Challenge server. For the *raidcli* command to function, the agent—an interpreter between the command-line interface and the Challenge RAID storage system—must be running.

Note: Although the directory and command are *raidcli*, the command is valid for all RAID levels.

The synopsis of the *raidcli* command is

```
raidcli [-vp] [-d device] parameter [optional_arguments]
```

In this syntax, options and variables have the following meanings:

-v   Enables verbose return.

-p   Parses the *raidcli* command without calling the API. If the string does not parse correctly, an error message is printed to *stderr*; otherwise there is no output.

-d *device*   Target RAID device. Use *raidcli getagent* for a list of RAID devices. This switch must be present for all *raidcli* management and configuration commands unless the environment variable indicates otherwise. This switch overrides an environment variable.

The environment variable RaidAgentDevice is the default value for the device when none is specified with the -**d** flag. If RaidAgentDevice is not set and no -**d** switch is present, an error is generated on all commands that need device information. For example:

```
setenv RaidAgentDevice sc4d2l0
setenv RaidAgentDevice /dev/scsi/sc4d2l0
```

The */dev/scsi* prefix is optional.

bind

The physical disk unit number is also known as the logical unit number, or LUN. The unit is a logical concept but is recognized as a physical disk unit by the operating system. The LUN is a hexadecimal number between 0 and 7.

Unlike standard disks, physical disk unit numbers (LUNs) lack a standard *geometry.* Disk capacity is not a fixed quantity between disk-array LUNs. The effective geometry of a disk-array LUN depends on the type of physical disks in the array and the number of physical disks in the LUN.

Note: Although *bind* returns immediate status for a RAID device, the bind does not complete for 15 to 60 minutes, depending on system traffic. Use *getlun* to monitor the progress of the bind; *getlun* returns the percent bound. When the bind is complete, each disk is noted as "Enabled" in the *getlun* output.

The command *raidcli bind* is used to group physical disks into RAID-0, RAID-1, RAID-1_0, RAID-3, or RAID-5 units or to create a hot spare:

```
raidcli –d device bind raid-type lun-number disk-names [optional-args]
```

Variables in this syntax are explained below.

-d *device*   Target RAID device, as returned by *raidcli getagent.*

raid-type   Choices are

- r0: RAID-0

- r1: RAID-1

- r1_0: RAID-1_0

- r3: RAID-3

- r5: RAID-5

- id; individual disk unit

- hs: hot spare

lun-number   Logical unit number to assign the unit (a hexadecimal number between 0 and 7).

disk-names   Indicates which physical disks to bind, in the format *bd*, where *b* is the physical bus name (a through e; be sure to use lowercase) and *d* is the device number on the bus (0 through 3). For example, a0 represents the device 0 on bus A, and e2 represents device 2 on bus E.

A RAID-0 bind requires a minimum of three (maximum 16) disks.

A RAID-1 bind requires two disks on separate buses.

A RAID-1_0 bind requires a minimum of four disks per bus. The maximum is 16 disks, grouped in sets of four on different buses. A RAID-1_0 bind requires separate buses for each member of the image pair. Select the disks in this order: first disk on first bus, first disk on second bus, second disk on first bus, second disk on second bus, third disk on first bus, third disk on second bus, and so on.

RAID-3 or RAID-5 bind also requires separate buses, each with five disk modules. Legal RAID-3 or RAID-5 bind configurations are

- a0 b0 c0 d0 e0

- a1 b1 c1 d1 e1

- a2 b2 c2 d2 e2

- a3 b3 c3 d3 e3

A hot spare bind requires one disk, which cannot be A0, B0, C0, D0, E0, or A3. The capacity of the hot spare must be at least as great as the capacity of the largest disk module it might replace.

All disk modules in a bind should have the same capacity to fully use the modules' disk space.

The optional arguments are as follows:

Note: You can enable auto-assign in RAIDGUI, but not in the command-line interface.

-r *rebuild-time*   Maximum time in hours to rebuild a replacement disk. Default is four hours; legal values are any number greater than or equal to zero.

A rebuild time of two hours rebuilds the disk more quickly but degrades response time slightly. A rebuild time of zero hours rebuilds as quickly as possible but degrades performance significantly.

If your site requires fast response time and you want to minimize degradation to normal I/O activity, you can extend the rebuilding process over a longer period of time, such as 24 hours. You can change the rebuild time later without damaging the information stored on the physical disk unit.

Caution: The LUN is not fault-tolerant during the rebuild; selecting a short rebuild time minimizes the chance of data loss.

-s *stripe-size*   Number of blocks per physical disk in a RAID-5 stripe. Default is 128; legal values are any number greater than zero.

The smaller the stripe element size, the more efficient the distribution of data read or written. However, if the stripe size is too small for a single host I/O operation, the operation requires accessing two stripes, thus causing the hardware to read and/or write from two disk modules instead of one. Generally, it is best to use the smallest stripe element size that will rarely force access to another stripe. The default stripe element size is 128 sectors. The size should be an even multiple of 16 sectors (8 KB).

Note: This value is ignored for hot spares.

-c *cache-flags*   Values are

- none: no caching

- read: read caching

- write: write caching

- rw: read and write caching

The default is none.

The following example binds disks A1, B1, C1, D1, and E1 into a RAID-5 logical unit with a logical unit number of 3, a four-hour maximum rebuild time, and a 128-block stripe size per physical disk, with read cache enabled:

```
raidcli -d sc4d2l0 bind r5 3 a1 b1 c1 d1 e1 -r 4 -s 128 -c
read
```

The following example binds A2 and B2 into a RAID-1 logical unit with a LUN number of 2 and a four-hour maximum rebuild time, with read cache enabled:

```
raidcli -d sc4d2l0 bind r1 2 a2 b2 -r 4 -c read
```

The following example binds disks A1, B1, C1, and D1 into a RAID-1_0 logical unit with a LUN number of 1, a four-hour maximum rebuild time, and a 128-block stripe size per physical disk, with read cache enabled:

```
raidcli -d sc4d2l0 bind r1_0 1 a1 b1 c1 d1 -r 4 -s 128 -c
read
```

The following example binds A2, B2, C2, D2, and E2 into a RAID-0 logical unit with a LUN number of 3, and a 128-block stripe size per physical disk, with read cache enabled:

```
raidcli -d sc4d2l0 bind r0 3 a2 b2 c2 d2 e2 -s 128 -c read
```

The following example binds disk E3 as a hot spare with a LUN number of 7:

```
raidcli -d sc4d2l0 bind hs 7 e3
```

There is no output for *raidcli* with the *bind* parameter. Errors are printed to *stderr*.

Note: For complete messages, it is recommended that you use the –v option.

chglun

The command *raidcli chglun* is used to change parameters for a LUN:

```
raidcli -d device chglun [-l lun] [-c cache-flags] [-d default-owner]
[-r rebuild-time] [-i idle-thresh] [-t idle-delay-time] [-w write-aside]
[-pf prefetch type] [-rt retention type] [-sc configure-read-cache]
[-sm segment-size/multiplier] [-pm prefetch-size/multiplier] [-mp max-prefetch]
[-pd min-prefetch] [-pc max-IO]
```

Note: Only root can use this parameter.

In this syntax, variables mean the following:

-l *lun*   Logical unit number to be changed

-c *cache-flags*   Values are

- none: no caching

- read: read caching

- write: write caching

- rw: read and write caching

The default is none.

-d *default-owner* Values are

- 1: change SP ownership of LUN

- 0: don't change ownership

If your storage system has dual SPs, see "Dual Interfaces, Load Balancing, and Device Names," in Chapter 4.

-r *rebuild-time*   Maximum time in hours to rebuild a replacement disk. Default is four hours; legal values are any number greater than or equal to 0.

-i *idle-thresh*   Maximum number of I/Os that can be outstanding to a LUN and have the LUN still be considered idle. Used to determine cache flush start time. Legal values are 0 to 255; the default is 0. A value of 255 leaves the previous setting unchanged.

165

-t *idle-delay-time* Amount of time in 100-ms intervals that a unit must be below *idle-thresh* to be considered idle. Once a unit is considered idle, any dirty pages in the cache can begin idle time flushing. Legal values are 0 to 255; the default is 20. A value of 255 leaves the previous setting unchanged.

-w *write-aside* The smallest write request size, in blocks, that will bypass the cache and go directly to the disks. Legal values are 16 to 65535; the default is 2048. A value of 65535 leaves the previous setting unchanged.

-pf *prefetch-type* Values are

- 0: disabled

- 1: constant length

- 2: variable length (default)

Note: See Table B-7 for an explanation of this and the remaining options for this command.

-rt *retention-type* Retention determines whether or not prefetched data has equal or favored priority over host-requested data when the read cache becomes full. Values are

- 0: all data treated equally

- 1: prefetched data retained over requested data (default)

-sc *configure-read-cache*
Values are

- 0: configure read cache according to parameters specified in the command string

- 1: configure read cache according to default settings

-sm *segment-size/multiplier*

Segment size or segment multiplier determines the size of the segments that make up a prefetch operation.

Legal values for segment size are 0 to 255; a value of 255 leaves the previous setting unchanged. Legal values for the segment multiplier are 1 to 32; the default is 4.

- For constant-length prefetching, specify the segment size in blocks of data to prefetch in one read operation from the LUN.

- For variable-length prefetching, specify the segment multiplier to calculate the size of the segment. For example:

```
sgmtblocks = readrequestsize * sgmtmultiplier
```

You can specify a segment size that is equal to or less (but not greater) than the prefetch size. You can specify a segment multiplier that is equal to or less (but not greater) than the prefetch multiplier. If you specify a segment multiplier that is equal to the prefetch multiplier, prefetch operations are not divided into segments.

-pm *prefetch-size/multiplier*

Prefetch size or prefetch multiplier determines the size of the segments that make up a prefetch operation. Legal values for prefetch size are 0 to 2048. Legal values for the prefetch multiplier are 1 to 32; the default is 4.

- For constant-length prefetching, specify the size of the prefetch data in blocks.

- For variable-length prefetching, specify the multiplier to calculate the prefetch length to fetch. For example:

```
prefblocks = readrequestsize * prefmultiplier
```

-mp *max-prefetch*

Maximum number of blocks to prefetch for variable-length prefetching.

- For constant-length prefetching, all prefetches are of the constant length specified by the prefetch size and this field is ignored. Legal values are 0 to 2048; the default is 512.

- For variable-length prefetching, specify the maximum number of blocks to prefetch in response to a read request.

-pd *min-prefetch*

Prefetch Disable specifies the minimum read request size at which the system disables prefetching. Legal values are 0 to 65535; the default is 129. Any value between 65280 and 65535 leaves the previous setting unchanged.

-pc *max-IO*   The maximum number of outstanding I/Os at which the LUN stops prefetching. Legal values are 0 to 100; the default is 40.

The following example changes LUN 3 to perform write caching and rebuild in four hours; it does not change the default owner:

```
raidcli -d sc4d2l0 chglun -l 3 -c write -d 0 -r 4
```

There is no output for *raidcli* with the *chglun* parameter. Errors are printed to *stderr.*

clearlog

The *clearlog* parameter destroys the contents of the Challenge RAID SP's error log. Its syntax is

```
raidcli -d device clearlog
```

Note: You must be root to use this parameter.

This command has no output.

clearstats

The *clearstats* parameter resets statistics logging on the Challenge RAID SP, resetting all log counters to 0. The command *raidcli clearstats* is used to reset statistics logging:

```
raidcli -d device clearstats
```

Note: You must be root to use this parameter.

This command has no output.

getagent

The command *raidcli getagent* is used to display information on devices controlled by the API:

```
raidcli getagent
```

The following is a sample output for one device; normally, the output would give information on all devices:

```
Agent Rev: 1.10
Name: Disk Array
Desc: RAID5 Disk Array
Node: sc4d2l0
Signature:0xf3b51700
Peer Signature: 0x657e0a00
Revision: 7.12.4
```

```
SCSI ID: 1
Prom Rev: 0x0076100
SP Memory: 64
Serial No: 94-7240-808
```

Table B-2 summarizes entries in the *raidcli getagent* output.

Table B-2   Output of *raidcli getagent*

Entry	Meaning
Agent Rev	Revision number of RAID agent
Name	ASCII string found in the agent configuration file that assigns a name to the node being accessed (see node description below)
Desc	ASCII string found in the agent configuration file that describes the node being accessed (see node description below)
Node	The */dev/scsi* entry that the agent uses as a path to the actual SCSI device; this value must be entered by the user for every CLI command (except *getagent*)
Signature	Unique 32-bit identifier for the SP being accessed through Node
Peer Signature	Unique 32-bit identifier for the other SP in the chassis; 0 if no additional SP is present
Revision	Revision of firmware currently running on the SP
SCSI ID	SCSI ID number
Prom Rev	Current PROM revision on the SP
SP Memory	Amount of DDRAM present on the SP
Serial No	12-digit ASCII string that uniquely identifies this subsystem

getcache

If you use storage system caching, you can use the *raidcli getcache* command to get information on cache activity. The information in this command's output, particularly the percentage of cache hits, may help you decide on the most efficient cache page size and whether a physical disk unit really benefits from caching.

The command *raidcli getcache* is used to display cache information:

```
raidcli -d device getcache
```

The following is sample output:

```
Usable Cache: 8
Page size: 2
Cache State: Enabled
Low Watermark; 75
High Watermark: 90
SP A Cache Pages: 2048
SP B Cache Pages: 2047
Unassigned Cache Pages: 0
Read Hit Ratio: 82
Write Hit Ratio: 74
Prct Dirty Cache Pages = 63
Prct Cache Pages Owned = 50
Prct Read Flushes = 10
Prct Write Flushes = 12
SPA Read Cache State enabled
SPB Read Cache State enabled
Minimum system memory allowed 2 MB
Physical memory size of spA = 8 MB
Physical memory size of spB =16 MB
System memory size of spA = 8 MB
System memory size of spB = 8 MB
Read Cache Size of spA = 3 MB
Read Cache Size of spB = 3 MB
Optimized Raid 3 memory size for spA = 2 MB
Optimized Raid 3 memory size for spB = 2 MB
```

171

Table B-3 summarizes entries in the *raidcli getcache* output.

Table B-3   Output of *raidcli getcache*

Entry	Meaning
Usable Cache	Size of cache in megabytes to use for caching, not greater than the SP memory size as displayed by raidcli getagent (see "getagent," earlier in this appendix). Valid values are 0, 8, and 64. Defaults are 0 (user has selected to disable the cache) and 8 (user has selected to enable the cache).
	The command-line interface does not let you specify more memory than you have. If you specify less than you have, the remaining memory is unused.
Page Size	User-specified page size in KB for caching; for example, 2 means a 2 KB page size.
Cache State	Enabled: SP is fully functional.
	Disabled: SP not capable of or configured for caching.
	Synching: SP is synchronizing its cache with the peer SP.
	Enabling: cache is in the process of becoming enabled.
	Quiescing: cache was enabled but system experienced a fault or was disabled by user. The system is now finishing processing outstanding I/O before saving the contents of the cache.
	Frozen: SP has finished quiescing and is now waiting for peer SP to finish quiescing or dumping.
	Dumping: SP is in the process of dumping contents of the cache to a private location on the disk.
	Initing: SP is initializing.
Low Watermark	Percentage of dirty cache pages that, when reached during flush operations, will cause the SP to cease flushing the cache.
	Valid values are 0 through 100; the default is 50, regardless of whether caching is enabled or disabled.

Table B-3 (continued)   Output of *raidcli getcache*

Entry	Meaning
High Watermark	Percentage of dirty cache pages which, when reached during flush operations, will cause the SP to cease flushing the cache.
	Valid values are 0 through 100; the setting must be greater than the low watermark. The default is 75, regardless of whether caching is enabled or disabled.
SP A Cache Pages SP B Cache Pages	Total number of cache pages currently owned by SP A; each page has the cache page size selected when storage-system caching was set up. This number equals the cache size divided by the cache page size, minus space for checksum tables. If the storage system has two working SPs, they divide the total number of pages between them. If an SP is idle for a long period or fails, the active SP may increase its share of pages.
Unassigned Cache Pages	Number of dirty cache pages not owned by A or B. These can occur when a unit is broken and there are no disks to which to flush the dirty pages.
Read Hit Ratio	Percent of read requests to the SP that can be satisfied from the cache without requiring disk access.
Write Hit Ratio	Percent of write requests to the controller that can be satisfied with the cache without requiring disk access. A write cache hit occurs when the SP finds and modifies data in the write cache memory, which usually saves a write operation.
	For example, a hit that occurs when a page is sought in a RAID 5 LUN eliminates the need to read, modify, and write the data. High ratios are desirable because each hit indicates at least one disk access that was not needed.
Prct Dirty Cache Pages	Percentage of pages that have been modified in the SP's write cache, but that have not yet been written to disk. A high percentage of dirty pages means the cache is handling many write requests.
Prct Cache Pages Owned	Percentage of cache pages owned by this controller when pages from other SP are taken into account.

Table B-3 (continued)   Output of *raidcli getcache*

Entry	Meaning
Prct Read Flushes	Percentage of reads to SP that caused the cache to flush a page(s).
Prct Write Flushes	Percentage of writes to SP that caused the cache to flush a page(s).
SP Read Cache State	Enabled or Disabled.
Minimum system memory allowed	AMD-based SP: 2 MB; PowerPC-based SP: 4 MB.
Physical memory size of SP	Number of MB in SP's physical memory.
System memory size of SP	Number of MB in SP's system memory.
Read Cache Size of SP	Number of MB in SP's read cache.
Optimized Raid 3 memory size for SP	N/A

getcontrol

The command *raidcli getcontrol* is used to get general system information:

```
raidcli -d device getcontrol
```

A sample output of this command follows:

```
System Fault LED: OFF
Statistics Logging: ON
System Cache: ON
Max Requests: 23
Average Requests: 5
Hard errors: 0
Total Reads: 18345
Total Writes: 1304
Prct Busy: 25
Prct Idle: 75
System Date: 1/5/1996
Day of the week: Friday
System Time: 12:43:54
read_requests: 145
```

```
write_requests: 34642049
blocks_read: 1633
blocks_written: 363696414
sum_queue_lengths_seen_by_arrivals: 401
arrivals_to_non_zero_queue: 400
hw_flush_on: 0
idle_flush_on: 5314
lw_flush_off: 0
write_Cache_flushes: 98
write_cache_blocks_flushed: 1150
Internal bus 1 busy ticks: 429
Internal bus 1 busy ticks: 429
Internal bus 2 busy ticks: 429
Internal bus 2 busy ticks: 429
Internal bus 3 busy ticks: 429
Internal bus 3 busy ticks: 429
Internal bus 4 busy ticks: 429
Internal bus 4 busy ticks: 429
Internal bus 5 busy ticks: 429
Internal bus 5 busy ticks: 429
Controller busy ticks: 1403792
Controler idle ticks: 3775334
```

In the *getcontrol* output,

- statistics logging is always turned on by the agent

- hard errors are those returned to the host

- total reads and writes are the totals as seen by the SP

- system time is in 24-hour format

175

getcrus

The command *raidcli getcrus* is used to display state information on every field-replaceable unit in the Challenge RAID storage system except disks in the disk modules:

```
raidcli -d device getcrus
```

Sample output of this command follows.

```
FANA State: Present
FANB State: Present
VSCA State: Present
VSCB State: Present
VSCC State: Present
SPA State: Present
SPB State: Present
BBU State: Present
```

Table B-4 interprets items in this output. Values for all entries of the output are Present or Not Present, except as noted.

Table B-4   Output of *raidcli getcrus*

Output	Meaning
FANA, FANB	Fan banks A and B.
VSCA	Power supply (voltage semi-regulated converter).
VSCB	Optional second power supply.
SPA	Storage control processor.
SPB	Optional second storage-control processor.
BBU	Battery backup unit, which has three states: Faulted (removed), Charging, and Present (fully charged or charging).
	If the battery backup unit takes longer than an hour to charge, it shuts itself off and transitions to the "Faulted" state.

getdisk

The command *raidcli getdisk* is used to display information on all disks in the system, whether bound or not:
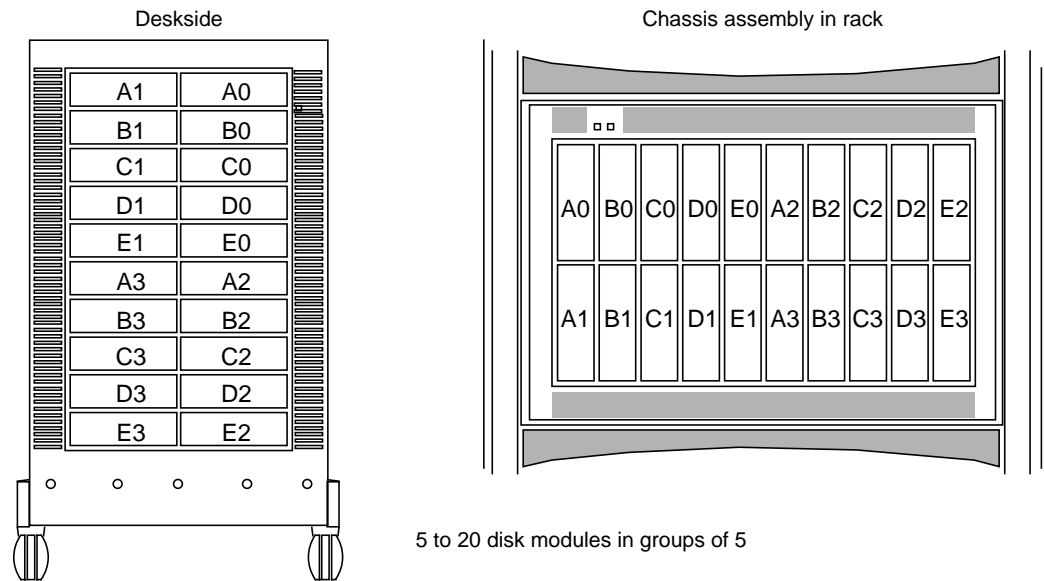
```
raidcli -d device getdisk
```

For information on a particular disk, use

```
raidcli -d device getdisk [diskposition]
```

In this command, *diskposition* has the format *bd*, where *b* is the bus on which the disk is located (a through e) and *d* is the device number (0 through 3). Figure B-1 diagrams disk module locations.

Figure B-1   Disk Module Locations

For example, the following command retrieves information for disk A0:

```
raidcli -d sc4d2l0 getdisk a0
```

A sample output of this command follows.

```
A0 Vendor Id: <manufacturer>
Ao Product Id: <part number>
A0 Lun: 0
A0 State: Bound and Not Assigned
A0 Hot Spare: NO
A0 Prct Rebuilt: 100
A0 Prct Bound: 100
A0 Serial Number: 00019699
A0 Capacity: 0x000f42a8
A0 Private: 0x00009000
A0 Bind Signature: 0x1c4eb2bc
A0 Hard Read Errors: 0
A0 Hard Write Errors: 0
A0 Soft Read Errors: 0
A0 Soft Write Errors: 0
A0 Read Retries: 0
A0 Write Retries: 0
A0 Remapped Sectors: 0
A0 Number of Reads: 1007602
A0 Number of Writes: 1152057
```

Table B-5 interprets items in this output.

Table B-5   Output of *raidcli getdisk*

Output	Meaning
Vendor Id	Manufacturer of disk drive
Product Id	Part number of disk
Lun	Logical unit number to which this disk is bound

Table B-5 (continued)   Output of *raidcli getdisk*

Output	Meaning
State	Removed: disk is physically not present in the chassis or has been powered off
	Off: disk is physically present in the chassis but is not spinning
	Powering Up: disk is spinning and diagnostics are being run on it
	Unbound: disk is healthy but is not part of a LUN
	Bound and Not Assigned: disk is healthy, part of a LUN, but not being used by this SP
	Rebuilding: disk is being rebuilt
	Enabled: disk is healthy, bound, and being used by this SP
	Binding: disk is in the process of being bound to a LUN
	Formatting: disk is being formatted
Hot Spare	YES or NO
Prct Rebuilt	Percentage of disk that has been rebuilt
Prct Bound	Percentage of disk that has been bound
Serial Number	Serial number from disk inquiry command
Capacity	Actual disk capacity in blocks
Private	Amount of physical disk reserved for private space
Bind Signature	Unique value assigned to each disk in a logical unit at bind time
Hard Read Errors	Number of hard errors encountered on reads for this disk
Hard Write Errors	Number of hard errors encountered on writes for this disk
Soft Read Errors	Number of soft errors encountered on reads for this disk
Soft Write Errors	Number of soft errors encountered on writes for this disk
Read Retries	Number of retries occurring during reads
Write Retries	Number of retries occurring during writes
Remapped Sectors	Number of sectors that have been remapped
Number of Reads	Number of reads this disk has seen
Number of Writes	Number of writes this disk has seen

getlog

The SP maintains a log of event messages in its memory. These events include hard errors, startups, and shutdowns involving disk modules, fans, SPs, power supplies, and the battery backup unit. Periodically, the SP writes this log to disk to maintain it when SP power is off. The log can hold over 2,000 event messages; it has a filter feature that lets you select events by device or error message.

The event messages are in chronological order, with the most recent messages at the end. The command *raidcli getlog* is used to display the entire log:

```
raidcli -d device getlog [+N | -N] [-v]
```

The **+N** argument displays the newest *n* entries in the log, starting with the oldest entry:

```
raidcli -d device getlog +N
```

The **-N** argument displays the oldest *n* entries in the log, starting with the oldest entry:

```
raidcli -d device getlog -N
```

The following is a possible output of the command `raidcli getlog +5`:

```
12/17/95 09:59;51 A3: (A07) Cru Powered Down [0x47]
12/17/95 09:59;51 A3: (608) Cru Ready [0x0]
12/17/95 09:59;51 A3: (603) Cru Rebuild Started [0x0]
12/17/95 09:59;51 A3: (604) Cru Rebuild Complete [0x0]
12/17/95 09:59;51 A3: (602) Cru Enabled [0x0]
```

These entries show that a field-replaceable unit (disk module, fan unit, battery backup unit, SP) has failed, been replaced, been rebuilt, and been enabled.

At the tail of each log entry is an error code in brackets that gives diagnostic information when it is available. Table B-6 summarizes the error codes.

Table B-6   *getlog* Error Codes

Error Code	Explanation
0	No error
1	Invalid command-line parameters
2	Too few parameters
3	Too many parameters
4	Invalid bind type
5	Invalid LUN number
6	Invalid rebuild time
8	Invalid number of disks in *bind* command
9	Valid disk names are of format a0, b1, ... e3, etc.
10	Invalid stripe size
11	Invalid disk name
12	Invalid cache flags
13	Invalid default owner
14	Incorrect number of *chglun* parameters
15	Unable to determine name of target host machine
16	Enable-disable flag invalid
17	Invalid usable cache size
18	Invalid page size
19	Invalid watermark value
20	High watermark less than low watermark
21	No device name listed
22	Invalid idle threshold

Table B-6 (continued)   *getlog* Error Codes

Error Code	Explanation
23	Invalid idle delay
24	Invalid write aside size
25	Disks must be on separate buses for bind
26	The agent was unable to configure any devices
27	LUN does not exist
28	LUN already exists
29	Cannot get current working directory for *firmware* command
50	Agent encountered an error during SCSI execution
51	Agent encountered an error during serial port execution
52	Agent returned an operating system error
53	Agent returned an internal agent error code
54	Cannot communicate with agent
55	RPC error encountered

More detailed descriptions of the above errors can be seen when verbose mode is enabled (-v).

getlun

The command *raidcli getlun* is used to display information on a logical unit and the components in it. The syntax is

```
raidcli -d device getlun lun-number
```

The following example displays information about LUN 0:

```
raidcli -d sc4d2l0 getlun 0
```

Note: If numbers in the output look unreasonable, particularly for numerical statistics such as Queue Max, use *clearstats* to reset all statistics.

Below is a truncated output for a RAID-5 group of five disks.

Note: Information on individual disks is not displayed unless statistics logging is enabled with *raidcli setstats*. See "getcontrol" earlier in this appendix.

```
Prefetch size multiplier = 4
Segment size/multiplier = 4
Maximum prefetch = 512
Prefetch Disable = 129
Prefetch idle content = 40

Variable length prefetching

Prefetch data retained

Read cache configured according to specified parameters
Read Histogram[1] 194560Write Histogram[1] 1945 60
Write Histogram[6053104] 1945 60
Write Histogram[10186] 1945 60
Write Histogram[6053103] 1945 60
Write Histogram[12109245] 1945 60
Write Histogram[0] 1945 60
Write Histogram[0] 1945 60
Write Histogram[0] 1945 60
Write Histogram[0] 1945 60
Write Histogram[0] 1945 60
Read Histogram[5] 1945 60
Read Histogram[0] 1945 60
Read Histogram[2] 1945 60
Read Histogram[2] 1945 60
Read Histogram[0] 1945 60
Read Histogram[0] 1945 60
Read Histogram[0] 1945 60
Read Histogram[0] 1945 60
Read Histogram[0] 1945 60

Read histogram overflows 0
Write Histogram[1] 194560
Write Histogram[6053104] 194560
Write Histogram[10186] 194560
Write Histogram[6053103] 194560
Write Histogram[12109245] 194560
Write Histogram[0] 194560
Write Histogram[0] 194560
Write Histogram[0] 194560
```

```
                         Write Histogram[0] 194560
                         Write Histogram[0] 194560

                         Write histogram overflows 0
                         Read requests 10
                         Write Requests 24225638
                         Blocks read 61
                         Blocks written
                         Read cache hits 4
                         Read cache misses 6
                         Prefetched blocks 51
                         Unused prefetched blocks 0
                         Write cache hits 24225043
                         Forced flushes 0

                         Type: RAID5
                         Stripe size: 128
                         Capacity: 0x10000
                         Current owner: YES
                         Auto-trespass: Disabled
                         Auto-assign; Enabled
                         Write cache: Disabled
                         Read cache: Disabled
                         Idle Threshold: 0
                         Idle Delay Time: 20
                         Write Aside Size: 2048
                         Default Owner: YES
                         Rebuild Time: 0
                         Read Hit Ratio: 0
                         Write Hit Ratio: 0
                         Prct Reads Forced Flushed: 0
                         Prct Writes Forced Flushed: 0
                         Prct Rebuilt: 100
                         Prct Bound: 100

                         A0 Enabled
                         A0 Reads: 62667
                         A0 Writes: 29248
                         A0 Blocks Read: 3212517
                         A0 Blocks Written: 471642
                         A0 Queue Max: 26
                         A0 Queue Avg: 1
                         A0 Avg Service Time: 14
                         A0 Prct Idle: 100
                         A0 Prct Busy: 0
                         A0 Remapped Sectors: 0
```

```
A0 Read Retries: 50
A0 Write Retries: 0

B0 Enabled
B0 Reads: 66946
[etc.]

C0 Enabled
C0 Reads: 69342
[etc.]

D0 Enabled
D0 Reads: 68558
[etc.]

E0 Enabled
Eo Reads: 69721
[etc.]
```

Table B-7 summarizes entries in the *raidcli getlun* output.

Table B-7　Output of *raidcli getlun*

Entry	Meaning
Prefetch size/multiplier	For prefetching (read-ahead caching), prefetch size and multiplier determine the amount of data prefetched for one host read request. For constant-length prefetching, the prefetch size is the number of blocks to prefetch. The range for the prefetch size is 1-2048 blocks. The range for the prefetch multiplier is 1-32; the default is 4. For more details, see "Changing LUN Parameters" in Chapter 8.
Segment size/multiplier	Segment size and multiplier determine the size of the segments that make up a prefetch operation. An SP reads one segment at a time from the LUN because smaller prefetch requests interfere less with other host requests. The range for the segment size is 0-254 blocks. The range for the prefetch multiplier is 1-32; the default is 4.
Maximum prefetch	Maximum number of blocks to prefetch for variable-length prefetching. The range is 1-2048; the default is 512.
Prefetch Disable	Number read request sectors that disables prefetch; this number is so large that prefetching data would not be beneficial; for example, if the amount of requested data is equal to or greater than the size of the read cache. The range is 1-65279; the default 129.

Table B-7 (continued)   Output of *raidcli getlun*

Entry	Meaning
Prefetch idle content	Leave this parameter set to 40 (the default).
Variable/Constant length prefetching	Prefetch type: determines whether to prefetch data of a variable or constant length or to disable prefetching. The default is variable-length prefetching.
Prefetch data retained	Retention determines whether or not prefetched data has equal or favored priority over host requested data when the read cache becomes full. The default is Favor prefetch, which is suitable for most applications.
Type	RAID0: nonredundant individual access array RAID1: mirrored pair RAID10: mirrored RAID-0 group (RAID-1_0) RAID3: parallel access array RAID5: individual access array DISK: individual disk unit SPARE: hot spare
Stripe Size	Sectors per disk per stripe with which the unit was bound
Capacity	Number of sectors total for use by user
Current owner	YES if this SP owns the unit; NO if it does not Set with *chglun*; see "Dual Interfaces, Load Balancing, and Device Names" in Chapter 4
Auto-trespass	Always Disabled
Auto-assign	Enabled or Disabled; Disabled is recommended
Write Cache	Enabled means this LUN is write caching; otherwise, Disabled
Read Cache	Enabled means this LUN is read caching; otherwise, Disabled
Idle Threshold	Maximum number of I/Os outstanding; used to determine cache flush start time; set with *chglun*
Idle Delay Time	Amount of time in 100-ms intervals that unit is below idle threshold; set with *chglun*
Write Aside Size	Smallest write-request size in blocks that can bypass the cache and go directly to the disk; set with *chglun*

Table B-7 (continued)   Output of *raidcli getlun*

Entry	Meaning
Default Owner	YES if this SP is the default owner (not necessarily current owner) of this LUN, otherwise, NO
Rebuild Time	Amount of time in hours in which a rebuild should be performed. 0 means rebuild as fast as possible, but means a degradation in host I/O performance
Read Hit Ratio	Percentage of read requests to the controller that can be satisfied from the cache without requiring disk access
Write Hit Ratio	Percentage of write requests to the cache that can be satisfied with the cache without requiring a disk access
Prct Reads Forced Flushed	Percentage of read requests that flushed the cache
Prct Writes Forced Flushed	Percentage of write requests that flushed the cache
Prct Rebuilt	Percentage complete during a rebuild
Prct Bound	Percentage complete during a bind
Diskname State	Enabled, Binding, etc. (same as for *getdisk*)
Diskname Reads	Total number of reads this disk has done
Diskname Writes	Total number of writes this disk has done
Diskname Blocks Read	Total number of blocks this disk has read
Diskname Blocks Written	Total number of blocks this disk has written
Diskname: Queue Max	Maximum number of I/Os queued up to this drive
Diskname: Queue Avg	Average number of I/Os queued up to this drive
Diskname: Avg Service Time	Average service time in milliseconds
Diskname: Prct Idle	Percentage of time disk is not servicing request
Diskname: Prct Busy	Percentage of time disk is servicing request
Diskname: Remapped Sectors	Number of remaps that have occurred on this disk
Diskname Read Retries	Number of read retries that have occurred on this disk
Diskname Write Retries	Number of write retries that have occurred on this disk

setcache

The cache parameters you specify for the entire storage system are the cache size of 8 or 64 MB, depending on the amount of memory the SP has, and the cache page size, as 2, 4, 8, or 16 KB.

The command *raidcli setcache* is used to set up caching:

```
raidcli -d device setcache enable | disable [-u usable] [-p page]
[-l low_watermark] [-h high_watermark]
[-sma system-memory-SPA] [-smb system-memory-SPB]
[-sta read-cache-state-SPA] [-stb read-cache-state-SPB]
[-rca read-cache-size-SPA] [-rcb read-cache-size-SPB]
```

In this syntax, variables have the following meanings:

enable | disable

> Enter 1 to enable write caching or 0 to disable write caching.

> **Note:** When enable/disable is the only option used with *setcache*, 1 enables system read and write cache and 0 disables system read and write cache.

> Using *setcache* 1 with options takes 45 seconds while cache states are disabled; cache dumping might take some time.

-u *usable*　Size in megabytes to use for caching, not greater than the SP memory size as displayed by *raidcli getagent* (see "getagent," earlier in this appendix). Valid values are 0 through 64 in increments of 1 MB.

> The command-line interface does not let you specify more memory than you have. If you specify less than you have, the remaining memory is unused.

> **Note:** For mirrored caching, both SPs must have the same amount of cache memory in order for caching to be preserved in the event of shutdown or other power loss.

-p *page*　Size in KB of pages into which to partition the cache. Valid sizes are 2, 4, 8, and 16. The default is 2, regardless of whether caching is enabled or disabled.

-l *low*　Percentage of cache full that discontinues flushing. Valid values are 0 through 100; the default is 50, regardless of whether caching is enabled or disabled.

-h *high*   Percentage of cache full that initiates flushing. Valid values are 0 through 100; the setting must be greater than the low watermark. The default is 75, regardless of whether caching is enabled or disabled.

-sma *system-memory-SPA*, -smb *system-memory-SPB*
                   Set the system memory on SP A or B, respectively.

                   Note: This option reboots each SP in the system. Rebooting takes several minutes, during which *raidcli getagent* is the only command accepted. When the reboot completes, stop and restart the agent before any *raidcli* commands are accepted.

-sta *read-cache-state-SPA*, stb *read-cache-state-SPB*
                   Set the read cache state for SP A or B, respectively: 0 to disable, 1 to enable.

-rca *read-cache-size-SPA*, -rcb *read-cache-size-SPB*
                   Set the read cache size in MB for SP A and B, respectively.

Note: The sum of an SP's write cache size, system memory size, and read cache size must be less than or equal to the SP's physical memory size.

This command has no output.

You can change the cache size, the cache page size values, or the type of caching for any physical disk unit without affecting the information stored on it. Follow these steps:

1. Disable the cache:

    raidcli -d *device* **setcache 0**

2. Wait for the cache memory to be written to disk, which may take several minutes.

3. Reassign cache size and re-enable caching:

    raidcli -d *device* **setcache 1 -u 64 [parameters]**

Note: Before changing a cache parameter, you must always disable the cache.

The following example enables the system cache with an 8 MB cache partitioned into 8 KB pages, with a 50% low watermark value and a 75% high watermark value:

```
setcache -d sc4d2l0 1 -u 8 -p 8 -l 50 -h 75
```

setstats

The *setstats* parameter sets statistics logging:

```
raidcli setstats [-on | -off]
```

If no optional switches are present, this command returns the current state of statistics logging.

You must be root to use this parameter.

The command *raidcli setstats -on* enables statistics logging; *raidcli setstats -off* disables it. With optional switches, this command has no output.

If statistics logging is off, certain options in other commands are disabled. The option descriptions for these other commands specify whether statistics logging must be on for their options to be valid.

unbind

The *unbind* parameter deconfigures physical disks from their current logical configuration into LUNs.

Caution: This parameter destroys all data on the LUN (disk group).

You must be root to use this parameter. The command prompts you for verification before issuing the unbind, unless you override it with the -o flag.

The command *raidcli unbind* is used to unbind disks:

```
raidcli -d device unbind lun-number [-o]
```

In this syntax, variables and options mean:

lun-number   Number of the logical unit (LUN) to deconfigure.

-o   When *raidcli unbind* is entered, a prompt appears asking the user for verification before the unbind is issued. This flag disables this prompt.

This command has no output.

The following example destroys LUN 3 and frees its disks to be reconfigured, with no prompting to the user:

```
unbind -d sc4d2l0 3 -o
```

Storage-Control Processor Event-Log Error Codes

This appendix lists the hexadecimal error codes and messages that can appear in the storage-control processor unsolicited event log. The messages are given in numerical order.

SP error codes fall into these groups:

*   The 0x600-series codes are informational codes that do not indicate an error condition.

*   The 0x900-series and 0xa00-series codes indicate a serious error condition.

The 0x800-series codes indicate transient events, typical of mass storage systems, from which the storage system recovers. These messages are not included in this appendix.

Many 0x801 messages, however, indicate a problem with either a disk module or SCSI bus. If these messages refer to the same disk module, that disk module is reaching the end of its useful life. If they refer to disk modules on the same SCSI bus, the SCSI bus or components that affect the bus are becoming problematic.

```
0x601 (SP Powerup; Rev 0x%08x)
```

> The system is coming up.

```
0x602 CRU Enabled
```

> The specified disk module has been enabled and is ready for use. This message appears after you rebuild or register the LUN to which the module belongs.

0x603 CRU Rebuild Started

>The storage system started rebuilding the RAID 5, 3, 1, or 1/0 LUN to which the disk module belongs, or the disk module is a hot spare.

0x604 CRU Rebuild Complete

>The storage system has finished rebuilding a RAID 5, 3, 1, or 1/0 LUN.

0x608 CRU Ready

>The specified drive module is powered on and ready for binding or registering the LUN to which it belongs.

0x610

>The disk module could not be physically formatted, and thus cannot be used in the storage system. Make sure the disk module is a valid model. If the model is valid, then consult your service provider for recovery steps.

0x621

>The storage system has begun the background checkpoint verification of the accuracy and completeness of the disk module parity check data. This message may appear after you replace an SP or transfer control of LUNs from one SP to another.

0x622

>The storage system has completed background checkpoint verification of the accuracy and completeness of the parity check data in a RAID 5, 3, 1/0 or 1 LUN.

0x630 Fan Installed

>The storage system has detected that a fan module has been installed or replaced.

0x631

>The storage system has detected that a VSC has been installed or replaced.

0x633

> The storage system has detected an increase in a fan's speed, perhaps because temperature rose or another fan failed.

0x634

> Speed of a fan has returned to normal.

0x635

> Virtual sector data error. The storage system has detected a data inconsistency in a disk sector.

0x636

> BBU was removed from the storage-system chassis.

0x637 BBU Recharging

> BBU is recharging.

0x638 BBU Enabled

> BBU has become ready.

0x639 System Cache Enabled

> Storage-system cache has become ready.

0x640

> The storage system has finished reconstructing the disk mirror.

0x641

> Background rebuild operation has aborted before it was complete.

0x643 SP Initializing

> The storage system began formatting this disk module.

`0x644 SP Inserted`

> A peer (second) SP was detected in this chassis.

`0x650 CRU Signature Error`

> CRU signature error occurred.

`0x654 Cache Dumping`

> The storage system has started dumping the storage-system cache to the vault disks.

`0x657 Cache Dumping Completed`

> The storage system has finished dumping the system cache to the vault disks.

`0x658`

> Storage-system caching was enabled by the storage system or system operator.

`0x659 Cache Disabled By User`

> Storage-system caching was disabled by the storage system or system operator. The storage system disables caching if the BBU is not fully charged or an SP, vault disk, or fan fails; see error 0x908.

`0x660 AC Power Failure`

> After an AC power failure, the storage system dumped the cache to the vault and turned off power to the BBU (it does this to minimize drain on the BBU).

`0x663`

> BBU self-test started.

`0x664 Cache Disabled By BBU Test`

> In preparation for its weekly BBU test, the storage system disabled caching. This message is followed by a 0x663 message.

```
0x665
0x666
```

> The storage system detected change to a single SP (0x665) or dual SP (0x666) configuration.

```
0x667
0x668
```

> Host cache is recovering or has recovered. The storage system started recovering the cache (0x667) and finished recovering (0x668).

```
0x901 Hard SCSI Bus Error
```

> An abnormal SCSI bus or disk module event was detected and could not be cleared through retry operations. This error message is often followed by an explicit CRU message as in 0xa07.

```
0x903 Fan Removed
```

> The fan module shut down or was removed.

```
0x904 VSC Shutdown-Removed
```

> A VSC unit has been shut down or removed from the storage system.

```
0x905 Chassis overtemperature
```

> The storage system found internal temperature too high. It tries to correct an overtemperature condition by increasing fan speed. Check for any obvious problems, such as obstruction of cooling vents or excessive room temperature.

0x906 Unit Shutdown

> A failure in a CRU (which may be a fan or disk module), has made further access to the LUN impossible. If this unit has redundant CRUs (for example, it is a RAID 5 LUN), a failure in two CRUs is needed to produce this error. The SP shut down the LUN and the server can no longer access it.
>
> If this message appears along with the 0x905 or 0xa06 message, replacing a defective fan module may restore access to the LUN. If the problem is with disk modules, *do not* replace the disk modules; instead call your service provider.

0x907 Fatal Firmware Error

> A fatal firmware error has occurred; as a result, the program running in the SP has reset the SP. The SP was restarted and continued normally. Consult your service provider.

0x908 Fault - Cache Disabling

> The storage system is disabling caching because of a system fault. The problem might be that
>
> • the BBU is not ready (not present and fully charged)
>
> • one or more vault disks is missing or being rebuilt
>
> • a fan fault occurred
>
> • an SP failed
>
> To recover, either identify the problem and fix it or wait for the storage system to fix it (for example, wait for the BBU to reach full charge or for the vault disks to be rebuilt). When the fault no longer exists, the storage system automatically re-enables storage-system caching.

```
0x909 Vault Dump Failure
```

A fault caused the storage system to try and dump the vault. The cache dump failed because two or more vault disks are missing or have failed.

Try replacing one or more disk modules in the vault (for 20 slot, A0, B0, C0, D0, or E0; for 10 slot, A0-A4). A power failure, or double SP failure, while the vault is failed and the caching is enabled makes any LUN that has pages in the cache inaccessible; for any such LUN, you need to replace the bad modules, unbind, rebind, software format, and load the lost data from backup. At system power-on, error 0x90A occurs for the inaccessible LUN(s).

```
0x90A Can't Assign - Cache Dirty
```

The storage system cannot enable caching for the LUN or the storage system because the cache contains modified unwritten (dirty) pages for the LUN that it cannot write to the LUN. This error condition can happen, for example, if two disk modules failed in a LUN that was using caching. This error may occur on power-on after the error condition that caused message 0x909.

Make sure there are usable disk modules in all vault slots (20-slot chassis, A0, B0, C0, D0, or E0; 10-slot chassis, A0-A4) and cycle power to restore the cache. If the vault is not the problem, you may need to unbind the LUN to which the dirty pages are destined. Then, you can enable caching for the LUN and the storage system. Sooner or later, you must replace the failed module(s), rebind and soft format the failed LUN, and reload from backup.

0x90B Cache Unit Failed

> The storage system cannot define the cache because the existing cache contains modified unwritten (dirty) pages.

> This error can occur if you try to change cache parameters while the caching is active; if so, disable caching, wait for caching to be disabled, and retry. If the problem does not result from changing cache parameters, check for one or more failed LUNs and if you find one, fix it. If that is not the problem, you may need to unbind the LUN(s) to which the unwritten pages belong; the ID of the LUN(s) is part of the accompanying 0x90A error message.

0x90C Image Larger Than Memory

> The cache was dumped to the vault, but cannot be restored to SP memory because an SP has too little memory to accept the cache image. This can happen if an SP fails and you replace it with an SP that has less memory than the one you removed.

> To recover, remove the SP that has the inadequate amount of memory, insert the correct amount of memory on it, and reinsert it.

0x90D BBU Removed

> A BBU failed or was removed. The cache is dumped to the vault; caching is disabled; and the cache is flushed to disk. Caching cannot be enabled until the problem is corrected either by replacing the BBU, if it failed, or by reinstalling it. When the fault is fixed, caching is re-enabled automatically.

0x90E BBU Disabled, Says Ready

> BBU test was unable to turn off the BBU. The BBU is probably faulty. The cache is dumped to the vault; caching is disabled; and the cache is flushed to disk. Caching cannot be enabled until the problem is fixed. Replace the BBU. When the fault is fixed, caching is re-enabled automatically.

200

```
0x90F Host Cache Recovered With Errors
```

A non-mirrored cache recovery failed to recover the cache
pages for some, but not all, cached LUNs. It does not apply
to a storage system with two SPs. Contact your service
provider.

```
0x910 Host Cache Recovery Failed
```

A non-mirrored cache failed to recover information for all
cached LUNs. It does not apply to a storage system with
two SPs. Contact your service provider.

```
0x920 (Hard Media Error)
```

The sector remap failed. The SP could not remap a defective
sector on the drive module. This is a hard media error. The
disk's remap area may be full. The disk is unusable; replace
it. If the disk was an individual unit or part of an
unmirrored RAID-0 group, you will need to bind the
replacement disk into a replacement, soft format it, and
reload its filesystem's data from backup.

```
0x921 Host Vault Load Failed
```

The SP encountered errors while trying to load the cache
image from disk. This message may indicate multiple disk
failures. Probably any LUN with write-cached pages will be
inaccessible and must be unbound. To identify such a LUN,
look for a message 90A "Can't Assign" and read that
message in this table.

```
0x922 Host Vault Load Inconsistent
```

The SP found inconsistencies in the cache image on disk.
This may indicate a failure or abort of the cache dump.
Probably any LUN with write-cached pages will be
inaccessible and must be unbound. To identify such a LUN,
look for a message 90A "Can't Assign" and read that
message in this table.

`0x923 Host Vault Load Failed Bitmap Ok`

> The SP successfully read the control portion of the cache image on disk, but found the data portion to be incomplete. This means that a failure or abort occurred during the cache dump. Probably any LUN with write-cached pages will be inaccessible and must be unbound. To identify such a LUN, look for a message 90A "Can't Assign" and read that message in this table.

`0x924 Host Vault Disks Scrambled`

> The SP found that the order of the vault disks containing the cache image was different from their order when the cache image was dumped to disk. This means the disks were swapped at power-off. You must restore the disks to their original order before the SP can load the cache image.

`0xa02 Failed SCSI Bus`

> An internal SCSI bus has failed. The CRU number displayed corresponds to the bus number (A0 means bus A, B0 means bus B, and so on). The failure resulted from a bad cable or cable connection, bad terminator, bad SCSI chip on an SP, or a bad device. All disk modules on that internal bus are now inaccessible by the SP. A RAID 5, 3, 1, or 1/0 LUN, or software mirror can continue if the other disk module(s) are on other internal buses.

> It is unlikely that the other SP (if any) will be able to use the bus. Call your Silicon Graphics service provider.

`0xa05 NOVRAM Uninitialized`

> The nonvolatile memory on the SP is not initialized. No server I/O operations can occur. Consult your service provider.

0xa06 Chassis Shutdown

A second VSC failure has occurred, ora fan module has been inoperative for more than two minutes. The SP is powering off all modules in the chassis. Someone must correct the problem—perhaps by inserting a new fan module—before powering on again.

0xa07 CRU Powered Down

The specified disk module has been powered off by the SP, has failed, or has been removed from the chassis.

0xa08 Data Base Sync Error

The SP cannot determine the correct virtual configuration of all LUNs in the storage system. Some LUNs may be unusable. Contact your service provider.

0xa09 Drive Too Small

For a redundant LUN, a replacement disk module was inserted, but it has a smaller capacity than the other disk module(s) in the LUN. The rebuild operation cannot begin until someone moves the replacement disk module and inserts a module of the correct size.

0xa11 SP Removed

The other SP in this chassis has failed. You can force the working SP to take over the failed SP's LUNs via the secondary route.

0xa12 Hard Cache Memory Error

The storage system detected a hard error in SP memory in a storage system with non-mirrored caching. This error does not apply to a storage system with two SPs.

Index

F

fan module
  closing, 72
  opening, 71
  replacing, 120
  status, 47, 53
field-replaceable unit. *See* component
filesystem capacity and RAID, 74

G

getagent, 42-43, 169-170
getcache, 134-135, 171-174
getcontrol, 43, 174-175
getcrus, 46-47, 176
getdisk, 43-46, 177-179
getlog, 47-48, 180-182
getlun, 91, 182-187
graphical user interface.*See* RAIDGUI
grouping disk modules. *See bind*

H

high availability, 18, 88, 90
.hosts file, 50
hotline, 102, 105, 113, 119
hot spare, 22-23
  binding, 78, 161

I

idle
  delay time, 166
  threshold, 165
individual disk unit, 22
  binding, 78
IRIX version, xiii

L

Licensed Internal Code, 63
logical unit number, 76
  See also LUN
log. *See* event log, SP event log
LUN, 76
  button, 93
  changing configuration, 145
  changing parameters, 149-153
  getting information, 91-98
  maximum number, 78, 87
  number, assigning, 78, 87
  ownership, 88, 146, 149
    load balancing, 99
  See also bind
  transferring control, 146-149

M

model number for disk module, 10
multiple CHALLENGE RAID storage systems, 4

O

operation, 39-72

Tell Us About This Manual

As a user of Silicon Graphics documentation, your comments are important to us. They help us to better understand your needs and to improve the quality of our documentation.

Any information that you provide will be useful. Here is a list of suggested topics to comment on:

- General impression of the document
- Omission of material that you expected to find
- Technical errors
- Relevance of the material to the job you had to do
- Quality of the printing and binding

Please include the title and part number of the document you are commenting on. The part number for this document is 007-2532-006.

Thank you!

Three Ways to Reach Us

The **postcard** opposite this page has space for your comments. Write your comments on the postage-paid card for your country, then detach and mail it. If your country is not listed, either use the international card and apply the necessary postage or use electronic mail or FAX for your reply.

If **electronic mail** is available to you, write your comments in an e-mail message and mail it to either of these addresses:

- If you are on the Internet, use this address: techpubs@sgi.com
- For UUCP mail, use this address through any backbone site: *[your_site]*!sgi!techpubs

You can forward your comments (or annotated copies of manual pages) to Technical Publications at this **FAX** number:

415 965-0964