

**CHALLENGE® RAID**  
**Owner's Guide**

Document Number 007-2532-004

## 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, Carl Strasen, Brad Eacker, and Paul Tsien

Cover design and illustration by Rob Aguilar, Rikk Carey, Dean Hodgkinson, Erik Lindholm, and Kay Maitz

© Copyright 1995, Silicon Graphics, Inc.— All Rights Reserved

This document contains proprietary and confidential information of Silicon Graphics, Inc. The contents of this document may not be disclosed to third parties, copied, or duplicated in any form, in whole or in part, without the prior written permission of Silicon Graphics, Inc.

## RESTRICTED RIGHTS LEGEND

Use, duplication, or disclosure of the technical data contained in this document by the Government is subject to restrictions as set forth in subdivision (c) (1) (ii) of the Rights in Technical Data and Computer Software clause at DFARS 52.227-7013 and/or in similar or successor clauses in the FAR, or in the DOD or NASA FAR Supplement. Unpublished rights reserved under the Copyright Laws of the United States. Contractor/manufacturer is Silicon Graphics, Inc., 2011 N. Shoreline Blvd., Mountain View, CA 94043-1389.

Silicon Graphics, CHALLENGE, and IRIS are registered trademarks and IRIX, XFS, IRIS FailSafe, POWER CHALLENGE, and POWER Channel are trademarks of Silicon Graphics, Inc. Oracle Parallel Server and OPS are trademarks of Oracle Corporation.

---

# Contents

**List of Figures** vii

**List of Tables** ix

**About This Guide** xi

- 1. Features of the CHALLENGE RAID Storage System** 1
  - Storage System Components 5
    - SCSI-2 Interface 5
    - CHALLENGE RAID Storage-Control Processor 6
    - Storage System Chassis 8
    - Disk Modules 10
  - Data Availability and Performance 11
    - Data Redundancy 11
    - Enhanced Performance: Disk Striping 12
    - Enhanced Performance: Storage System Caching 12
    - Data Reconstruction and Rebuilding After Disk Module Failure 13
  - RAID Levels 14
    - RAID-0 Group: Nonredundant Array 14
    - RAID-1: Mirrored Pair 15
    - RAID-1\_0 Group: Mirrored RAID-0 Group 16
    - RAID-5: Individual Access Array 18
  - RAID Hot Spare 20
  - Using the CHALLENGE RAID Command Line Interface 21
- 2. Storage System Configurations** 23
  - Basic Configuration 25
  - Dual-Interface/Dual-Processor Configuration 26
  - Split-Bus Configuration 28
  - Dual-Bus/Dual-Initiator Configuration 31

- 3. **Operating the Storage System** 35
  - Checking CHALLENGE RAID Storage System Status 36
    - Using the *raid5* Command 37
    - Getting Device Names With *getagent* 38
    - Getting General System Information 39
    - Getting Information About Disks 39
    - Getting Information About Other Components 42
    - Displaying the CHALLENGE RAID Unsolicited Event Log 43
  - Shutting Down the CHALLENGE RAID Storage System 44
  - Restarting the CHALLENGE RAID Storage System 45
- 4. **Configuring Disks** 49
  - Binding Disks Into RAID Units 49
  - Getting Disk Group (LUN) Information 53
  - Changing LUN Parameters 56
  - Dual Interfaces, Load Balancing, and Device Names 57
- 5. **Maintaining Disk Modules** 59
  - Identifying and Verifying a Failed Disk Module 59
  - Setting Up the Workplace for Replacing or Installing Disk Modules 62
  - Replacing a Disk Module 63
    - Ordering Replacement Disk Modules 63
    - Unbinding the Disk 64
    - Removing a Failed Disk Module 65
    - Installing a Replacement Disk Module 69
    - Updating the Disk Module Firmware 71
  - Installing an Add-On Disk Module Array 72
    - Ordering Add-On Disk Module Arrays 72
    - Installing Add-On Disk Modules 73
    - Creating Device Nodes and Binding the Disks 77

<b>6.</b>	<b>Identifying Failed System Components</b>	<b>79</b>
	Power Supply	80
	Fan Module	80
	Battery Backup Unit	81
	Storage-Control Processor	81
	Diagnosing SP Failure	81
	Using the Auto-Reassign Capability	82
<b>7.</b>	<b>Caching</b>	<b>85</b>
	Setting Cache Parameters	86
	Viewing Cache Statistics	87
	Upgrading CHALLENGE RAID to Support Caching	90
	Changing Unit Caching Parameters	91
<b>A.</b>	<b>Technical Specifications</b>	<b>93</b>
<b>B.</b>	<b>The <i>raid5</i> Command Line Interface</b>	<b>97</b>
	<i>bind</i>	99
	<i>chglun</i>	102
	<i>clearlog</i>	104
	<i>clearstats</i>	104
	<i>firmware</i>	104
	<i>getagent</i>	105
	<i>getcache</i>	106
	<i>getcontrol</i>	109
	<i>getcrus</i>	109
	<i>getdisk</i>	110
	<i>getlog</i>	114
	<i>getlun</i>	116
	<i>setcache</i>	120
	<i>unbind</i>	121
	<b>Index</b>	<b>123</b>



---

## List of Figures

<b>Figure 1-1</b>	CHALLENGE RAID Storage System, Deskside Version: Front View	2
<b>Figure 1-2</b>	CHALLENGE RAID Rack	3
<b>Figure 1-3</b>	Computing SCSI Cable Length Example: Single-Host Configuration Only	4
<b>Figure 1-4</b>	CHALLENGE RAID Server With One SP	6
<b>Figure 1-5</b>	SPs Connected to the Same CHALLENGE Chassis	7
<b>Figure 1-6</b>	SPs Connected to Different CHALLENGE Chassis	7
<b>Figure 1-7</b>	Disk Module Locations (Chassis Front View)	8
<b>Figure 1-8</b>	SCSI-2 Bus and Internal Buses (Front View)	9
<b>Figure 1-9</b>	Disk Modules and Status Lights	10
<b>Figure 1-10</b>	RAID-1 Mirrored Pair (Hardware Mirrored Pair)	15
<b>Figure 1-11</b>	Distribution of User Data in a RAID-1_0 Group	17
<b>Figure 1-12</b>	Distribution of User and Parity Data in a RAID-5 Group	19
<b>Figure 1-13</b>	Hot Spare Example	21
<b>Figure 2-1</b>	Dual-Interface/Dual-Processor Configuration Example	27
<b>Figure 2-2</b>	Split-Bus Configuration Example	30
<b>Figure 2-3</b>	Dual-Bus/Dual-Initiator Configuration Example	33
<b>Figure 3-1</b>	CHALLENGE RAID Indicator Lights	36
<b>Figure 3-2</b>	Disk Module Locations	40
<b>Figure 3-3</b>	Turning Off (On) Power (Back of CHALLENGE RAID Chassis)	45
<b>Figure 3-4</b>	CHALLENGE RAID Indicator Lights	45
<b>Figure 3-5</b>	Unlocking the Fan Module	46
<b>Figure 3-6</b>	Enabling an SP's Power	47
<b>Figure 5-1</b>	Disk Module Status Lights	60
<b>Figure 5-2</b>	Disk Module Locations	61

<b>Figure 5-3</b>	Attaching the ESD Clip to the ESD Bracket on a Deskside Storage System 66
<b>Figure 5-4</b>	Attaching the ESD Clip to the ESD Bracket on a Rack Storage System 66
<b>Figure 5-5</b>	Pulling Out a Disk Module 67
<b>Figure 5-6</b>	Removing a Disk Module 68
<b>Figure 5-7</b>	Engaging the Disk Module Rail 69
<b>Figure 5-8</b>	Engaging the Disk Module Guide 70
<b>Figure 5-9</b>	Inserting the Replacement Disk Module 70
<b>Figure 5-10</b>	Marking the Label for Disk Module A0 74
<b>Figure 5-11</b>	Disk Drive Locations 75
<b>Figure 5-12</b>	Engaging the Disk Module Rail 75
<b>Figure 5-13</b>	Engaging the Disk Module Guide 76
<b>Figure 5-14</b>	Inserting a Disk Module 76
<b>Figure 6-1</b>	Unlocking the Fan Module 83
<b>Figure 6-2</b>	Opening the Fan Module 83
<b>Figure 6-3</b>	Disabling an SP's Power 84
<b>Figure B-1</b>	Disk Module Locations 111

---

## List of Tables

<b>Table 2-1</b>	CHALLENGE RAID Configurations	24
<b>Table 2-2</b>	Error Recovery: Basic Configuration	25
<b>Table 2-3</b>	Error Recovery: Dual Interface/Dual-Processor Configuration	26
<b>Table 2-4</b>	Error Recovery: Split-Bus Configuration	28
<b>Table 2-5</b>	Error Recovery: Dual-Bus/Dual-Initiator Configuration	31
<b>Table 3-1</b>	Output of <i>raid5 getagent</i>	38
<b>Table 3-2</b>	Output of <i>raid5 getdisk</i>	41
<b>Table 3-3</b>	Output of <i>raid5 getcrus</i>	43
<b>Table 4-1</b>	Output of <i>raid5 getlun</i>	54
<b>Table 5-1</b>	Ordering Replacement Disk Modules	63
<b>Table 5-2</b>	Ordering Add-On Disk Module Sets	72
<b>Table 6-1</b>	Field-Replaceable Units	79
<b>Table 7-1</b>	Output of <i>raid5 getcache</i>	88
<b>Table A-1</b>	CHALLENGE RAID Deskside Chassis Specifications	93
<b>Table A-2</b>	CHALLENGE RAID Rack Specifications	94
<b>Table B-1</b>	<i>raid5</i> Parameters	97
<b>Table B-2</b>	Output of <i>raid5 getagent</i>	105
<b>Table B-3</b>	Output of <i>raid5 getcache</i>	107
<b>Table B-4</b>	Output of <i>raid5 getcrus</i>	110
<b>Table B-5</b>	Output of <i>raid5 getdisk</i>	112
<b>Table B-6</b>	getlog Error Codes	114
<b>Table B-7</b>	Output of <i>raid5 getlun</i>	118



---

## About This Guide

The CHALLENGE<sup>®</sup> RAID storage system provides a compact, high-capacity, high-availability source of disk storage for the complete line of Silicon Graphics<sup>®</sup> CHALLENGE servers running IRIX<sup>™</sup> 5.3 and 5.3 with XFS<sup>™</sup>: CHALLENGE S, CHALLENGE DM, CHALLENGE L, and CHALLENGE XL. It also works with IRIX 6.1 and 6.2 on the POWER CHALLENGE L and XL.

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), and 5 are supported, as well as disks configured as hot spares. In addition, a basic CHALLENGE RAID storage system provides storage-system caching.

### 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.
- Appendix A, “Technical Specifications,” summarizes technical information for the CHALLENGE RAID deskside storage system.
- Appendix B, “The raid5 Command Line Interface,” lists and explains all parameters of the *raid5* command.

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

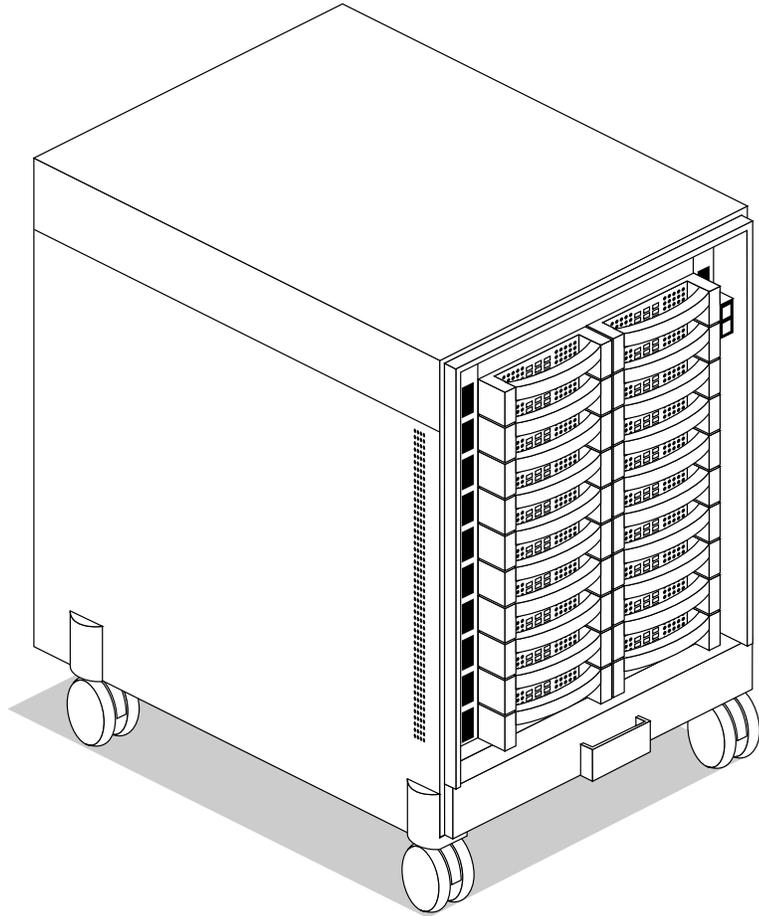
## Features of the CHALLENGE RAID Storage System

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)” (University of California, Berkeley, Report No. UCB/CSD/87/391). That paper defines various levels of RAID.

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

- CHALLENGE RAID storage system components
- data availability and performance
- RAID levels
- the RAID hot spare
- using the CHALLENGE RAID command line interface

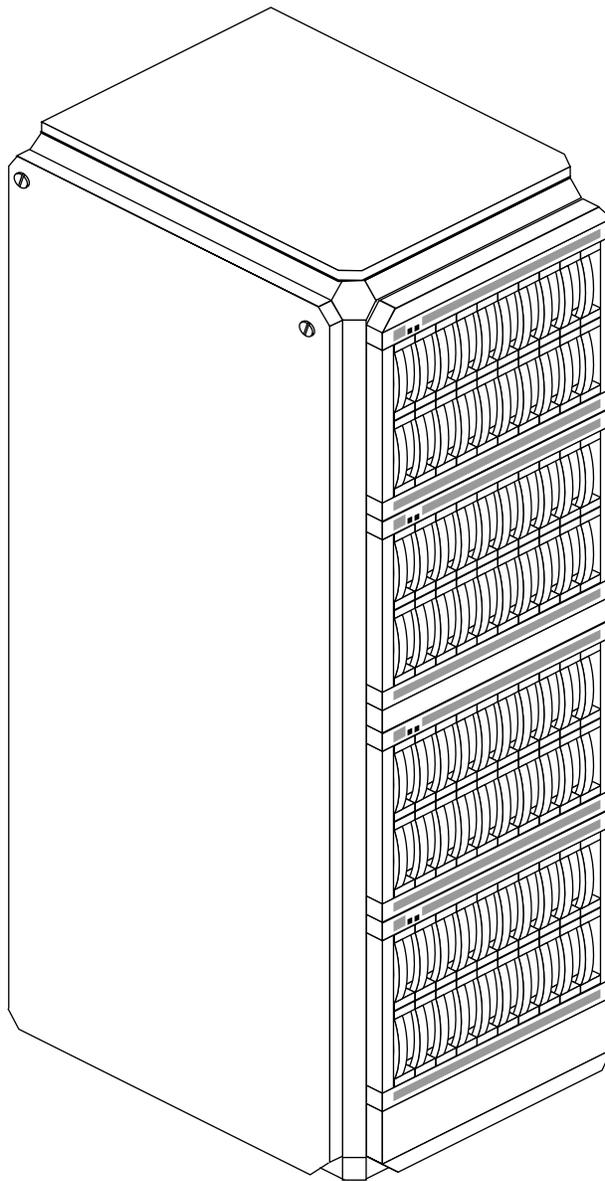
Figure 1-1 is an external view of the desktide version of the CHALLENGE RAID storage system.



**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

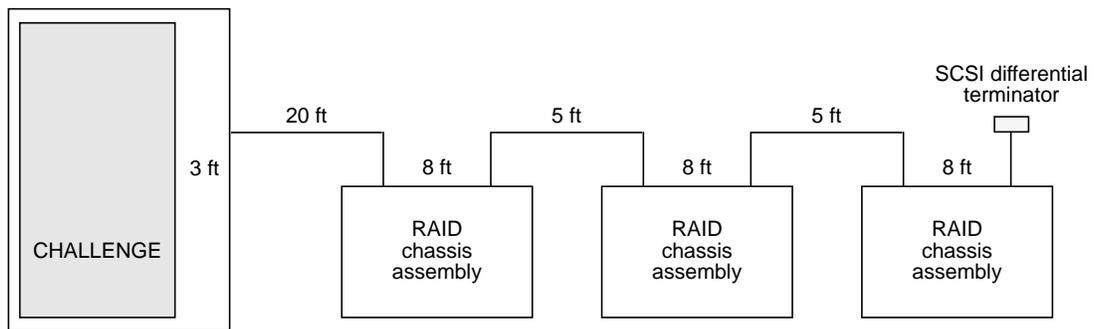
The CHALLENGE RAID storage system provides a compact, high-capacity, high-availability source of disk storage for your CHALLENGE S, DM, L, or XL server system. The storage system offers a large capacity of high-availability disk storage in multiple disk modules that you can replace when the storage system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

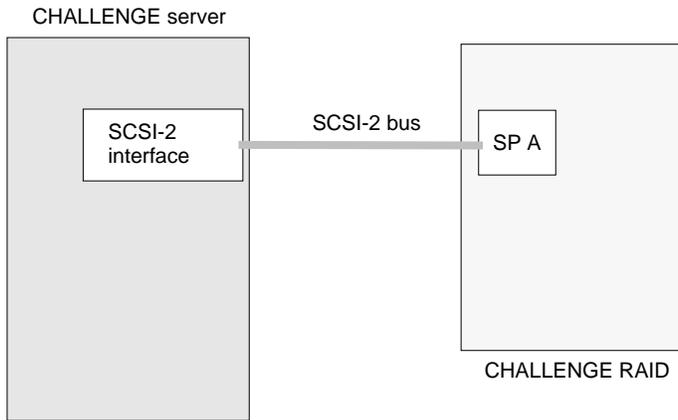
### SCSI-2 Interface

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.

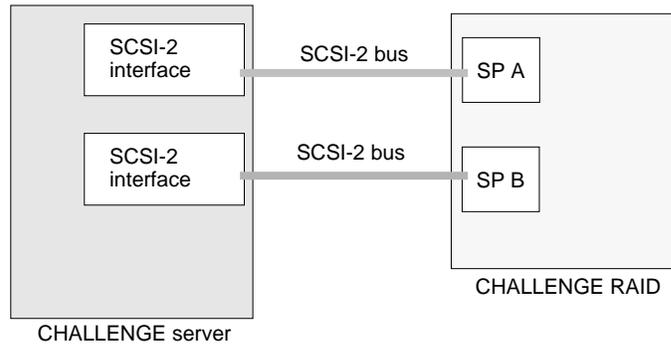
Figure 1-4 diagrams a CHALLENGE RAID storage system with one SP.



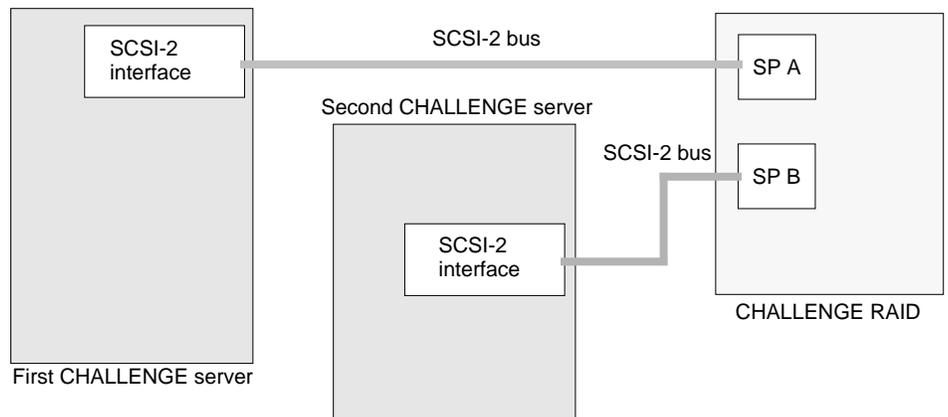
**Figure 1-4** CHALLENGE RAID Server With One SP

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 disks 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

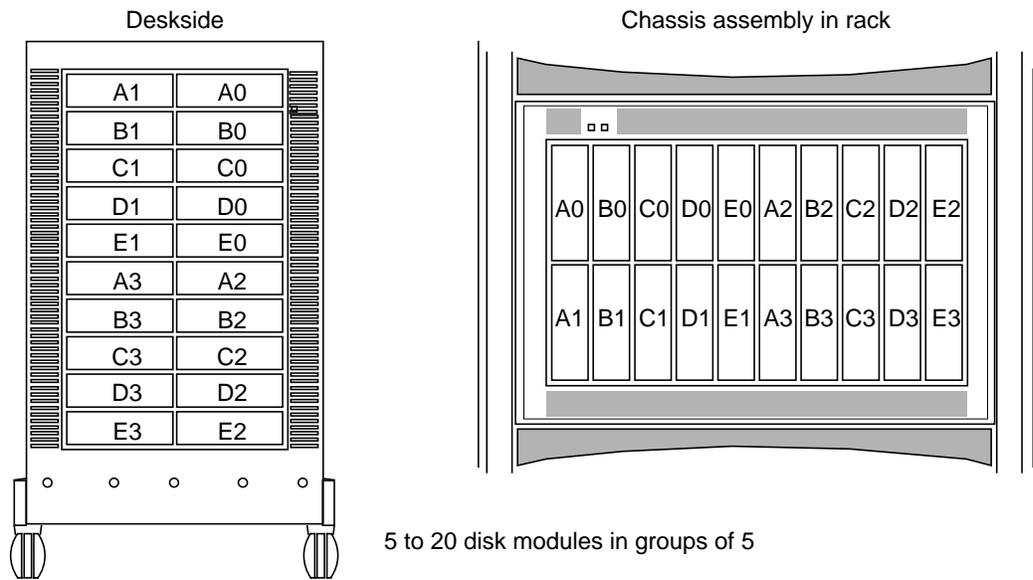
### Storage System Chassis

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:

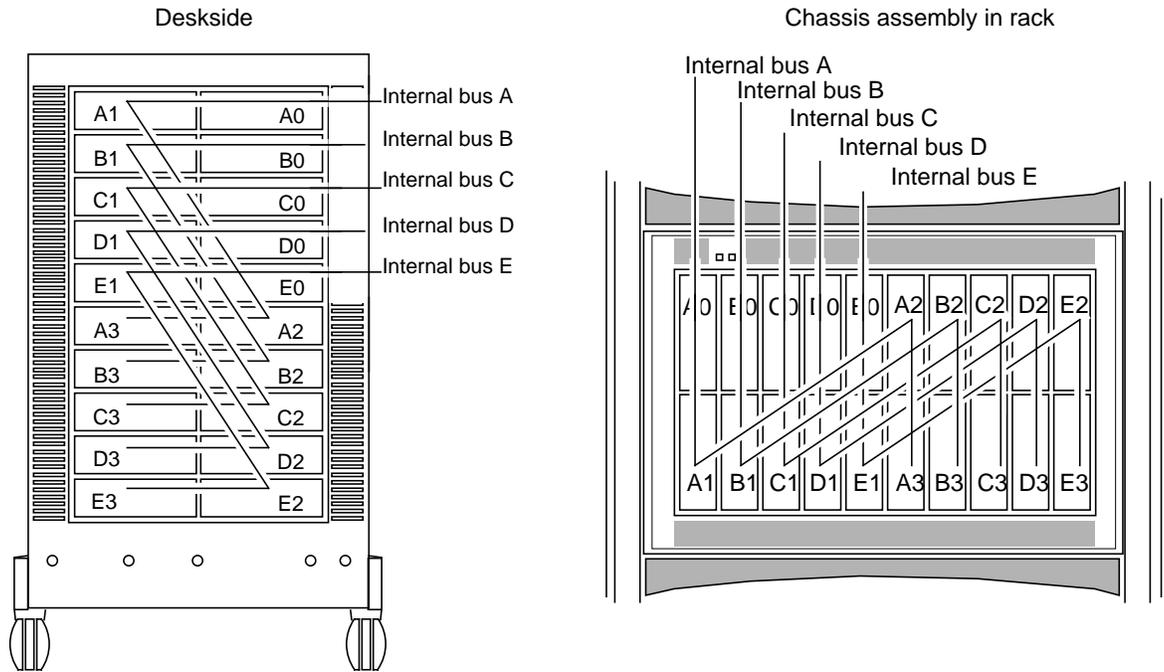
- modules A0, B0, C0, D0, and E0 (array 0)
- modules A1, B1, C1, D1, and E1 (array 1)
- modules A2, B2, C2, D2, and E2 (array 2)
- modules A3, B3, C3, D3, and E3 (array 3)

Figure 1-7 diagrams this placement. Individual disk modules have disk position labels attached.



**Figure 1-7** Disk Module Locations (Chassis Front View)

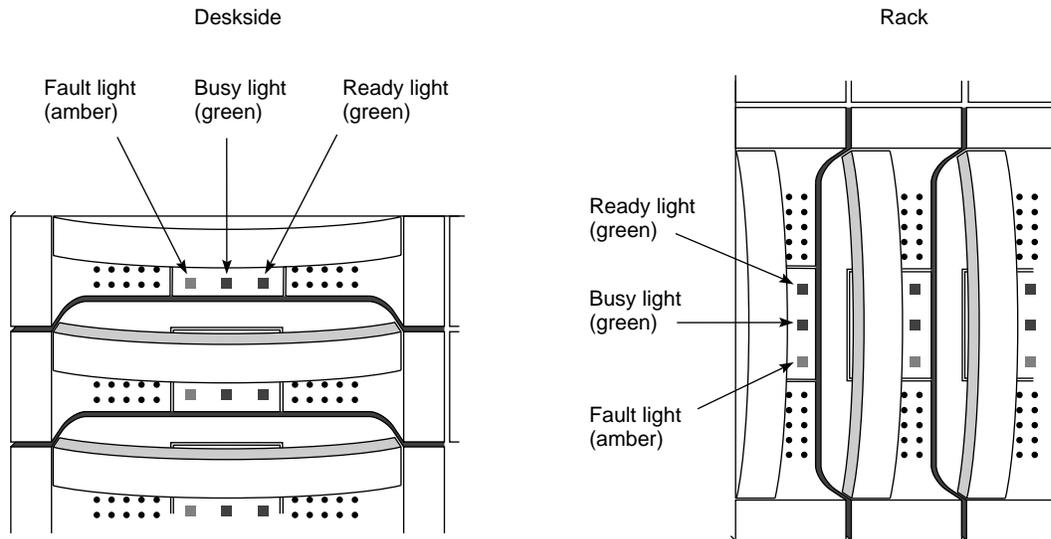
Through the SP, the SCSI-2 bus is split into five internal fast/narrow SCSI buses—A, B, C, D, and E—that connect the slots for the disk modules. For example, internal bus A connects the modules in slots A0, A1, A2, and A3, in that order. Figure 1-8 diagrams this configuration.



**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, and RAID-5.

Because the CHALLENGE RAID storage system has five internal SCSI-2 buses, RAID-5 provides redundancy for up to five groups of disk modules. A 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 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.

You enable storage system caching and specify basic cache parameters, and enable or disable read and write caches for individual disk units using the command line interface, as explained in Chapter 7.

**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 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-5 group: individual access array

**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.

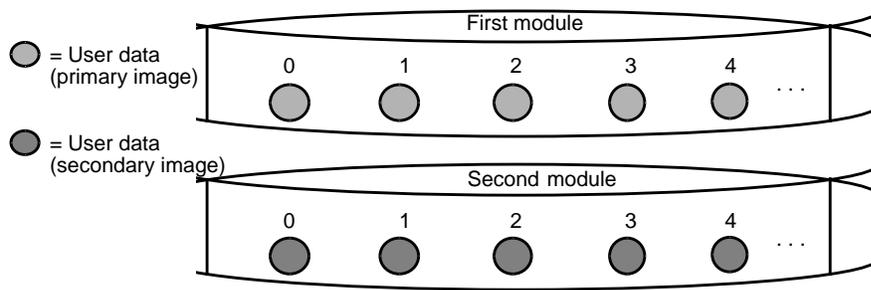
## RAID-1: Mirrored Pair

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 its 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)

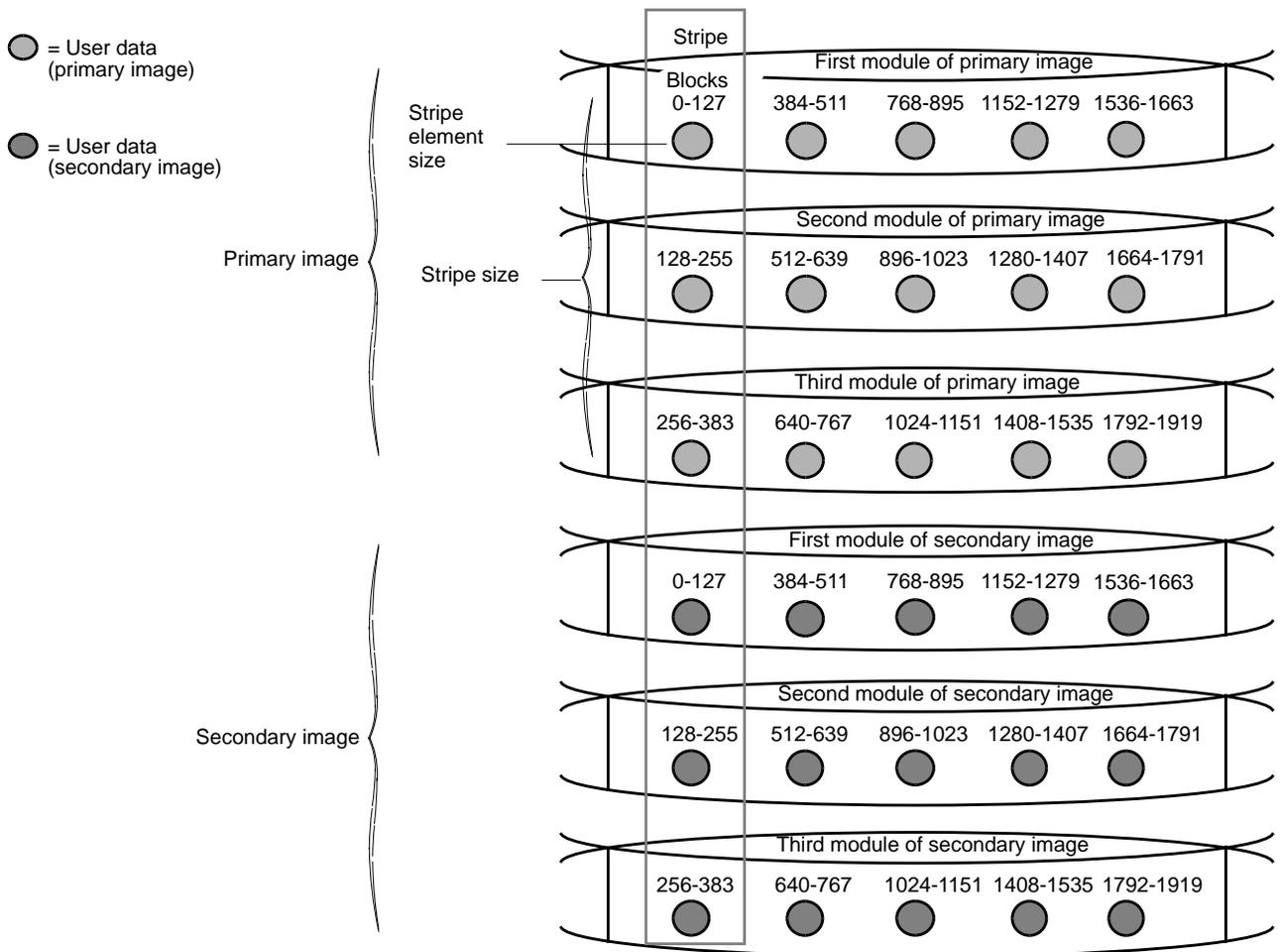
To achieve the maximum fault tolerance, configure the mirror with each disk module on a different internal SCSI bus; for example, the primary image on A0, the secondary image on B0, and so on.

### **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 modules, to the third mirrored image 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 a separate internal SCSI buses.

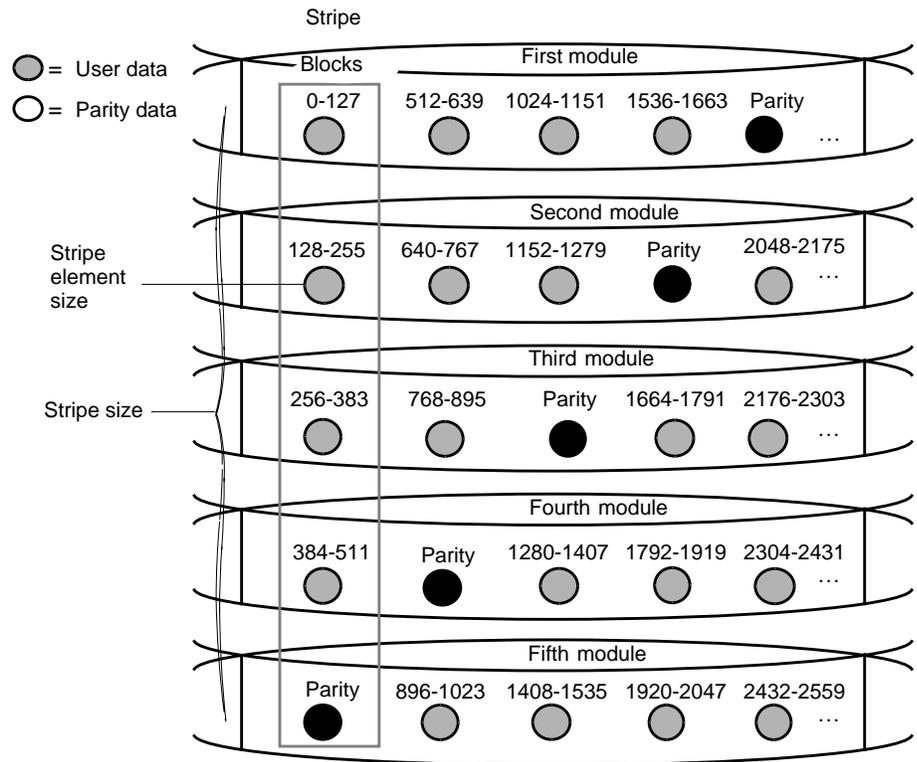
### **RAID-5: Individual Access Array**

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 1-12 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, and fourth, fifth, then back to the first, and so on.



**Figure 1-12** 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.

## 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-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.

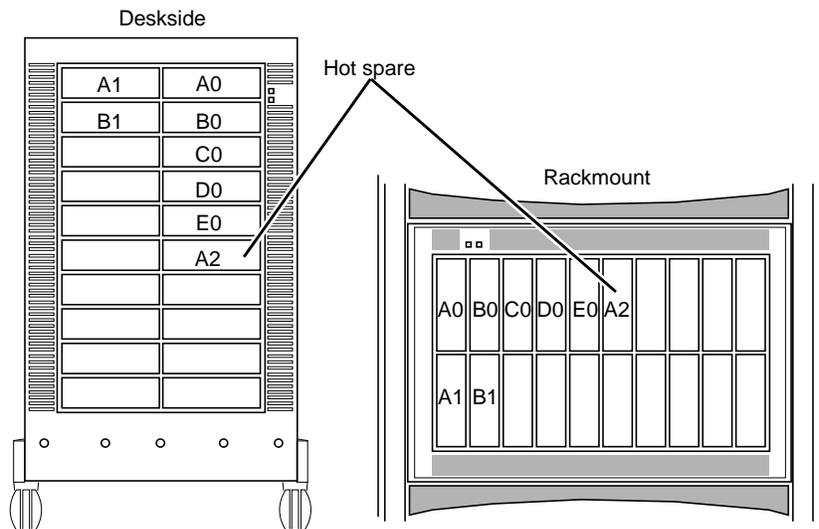
**Note:** The SP finishes rebuilding the disk module before it begins copying data, even if you replace the failed disk during the rebuild process.

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 may 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 1-13. 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 1-13** Hot Spare Example

## Using the CHALLENGE RAID Command Line Interface

Run the command line interface, `/usr/raid5/raid5`, in an IRIX window on your CHALLENGE server to

- bind (group) or unbind physical disks into a RAID-0, RAID-1, RAID-1\_0, or RAID-5 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)
- 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

**Note:** Although the directory and command are *raid5*, the command is valid for all RAID levels.

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.

---

## 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*: 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 fail to storage system disk units. Authorized service provider replaces interface, and system operator restarts operating system and applications.
SCSI-2 cable	No	I/O operations fail to storage system. 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.

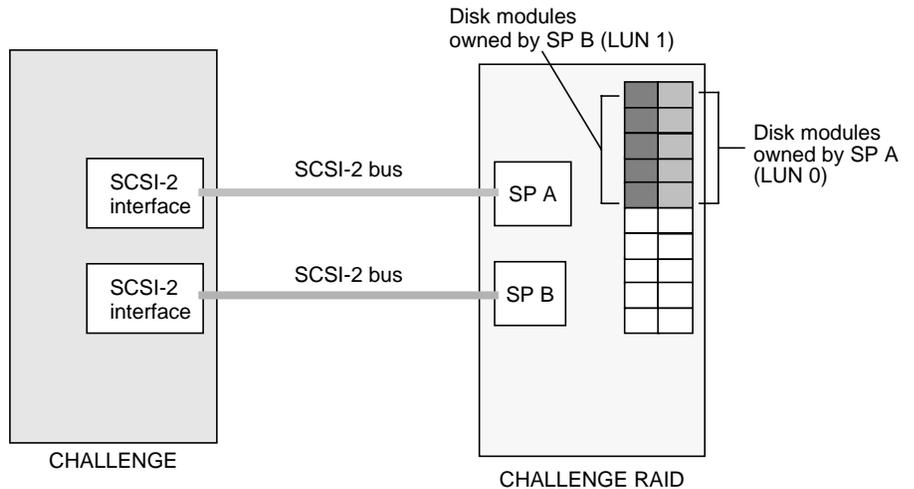
**Table 2-3** Error Recovery: Dual Interface/Dual-Processor Configuration

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 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** 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.

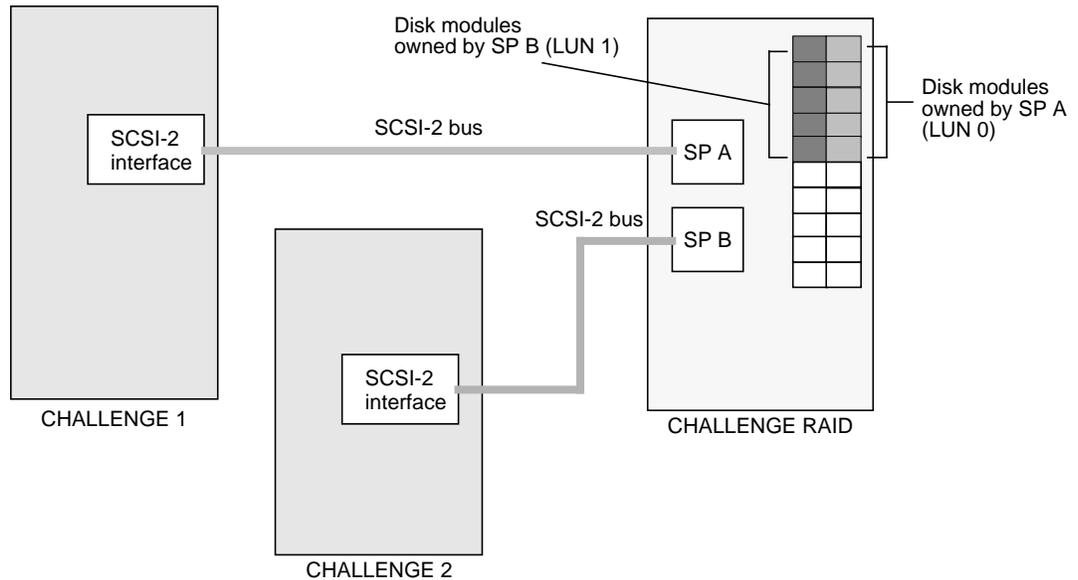
**Table 2-4** Error Recovery: Split-Bus Configuration

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 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.

**Table 2-4** (continued) Error Recovery: Split-Bus 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	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.

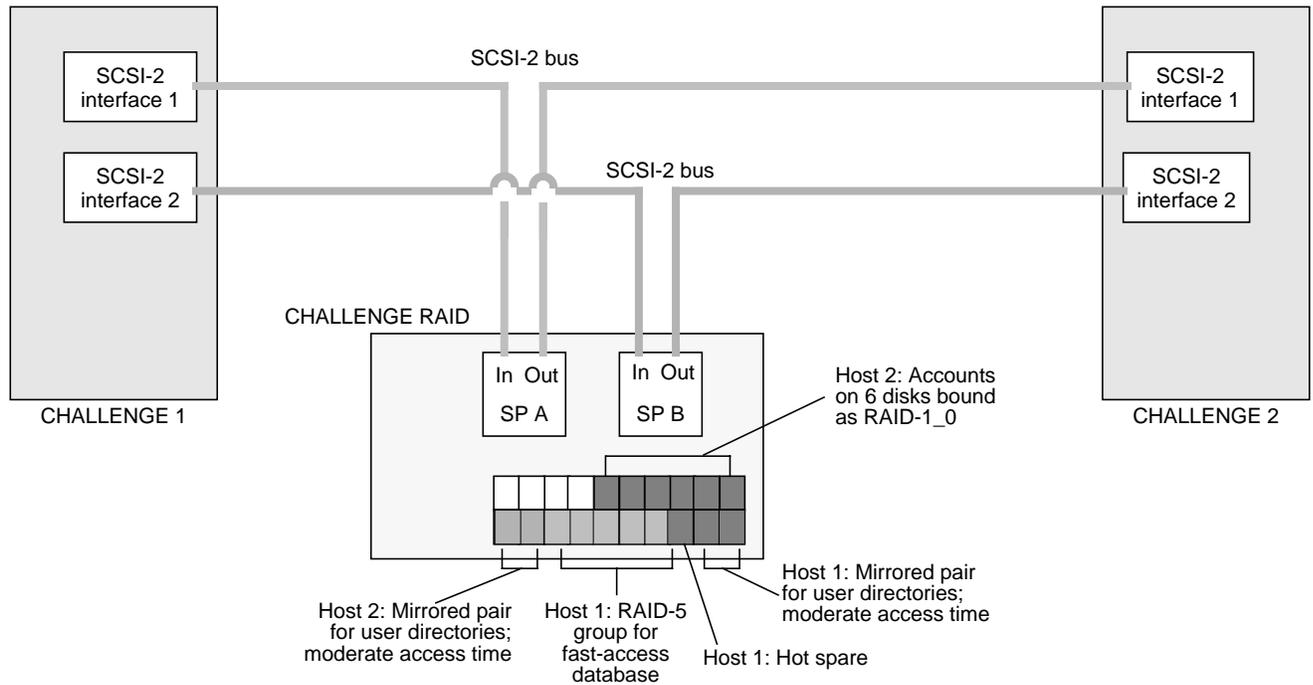
**Table 2-5** Error Recovery: Dual-Bus/Dual-Initiator Configuration

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 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** 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 run CHALLENGE RAID 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/raid5` command (command line interface, or CLI). Use `raid5` 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.

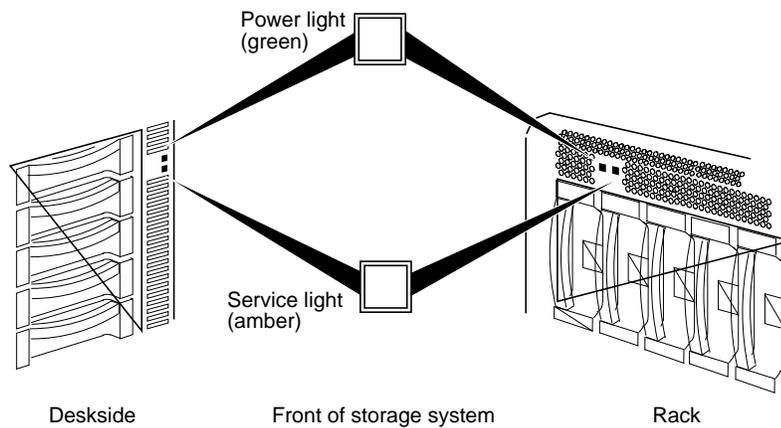
**Note:** Although the directory and command are `raid5`, the command is valid for all RAID levels.

Other chapters in this guide explain how to use the `raid5` 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 whether 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 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 *raid5* command in an IRIX shell, as explained in this section:

- using the *raid5* 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 *raid5* Command

The *raid5* command sends storage management and configuration requests to an application programming interface (API) on the CHALLENGE server. For the *raid5* 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 *raid5* command is

```
raid5 [-vp] [-d device] parameter [optional_arguments]
```

In this syntax, variables mean:

- |                  |  |
|------------------|--|
| -v               | Enables verbose return.  |
| -p               | Parses the <i>raid5</i> command without calling the API. If the string does not parse correctly, an error message is printed to <i>stderr</i> ; otherwise there is no output.  |
| -d <i>device</i> | Target RAID device. Use <i>raid5 getagent</i> for a list of RAID devices. This switch must be present for all <i>raid5</i> 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 *raid5* parameters.

### Getting Device Names With *getagent*

Use the *getagent* parameter with *raid5* to display information on devices controlled by the API:

```
raid5 getagent
```

Following is a sample output for one device; normally, the output would give information on all devices.

```
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 *raid5 getagent* output.

**Table 3-1** Output of *raid5 getagent*

Entry	Meaning
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 <i>/dev/scsi</i> 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 <i>getagent</i> ).
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.
Prom Rev	PROM revision present on the SP.

**Table 3-1** (continued) Output of *raid5 getagent*

Entry	Meaning
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

```
raid5 -d device getcontrol
```

A possible 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: 5/5/1995
Day of the week: Friday
System Time: 12:43:54
```

## Getting Information About Disks

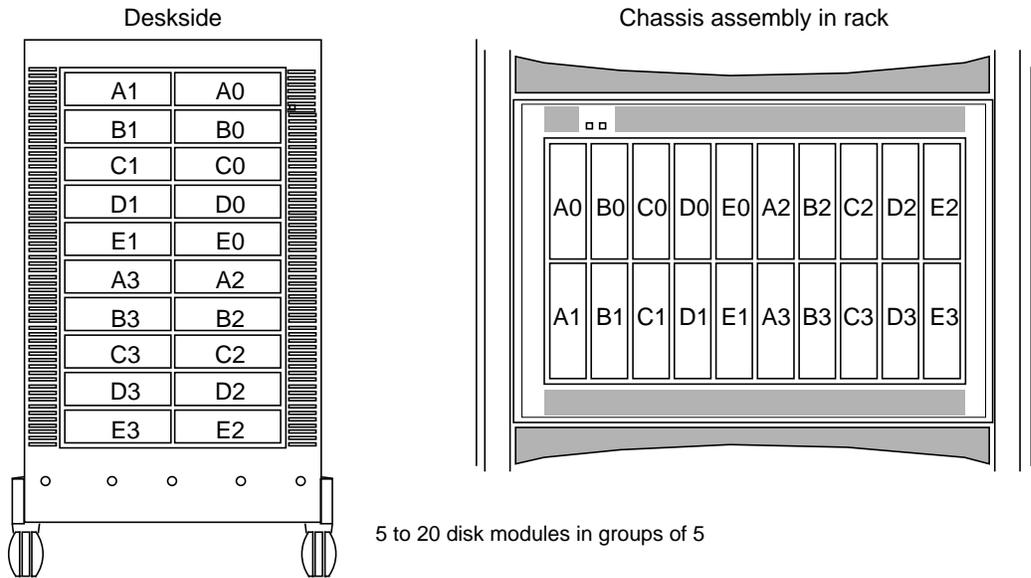
For information about all bound disks in the system, use this command in an IRIX shell:

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

For information on a particular disk, use

```
/usr/raid5/raid5 -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 lower case) 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 A2.

```
raid5 -d scsi4d210 getdisk a2
```

A sample output of this command follows:

```
A0 Vendor Id: SEAGATE
A0 Product Id: ST15150N
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 *raid5 getdisk*

Output	Meaning
Vendor Id	Manufacturer of disk drive
Product Id	2.1 GB disk: ST32550N 4.2 GB disk: ST15150N
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 Enabling: 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 *raid5 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

```
raid5 -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 *raid5 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

```
raid5 -d device getlog
```

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

```
raid5 -d device getlog +n
```

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

```
raid5 -d device getlog -n
```

Output of the command `raid5 -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

```
raid5 -d device clearlog
```

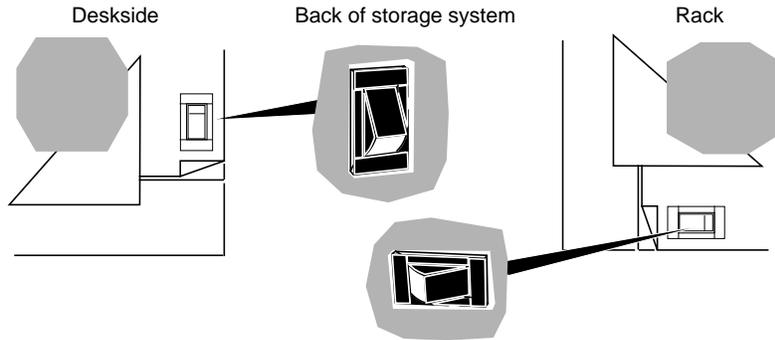
**Note:** You must be root to use the `clearlog` parameter.

## Shutting Down the CHALLENGE RAID Storage System

Follow these steps to shut down the CHALLENGE RAID storage system:

1. If you are using storage system caching, make sure that it is disabled; use `raid5 getcache` to check status. If necessary, disable it with `raid5 setcache disable`, as explained in Chapter 7.
2. Turn off the power switch on the back of the CHALLENGE RAID storage system, as shown in Figure 3-3.

**Note:** You do not need to disable the power for the SP(s).



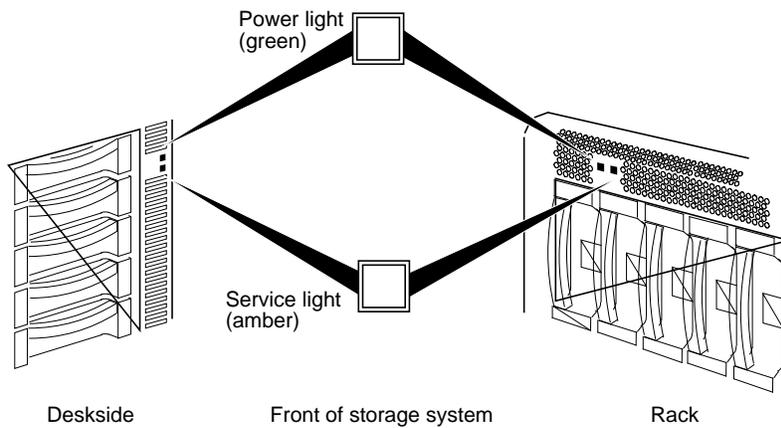
**Figure 3-3** Turning Off (On) Power (Back of CHALLENGE RAID Chassis)

## Restarting the CHALLENGE RAID Storage System

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

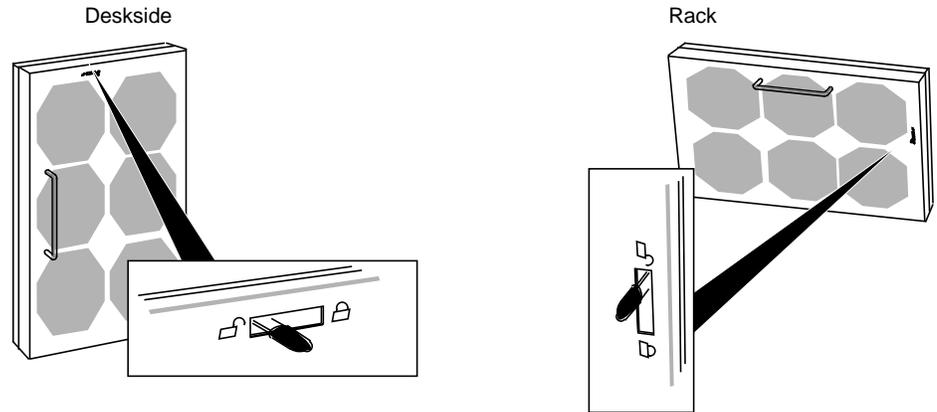
1. Turn on the storage system's power; see Figure 3-3.

The green power light on the front of the storage system turns on (see Figure 3-4) and the fans rotate.



**Figure 3-4** CHALLENGE RAID Indicator Lights

2. If the busy light on *none* of the drive modules lights 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 3-5.

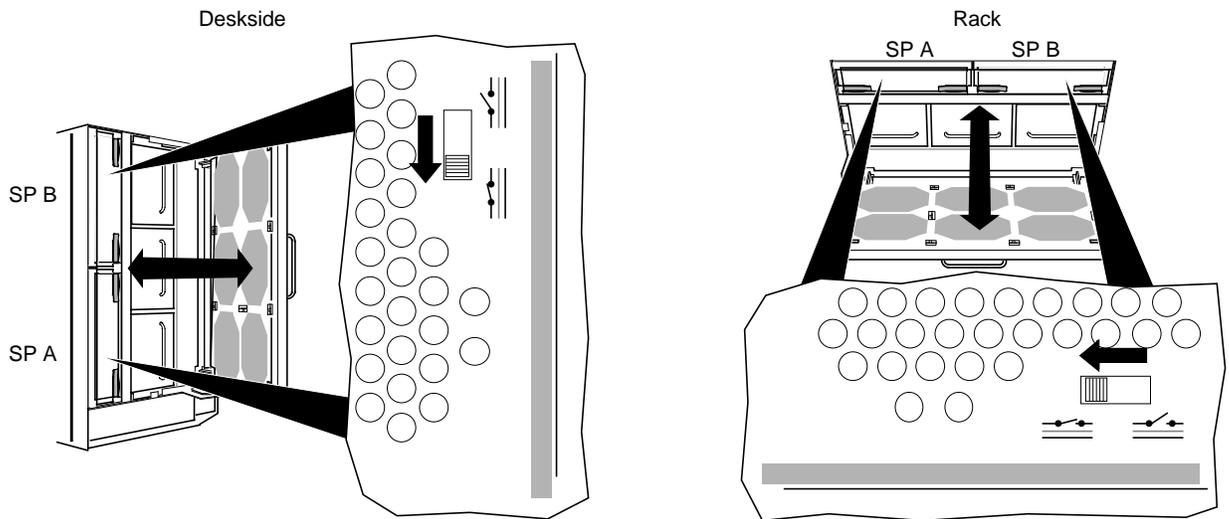


**Figure 3-5** Unlocking the Fan Module

3. Swing open the fan module.

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

4. Move the SP's power switch to the enable position, as shown in Figure 3-6.



**Figure 3-6** Enabling an SP's Power

5. Close the fan module by closing the fan module and moving the module's latch to the LOCK position.
6. Power on the CHALLENGE server(s).



---

## Configuring Disks

This chapter explains

- binding disks into RAID units
- getting disk group (LUN) information
- changing LUN parameters

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

### 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 F (15 decimal).

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.

To group physical disks into RAID-0, RAID-1, RAID-1\_0, or RAID-5 units or to create a hot spare, use as root:

```
raid5 -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 `raid5 getagent`; see “Getting Device Names With `getagent`,” in Chapter 3.

<i>raid-type</i>	<p>Choices are</p> <ul style="list-style-type: none"><li>• r0: RAID-0</li><li>• r1: RAID-1</li><li>• r1_0: RAID-1_0</li><li>• r5: RAID-5</li><li>• hs: hot spare</li></ul>
<i>lun-number</i>	<p>Logical unit number to assign the unit (a hexadecimal number between 0 and F).</p>
<i>disk-names</i>	<p>Indicates which physical disks to bind, in the format <i>bd</i>, where <i>b</i> is the physical bus name (a through e; be sure to use lower case) and <i>d</i> 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.</p> <p>A RAID-0 bind requires a minimum of three (maximum 16) disks.</p> <p>A RAID-1 bind requires two disks on separate buses.</p> <p>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: p1, s1, p2, s2, p3, s3, and so on.</p> <p>A RAID-5 bind also requires separate buses, each with five disk modules. Legal RAID-5 bind configurations are</p> <ul style="list-style-type: none"><li>• a0 b0 c0 d0 e0</li><li>• a1 b1 c1 d1 e1</li><li>• a2 b2 c2 d2 e2</li><li>• a3 b3 c3 d3 e3</li></ul> <p>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.</p>

The optional arguments are as follows:

**-r *rebuild-time*** Maximum time in hours to rebuild a replacement disk. Default is 4 hours; legal values are any number greater than or equal to 0.

A rebuild time of 2 hours rebuilds the disk more quickly but degrades response time slightly. A rebuild time of 0 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.

**-s *stripe-size*** Number of blocks per physical disk in a RAID stripe. Default is 128; legal values are any number greater than 0.

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.

**Note:** Although *bind* returns immediate status for a RAID device, the bind itself does not complete for 45 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 “Bound But Not Assigned” in the *getlun* output.

The following example binds disks A0, B0, C0, D0, and E0 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.

```
raid5 -d sc4d2l0 bind r5 3 a0 b0 c0 d0 e0 -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.

```
raid5 -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.

```
raid5 -d sc4d2l0 bind r1_0 1 a1 b1 c1 d1 -r 4 -s 128 -c read
```

The following example binds A3, B3, C3, D3, and E3 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.

```
raid5 -d sc4d2l0 bind r0 3 a3 b3 c3 d3 e3 -s 128 -c read
```

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

```
raid5 -d sc4d2l0 bind hs 7 e3
```

There is no output for *raid5* with the *bind* parameter. Errors are printed to *stderr*.

**Note:** For complete messages, it is recommended that you use the *-v* option.

## Getting Disk Group (LUN) Information

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

```
raid5 getlun lun-number
```

The following example displays information about LUN 3.

```
raid5 -d sc4d2l0 getlun 3
```

Following is truncated output for a RAID-5 group of five disks.

**Note:** Information on individual disks is not displayed unless statistics logging is enabled with *raid5 getcontrol*. See “Getting Information About Other Components” in Chapter 3 of this guide.

```
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
[etc.]

C0 Enabled
[etc.]

D0 Enabled
[etc.]

E0 Enabled
[etc.]

```

Table 4-1 summarizes entries in the *raid5 getlun* output.

**Table 4-1** Output of *raid5 getlun*

Entry	Meaning
Type	RAID0, RAID1, RAID10, RAID5, or 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 <i>chglun</i> ; see "Dual Interfaces, Load Balancing, and Device Names," later in this chapter
Auto-trespass	Always Disabled
Auto-assign	Always Enabled
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 <i>chglun</i>
Idle Delay Time	Amount of time in 100-ms intervals that unit is below idle threshold; set with <i>chglun</i>

Table 4-1 (continued) Output of *raid5 getlun*

Entry	Meaning
Write Aside Size	Smallest write-request size in blocks that can bypass the cache and go directly to the disk; set with <i>chglun</i>
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 <i>getdisk</i> )
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

**Table 4-1** (continued) Output of *raid5 getlun*

Entry	Meaning
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

## Changing LUN Parameters

To change parameters for a logical unit, use

```
raid5 -d device chglun -l lun [ -c cache-flags] [-d default-owner]
[-r rebuild-time] [-i idle-thresh] [-t idle-delay-time] [-w write-aside]
```

**Note:** Only root can use the *chglun* 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 storage-control processor ownership of LUN
- 0: don't change ownership

If your storage system has dual SPs, see “Dual Interfaces, Load Balancing, and Device Names,” later in this chapter.

*-r rebuild-time* Maximum time in hours to rebuild a replacement disk. Default is 4 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 any number greater than or equal to 0.
- 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 any number greater than or equal to 0.
- w *write-aside* The smallest write request size, in blocks, that will bypass the cache and go directly to the disks. Legal values are any number greater than or equal to 0.

The following example changes LUN 3 to perform write caching and rebuild in four hours; it does not change the default owner.

```
raid5 -d sc4d2l0 chglun -l 3 -c write -d 0 -r 4
```

There is no output for the *chglun* parameter. Errors are printed to *stderr*.

## 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.



## 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

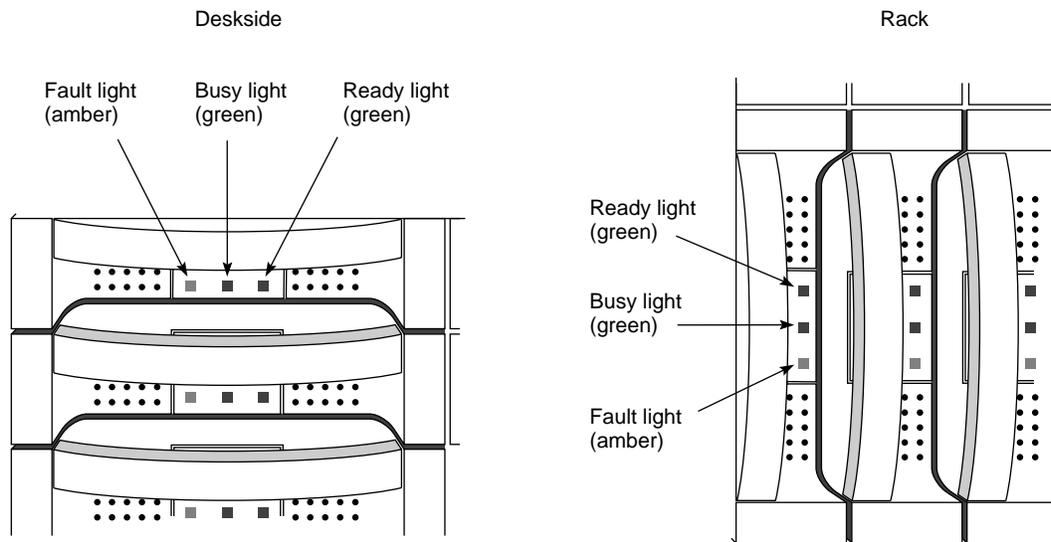
If you have determined that a module has failed by examining the cabinet fault light or by using the *raid5 getdisk* or *raid5 getcrus* command, as explained in Chapter 3 in this guide, you can replace the defective module and rebuild your data without powering off the CHALLENGE RAID storage system or 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:

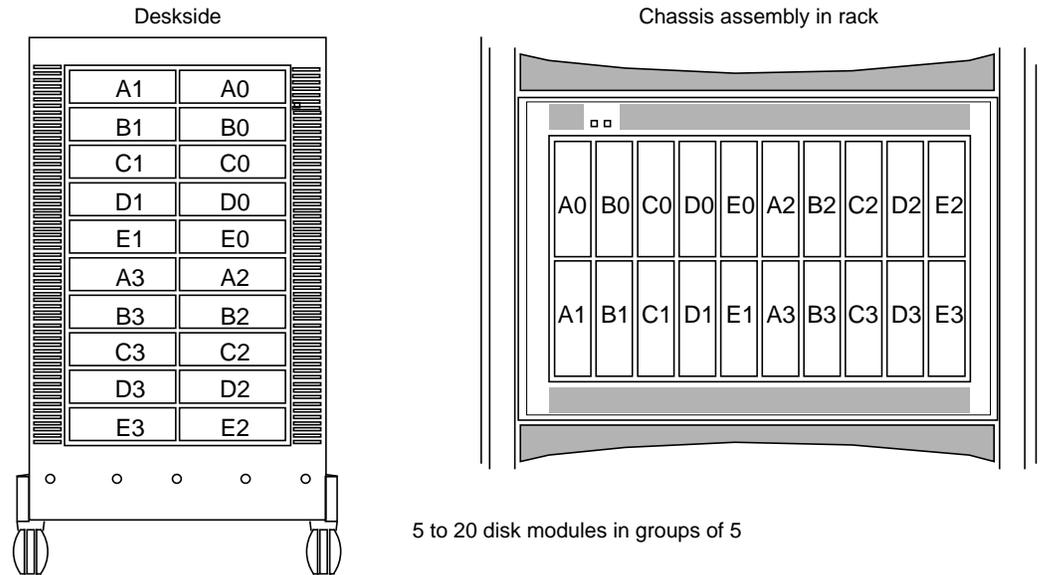
1. Look for the module with its amber fault light on. Figure 5-1 shows the fault indicator light and other lights on a disk module.



**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. If you replace a disk module, you must update the firmware as explained in "Updating the Disk Module Firmware," later in this chapter.



**Figure 5-2** Disk Module Locations

3. If you have not already checked the module status with *raid5 getdisk*, do so now; see “Getting Information About Disks” 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 *raid5 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

To avoid discharging electrostatic charge that has accumulated on your body discharges through the circuits, 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 or 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.

## Replacing a Disk Module

This section explains

- ordering replacement disk modules
- unbinding the disk
- removing the failed disk module
- installing the replacement disk module
- updating the disk module firmware

### 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-RAID5-1X2
Replacement 4.2 GB drive	P-S-RAID5-1X4

**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.

To unbind a disk, use the *unbind* parameter with the *raid5* command. Follow these steps:

1. In an IRIX window, use *raid5 getagent* to get the device name (node number):

```
raid5 getagent
```

2. If necessary, use *raid5 getdisk* to verify which the position of the failed disk.

3. Use *raid5 unbind* to unbind the failed disk:

```
raid5 -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:

```
raid5 -d sc4d210 unbind 3 -o
```

The disks in LUN 3 are now free to be reconfigured. Use *raid5 bind* to configure disks, as explained in Chapter 4.

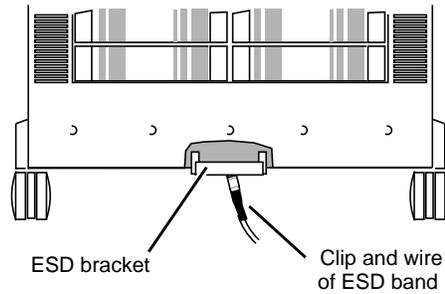
## Removing a Failed Disk Module

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.

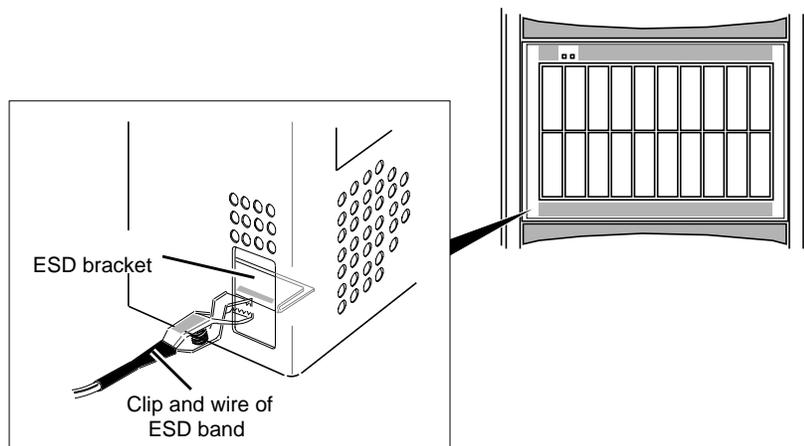
To remove a disk module, follow these steps:

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 desktside storage system.



**Figure 5-3** Attaching the ESD Clip to the ESD Bracket on a Deskside Storage System

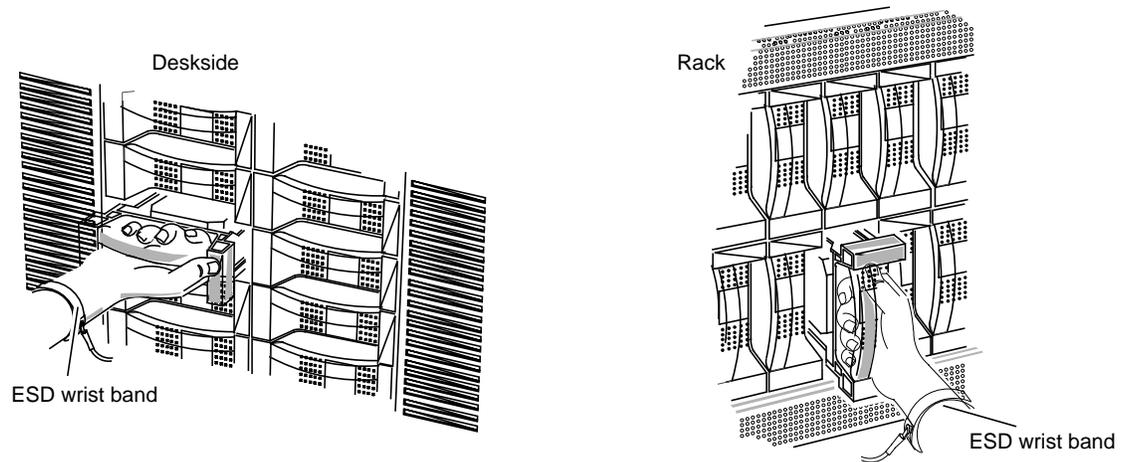
Figure 5-4 shows where to attach the clip on a rack storage system.



**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.



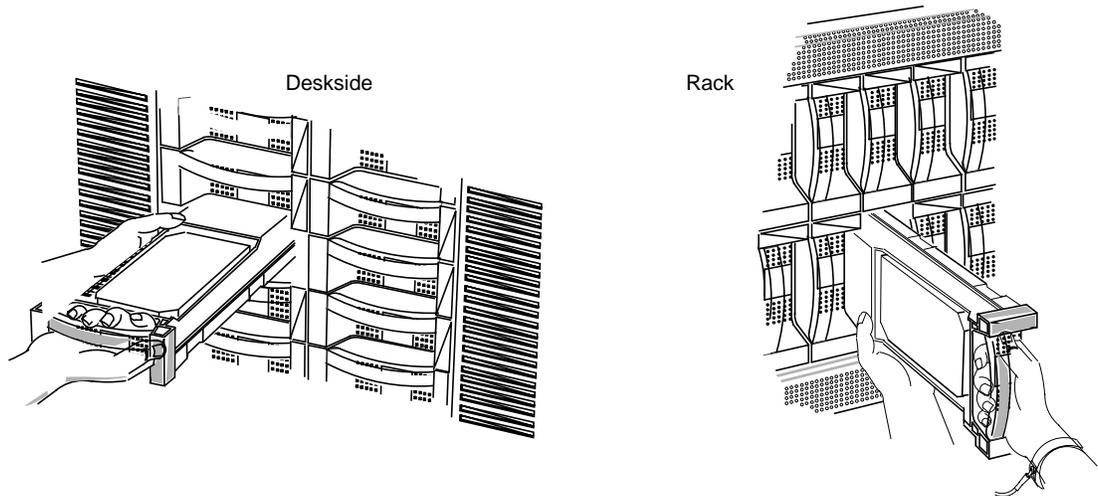
**Figure 5-5** Pulling Out a Disk Module

**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.

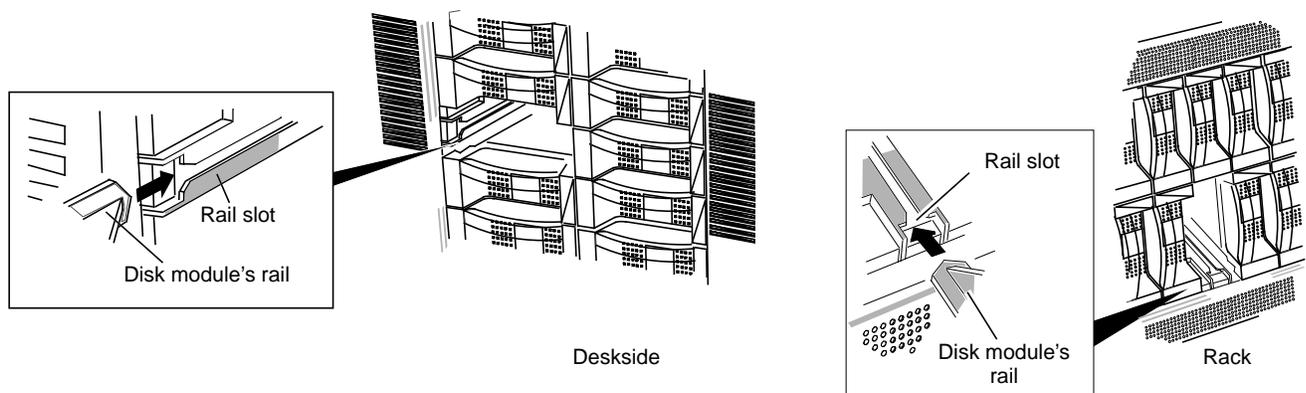
## Installing a Replacement Disk Module

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.

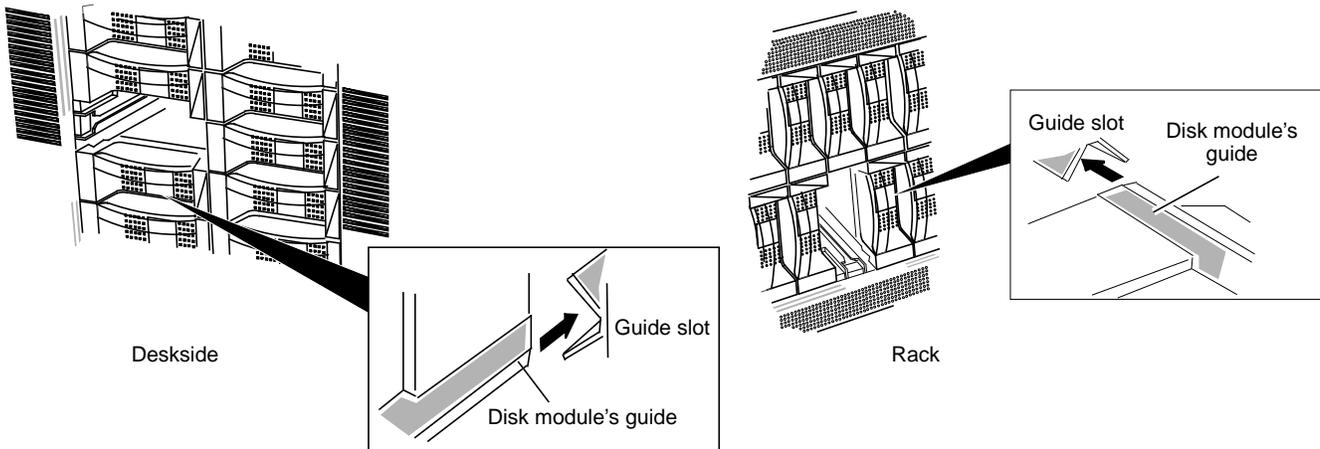
**Caution:** The disk module is extremely sensitive to shock and vibration. Even a slight jar can severely damage it.

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.



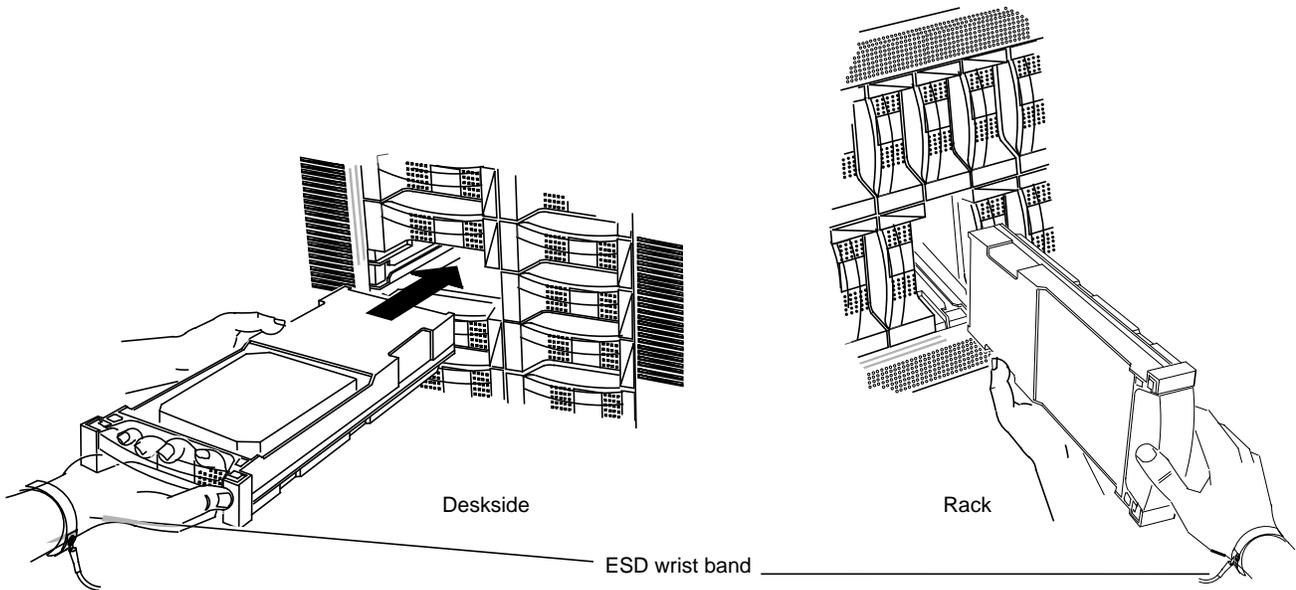
**Figure 5-7** Engaging the Disk Module Rail

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.

The default rebuild period is 4 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.

## Updating the Disk Module Firmware

After replacing a failed unbound disk module (A0, B0, C0, or A3), update the firmware on the CHALLENGE RAID SP. Follow these steps:

1. Quiesce the bus, disabling all applications. Make sure that only the RAID agent is running.
2. Type as root:

```
raid5 -d device firmware /usr/raid5/flarecode.bin
```

**Caution:** You must use this command every time you replace a failed unbound disk module (A0, B0, C0, or A3).

The image in the file given in the command contains microcode that runs on the SP and also a microcode image destined for the SP PROM, which runs the power-on diagnostics.

## Installing an Add-On Disk Module Array

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 module arrays 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 module arrays. Normal processing can continue while you install disk modules in arrays of five.

This section explains

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

### Ordering Add-On Disk Module Arrays

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

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-RAID5-5X2
Add-on five 4.2 BG drives	P-S-RAID5-5X4
Base array with five 2 GB drives	P-S-RAID5-B5X2
Base array with five 4 GB drives	P-S-RAID5-B5X4
Replacement 2.1 GB drive	P-S-RAID5-1X2
Replacement 4.2 GB drive	P-S-RAID5-1X4

**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 Add-On 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 add-on disk modules; see Figure 5-2.

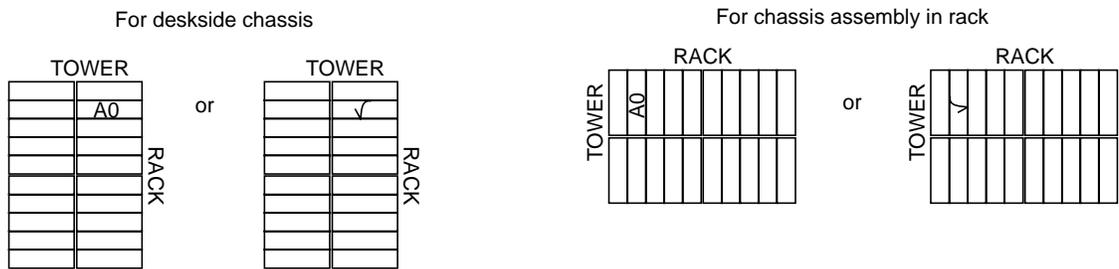


**Warning:** Although you need not complete each chassis assembly that is partially filled before installing more disk modules in the next chassis assembly, avoid making the rack top-heavy.

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
  - next, 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.
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 5-10 shows these two ways of labeling disk module A0.



**Figure 5-10** Marking the Label for Disk Module A0

For reference, Figure 5-11 diagrams all disk module locations.

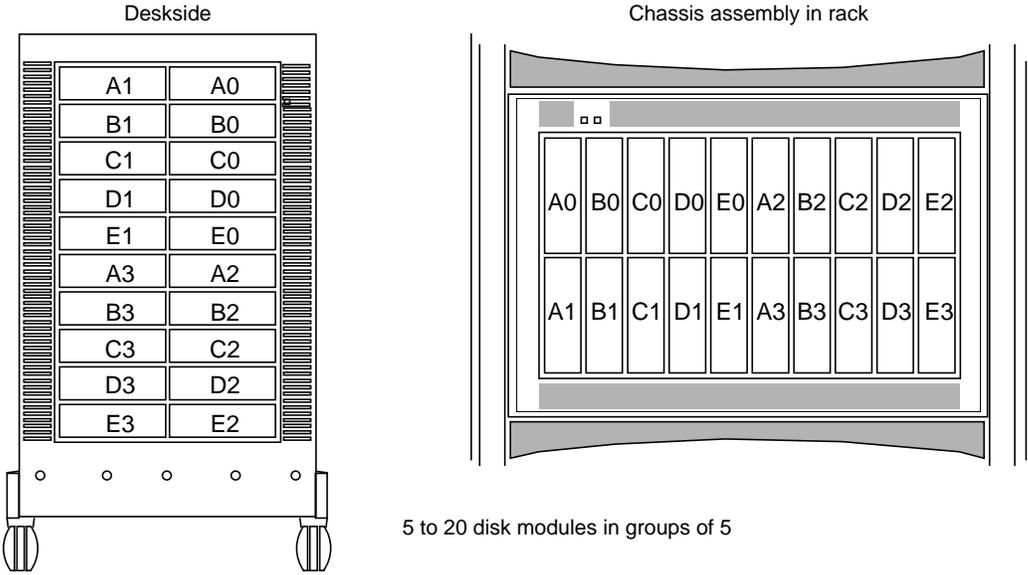


Figure 5-11 Disk Drive Locations

- 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.

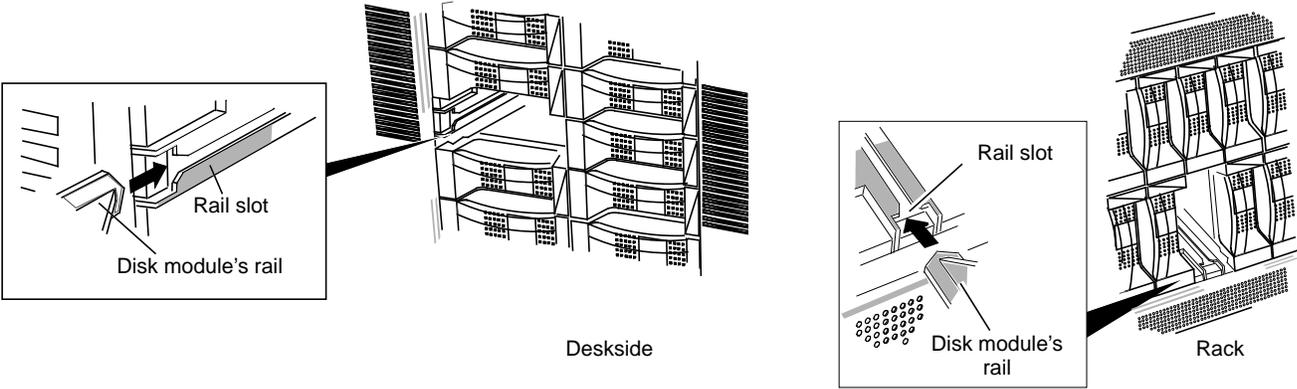
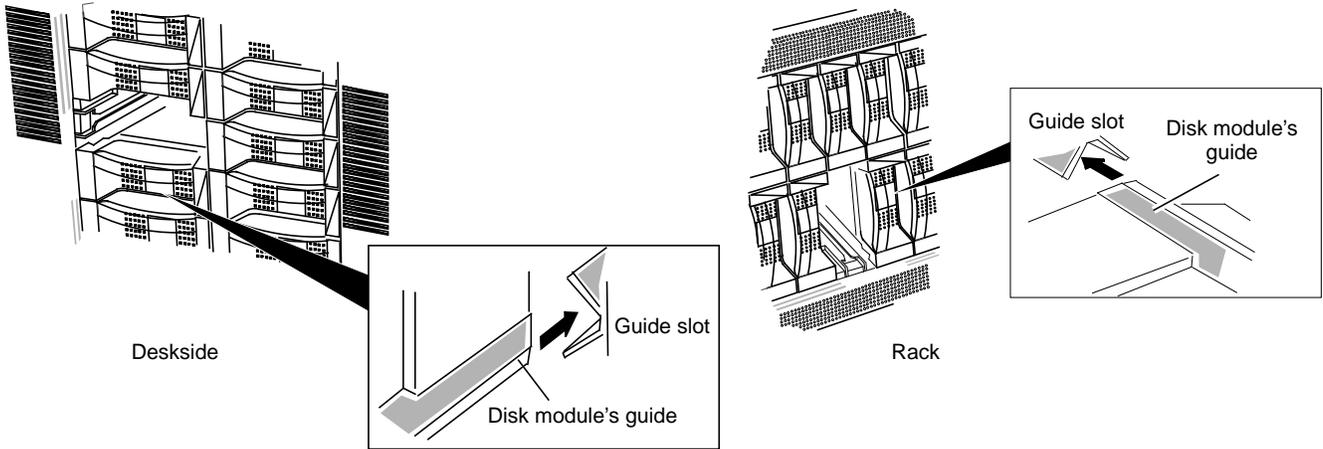


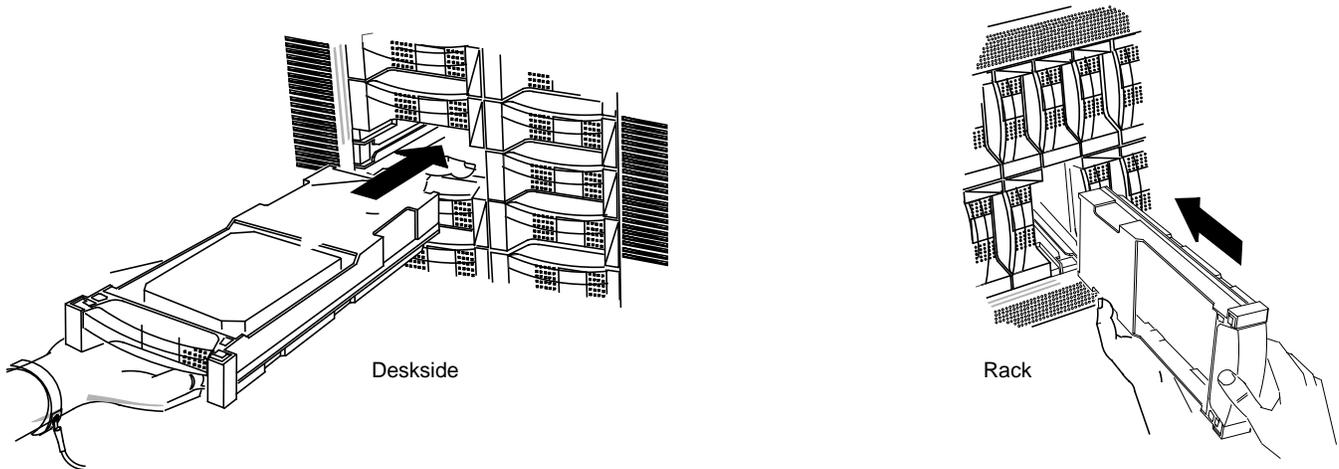
Figure 5-12 Engaging the Disk Module Rail

- 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

- 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 arrays 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 without rebooting. Also, in a system with two storage-control SPs which are used for primary and secondary paths, both SPs must be made aware of the new disks. Also, 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 to identify a failed 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, or in the output of the *raid5 getcrus* command, 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)

Table 6-1 lists Silicon Graphics marketing codes for replacement units for the CHALLENGE RAID storage system.

**Table 6-1** Field-Replaceable Units

Unit	Silicon Graphics Marketing Code
Power supply module	P-S-RAID5-PWR
Storage-control processor (SP)	P-S-RAID5-SP
Battery backup unit	P-S-RAID5-BBU
Fan module with 6 fans	P-S-RAID5-FAN
Add-on 8 MB SIMM cache kit	P-S-RAID5-C8
Add-on 64 MB SIMM cache kit	P-S-RAID5-C64

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

## 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-drive modules. If a second power supply fails, the entire chassis shuts down immediately.

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

## Fan Module

Each CHALLENGE RAID storage system has one fan module, containing six fans. 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, or if you see a fan fault in the *raid5 getcrus* command output, have the entire fan unit replaced as soon as possible by a Silicon Graphics System Service Engineer.

If the temperature within a CHALLENGE RAID chassis reaches an unsafe level, the power system shuts down the chassis 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 *raid5 getcache* command is "Disabled." Caching is not re-enabled until the battery backup unit is replaced with a fully charged one.

The *raid5 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 or if the battery backup unit state is shown as "Faulted" in the *raid5 getcrus* command output, have the battery backup unit replaced as soon as possible by a Silicon Graphics System Service Engineer.

## Storage-Control Processor

This section discusses

- diagnosing SP failure
- using the auto-reassign capability

### Diagnosing SP Failure

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.

For SP failure, the unsolicited event log contains the message “Peer Controller: Removed.” You can also use *raid5 getcrus* to verify the failed SP, as explained in Chapter 3. The *getcrus* output indicates the failed SP, for example, “SPA State: Not Present.”

### Using the Auto-Reassign Capability

If one of two SPs becomes physically unavailable to the system, that is, if an SP fails or becomes disconnected, the CHALLENGE RAID storage system can automatically reassign the LUNs it controls to the remaining SP.

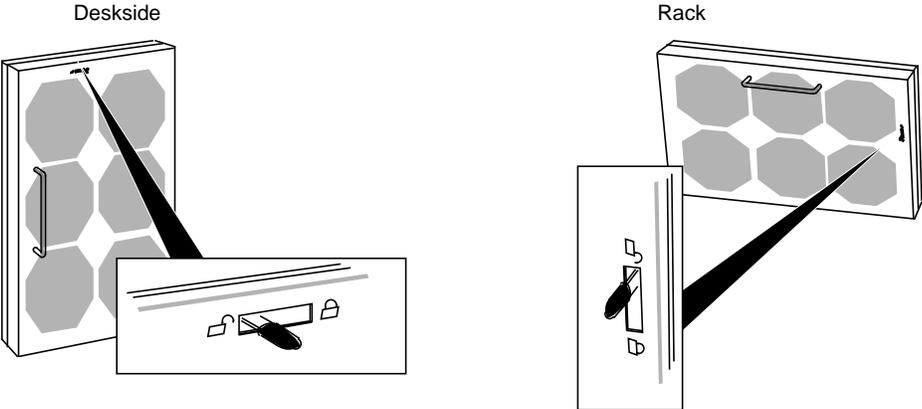
Note the following points:

- Auto-reassign capability cannot be used in the case of problems caused by a defective CHALLENGE-to-CHALLENGE RAID connection, such as a damaged cable.
- To enable auto-reassign, power off the failed SP. You do not need to power down the CHALLENGE RAID storage system.  
**Caution:** Do not power off the SP under any circumstances other than to enable auto-reassign.
- Powering off an SP requires opening the fan module at the back of the storage system. Because the fan module must not be left in the open position for more than two minutes, make sure you can complete this process within that time limit.

Follow these steps:

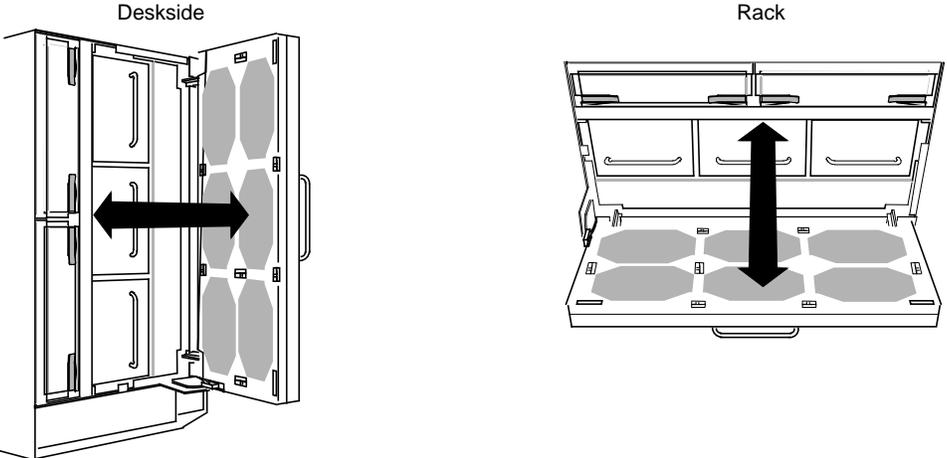
**Caution:** To prevent thermal shutdown of the storage system, never operate the storage system for more than two minutes with the fan module in the open position. Do not use this procedure under any circumstances other than to enable auto-reassign.

1. On the back of the CHALLENGE RAID storage system, move the fan module's latch to the UNLOCK position, as shown in Figure 6-1.



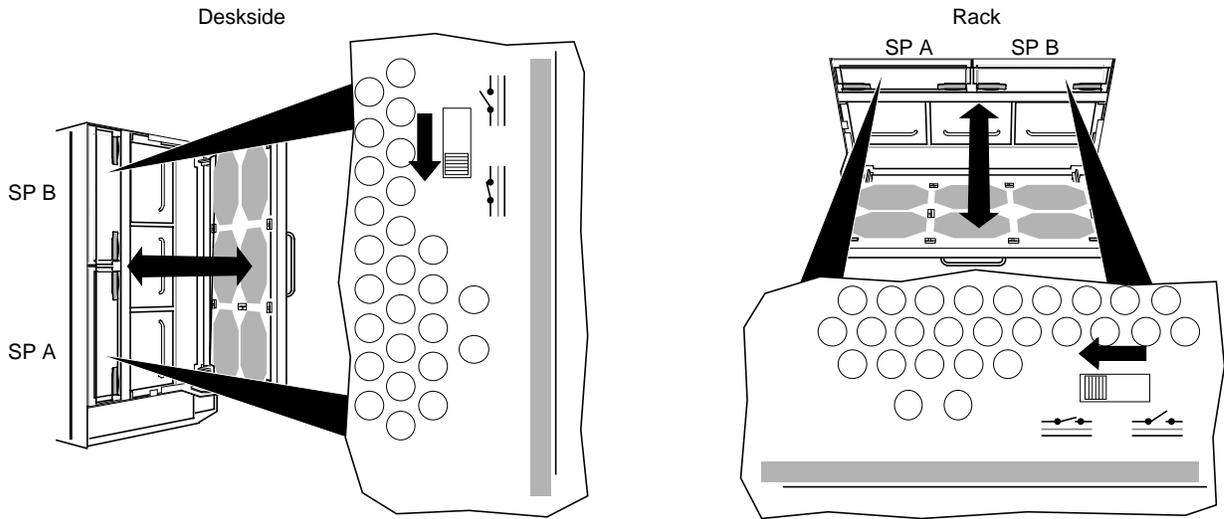
**Figure 6-1** Unlocking the Fan Module

2. Swing open the fan module, as shown in Figure 6-2.



**Figure 6-2** Opening the Fan Module

3. Disable the failed SP's power by sliding the switch to the Disable position, as shown in Figure 6-3.



**Figure 6-3** Disabling an SP's Power

**Caution:** Do not power off the SP under any circumstances other than to enable auto-reassign.

4. Immediately close and lock the fan module to let the SP cool.
5. Have the SP replaced as soon as possible by qualified service personnel. Leave the SP powered off in the meantime.

---

# Caching

Storage-system caching requires

- two storage-control processors, each with the same amount of cache memory
- fully charged storage system backup battery
- minimum of 8 MB of memory for each SP
- cache enabling using the *raid5 setcache* command, as explained in this chapter
- disk modules in slots A0, B0, C0, D0, and E0 as a fast repository for cached data

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

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 *raid5 setcache* command:

```
raid5 -d device setcache enable | disable [-u usable] [-p page] [-l low] [-h high]
```

In this syntax, variables mean:

*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 *raid5 getagent* (see “Getting Device Names With *getagent*,” in Chapter 3). Valid values are 0, 8, and 64:

- 0: user has selected to disable the cache
- other values: 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.

**Note:** For 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.

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:

```
raid5 -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:

```
raid5 -d device setcache 1 -u 64 [parameters]
```

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

## Viewing Cache Statistics

If you use storage-system caching, you can use the `raid5 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, caching is disabled until the failed module is replaced. If an SP fails, caching is disabled until the failed SP is replaced.

To display cache information, use

```
raid5 -d device getcache
```

Following is a 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
    
```

Table 7-1 summarizes entries in the *raid5 getcache* output.

**Table 7-1** Output of *raid5 getcache*

Entry	Meaning
Usable Cache	<p>Size of cache in megabytes to use for caching, not greater than the SP memory size as displayed by <i>raid5 getagent</i> (see “getagent,” in Appendix B). 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).</p> <p>The CLI does not let you specify more memory than you have. If you specify less than you have, the remaining memory is unused.</p>
Page Size	User-specified page size in KB for caching; for example, 2 means a 2 KB page size.
Cache State	<p>Enabled: SP is fully functional.</p> <p>Disabled: SP not capable of or configured for caching.</p> <p>Synching: SP is synchronizing its cache with the peer SP.</p> <p>Enabling: cache is in the process of becoming enabled.</p> <p>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.</p> <p>Frozen: SP has finished quiescing and is now waiting for peer SP to finish quiescing or dumping.</p> <p>Dumping: SP is in the process of dumping contents of the cache to a private location on the disk.</p> <p>Initing: the SP is initializing.</p>

Table 7-1 (continued)	Output of <i>raid5 getcache</i>
Entry	Meaning
Low 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 default is 50, regardless of whether caching is enabled or disabled.
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	Number of cache pages currently owned by SP A
SP B Cache Pages	Number of cache pages currently owned by SP B
Unassigned Cache Pages	Number of dirty cache pages not owned by A or B. This can happen 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
Prct Dirty Cache Pages	Percentage of cache pages that this SP owns which are dirty
Prct Cache Pages Owned	Percentage of cache pages owned by this controller when pages from other SP are taken into account
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)

## 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. See “Updating the Disk Module Firmware” in Chapter 5.
- 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, follow these steps to set up caching:

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

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

or

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

or

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

2. Specify the cache page size and cache size parameters.
3. Enable storage system caching with `raid5 -d device getcache`, as explained in “Setting Cache Parameters,” earlier in this chapter.

## Changing Unit Caching Parameters

This section explains how to change cache parameters.

**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.

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

1. Run *raid5 getcache* to determine if caching is enabled. If it is enabled, make sure both SPs are powered on; run *raid5 getcrus* to find out.
2. If caching is enabled, disable it with  

```
raid5 -d device setcache 0
```
3. Wait for the cache memory to be written to disk, which may take several minutes. Use *raid5 getcache* to check cache state (“Disabled”).
4. Follow instructions in “Setting Cache Parameters,” earlier in this chapter.



---

## 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
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	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**

<b>Classification</b>	<b>Specification</b>	<b>Value</b>
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	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 *raid5* Command Line Interface

This appendix is a complete guide to the `/usr/raid5` command (command-line interface, or CLI). It explains the *raid5* parameters, which are summarized in Table B-1.

**Table B-1** *raid5* Parameters

Parameter	Use
<i>bind</i>	Bind (group) physical disks
<i>chglun</i>	Change parameters for a currently bound group
<i>clearlog</i>	Destroy the contents of the RAID storage controller's error log
<i>clearstats</i>	Reset statistics logging on the RAID storage-control processor (SP), setting all log counters to 0
<i>firmware</i>	Update the firmware on the CHALLENGE RAID SP
<i>getagent</i>	Get names and descriptions of devices controlled by the SP
<i>getcache</i>	Get information about the storage system caching environment
<i>getcontrol</i>	Get general system information
<i>getcrus</i>	Display status information on all system components, such as the fan module, battery backup unit, and so on
<i>getdisk</i>	Display status information on one disk module or all disk modules in the system
<i>getlog</i>	Display the SP log, in which errors are stored, or display the head and tail of the log

**Table B-1 (continued)**      *raid5* Parameters

Parameter	Use
<i>getlun</i>	Display information about a disk group
<i>setcache</i>	Change information about the RAID controller's caching environment
<i>unbind</i>	Deconfigure physical disks from their current logical configuration, destroying all data on the logical unit (group)

The *raid5* command sends CHALLENGE RAID storage management and configuration requests to an application programming interface (API) on the CHALLENGE server. For the *raid5* 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 *raid5*, the command is valid for all RAID levels.

The synopsis of the *raid5* command is

```
raid5 [-vp] [-d device] parameter [optional_arguments]
```

In this syntax, options and variables mean:

- v                      Enables verbose return.
- p                      Parses the *raid5* 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 *raid5 getagent* for a list of RAID devices. This switch must be present for all *raid5* 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 sc4d210
setenv RaidAgentDevice /dev/scsi/sc4d210
```

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 F (15 decimal).

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 45 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.

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

```
raid5 -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 *raid5 getagent*.

`raid-type` Choices are

- r0: RAID-0
- r1: RAID-1

- r1\_0: RAID-1\_0
- r5: RAID-5
- hs: hot spare

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

*disk-names* Indicates which physical disks to bind, in the format *bd*, where *b* is the physical bus name (a through e) 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: p1, s1, p2, s2, p3, s3, 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.

The optional arguments are as follows:

*-r rebuild-time* Maximum time in hours to rebuild a replacement disk. Default is 4 hours; legal values are any number greater than or equal to 0.

A rebuild time of 2 hours rebuilds the disk more quickly but degrades response time slightly. A rebuild time of 0 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.

**-s *stripe-size*** Number of blocks per physical disk in a RAID-5 stripe. Default is 128; legal values are any number greater than 0.

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 A0, B0, C0, D0, and E0 into a RAID-5 logical unit with a logical unit number of 3, a four-hour maximum rebuild time, and a 128-block strip size per physical disk, with read cache enabled.

```
raid5 -d sc4d2l0 bind r5 3 a0 b0 c0 d0 e0 -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.

```
raid5 -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.

```
raid5 -d sc4d2l0 bind r1_0 1 a1 b1 c1 d1 -r 4 -s 128 -c read
```

The following example binds A3, B3, C3, D3, and E3 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.

```
raid5 -d sc4d2l0 bind r0 3 a3 b3 c3 d3 e3 -s 128 -c read
```

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

```
raid5 -d sc4d2l0 bind hs 7 e3
```

There is no output for *raid5* with the *bind* parameter. Errors are printed to *stderr*.

**Note:** For complete messages, it is recommended that you use the *-v* option.

## ***chglun***

To change parameters for a logical unit, use

```
raid5 -d device chglun [-l lun] [-c cache-flags] [-d default-owner]  
[-r rebuild-time] [-i idle-thresh] [-t idle-delay-time] [-w write-aside]
```

**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 4 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 any number greater than or equal to 0.

**-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 any number greater than or equal to 0.

**-w *write-aside*** The smallest write request size, in blocks, that will bypass the cache and do directly to the disks. Legal values are any number greater than or equal to 0.

The following example changes LUN 3 to perform write caching and rebuild in four hours; it does not change the default owner.

```
raid5 -d sc4d2l0 chglun -l 3 -c write -d 0 -r 4
```

There is no output for *raid5* with the *chglun* parameter. Errors are printed to *stderr*.

### ***clearlog***

The *clearlog* parameter destroys the contents of the CHALLENGE RAID SP's error log. To clear the event log, use

```
raid5 -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. To reset statistics logging, use

```
raid5 -d device clearstats
```

**Note:** You must be root to use this parameter.

This command has no output.

### ***firmware***

To update the firmware on the CHALLENGE RAID SP, type as root

```
raid5 -d device firmware /usr/raid5/flarecode.bin
```

**Note:** The bus must be quiesced, with all applications disabled and only the RAID agent running. You must use this command every time you replace a failed unbound disk module (A0, B0, C0, or A3).

In this string, the new firmware image */usr/raid5/flarecode.bin* contains microcode that runs on the SP and also a microcode image destined for the SP PROM, which runs the power-on diagnostics.

This command has no output.

**Note:** Once the microcode has been downloaded, each SP in the cabinet reboots. the reboot process takes several minutes, during which time no command line interface commands are accepted except *getagent*. When the reboot is complete, *getagent* returns the new firmware version and all command line interface commands are accepted.

## ***getagent***

Use the *getagent* parameter with *raid5* to display information on devices controlled by the API:

```
raid5 getagent
```

Following is a sample output for one device; normally, the output would give information on all devices.

```
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 B-2 summarizes entries in the *raid5 getagent* output.

**Table B-2** Output of *raid5 getagent*

Entry	Meaning
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)

**Table B-2 (continued)**      Output of *raid5 getagent*

Entry	Meaning
Node	The <i>/dev/scsi</i> 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 <i>getagent</i> )
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	PROM revision currently present 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 *raid5 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.

To display cache information, use

```
raid5 -d device getcache
```

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

```

Table B-3 summarizes entries in the *raid5 getcache* output.

**Table B-3** Output of *raid5 getcache*

Entry	Meaning
Usable Cache	<p>Size of cache in megabytes to use for caching, not greater than the SP memory size as displayed by <i>raid5 getagent</i> (see “<i>getagent</i>,” 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).</p> <p>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.</p>
Page Size	User-specified page size in KB for caching; for example, 2 means a 2 KB page size.
Cache State	<p>Enabled: SP is fully functional.</p> <p>Disabled: SP not capable of or configured for caching.</p> <p>Synching: SP is synchronizing its cache with the peer SP.</p> <p>Enabling: cache is in the process of becoming enabled.</p> <p>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.</p> <p>Frozen: SP has finished quiescing and is now waiting for peer SP to finish quiescing or dumping.</p> <p>Dumping: SP is in the process of dumping contents of the cache to a private location on the disk.</p> <p>Initing: SP is initializing.</p>

**Table B-3 (continued)**      Output of *raid5 getcache*

Entry	Meaning
Low Watermark	<p>Percentage of dirty cache pages that, when reached during flush operations, will cause the SP to cease flushing the cache.</p> <p>Valid values are 0 through 100; the default is 50, regardless of whether caching is enabled or disabled.</p>
High Watermark	<p>Percentage of dirty cache pages which, when reached during flush operations, will cause the SP to cease flushing the cache.</p> <p>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.</p>
SP A Cache Pages	Number of cache pages currently owned by SP A
SP B Cache Pages	Number of cache pages currently owned by SP B
Unassigned Cache Pages	<p>Number of dirty cache pages not owned by A or B. This can happen when a unit is broken and there are no disks to which to flush the dirty pages</p>
Read Hit Ratio	<p>Percent of read requests to the SP that can be satisfied from the cache without requiring disk access</p>
Write Hit Ratio	<p>Percent of write requests to the controller that can be satisfied with the cache without requiring disk access</p>
Prct Dirty Cache Pages	<p>Percentage of cache pages that this SP owns that are dirty</p>
Prct Cache Pages Owned	<p>Percentage of cache pages owned by this controller when pages from other SP are taken into account</p>
Prct Read Flushes	<p>Percentage of reads to SP that caused the cache to flush a page(s)</p>
Prct Write Flushes	<p>Percentage of writes to SP that caused the cache to flush a page(s)</p>

## ***getcontrol***

To get general system information, use

```
raid5 -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: 5/5/1995
Day of the week: Friday
System Time: 12:43:54
```

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

## ***getcrus***

For state information on every field-replaceable unit in the CHALLENGE RAID storage system except disks in the disk modules, use

```
raid5 -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 *raid5 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***

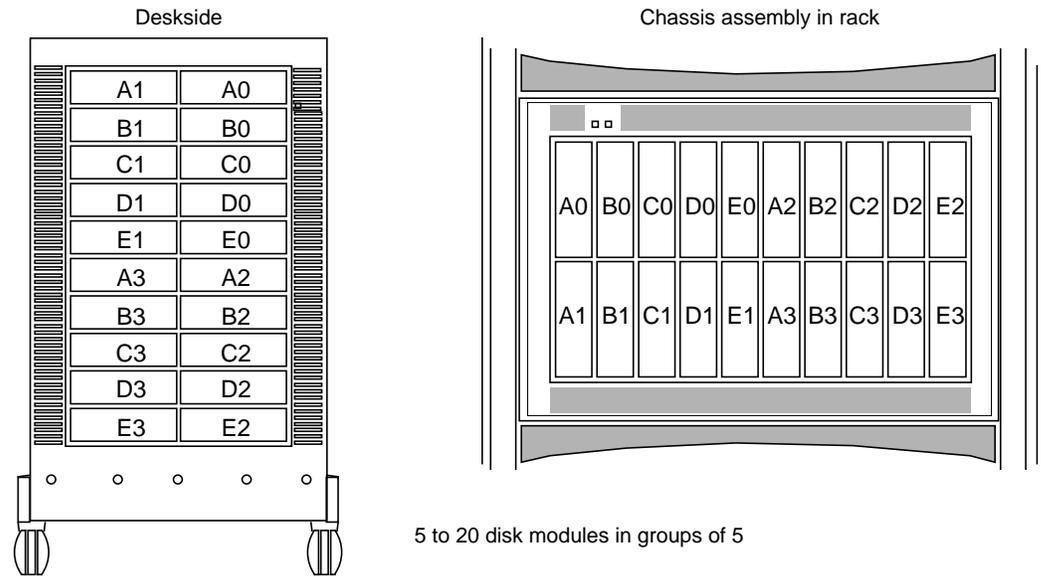
For information on all bound disks in the system, use

```
raid5 -d device getdisk
```

For information on a particular disk, use

```
raid5 -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 gets information disk A2.

```
raid5 -d sc4d2l0 getdisk a2
```

A sample output of this command follows.

```
A0 Vendor Id: SEAGATE
A0 Product Id: ST15150N
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 *raid5 getdisk*

Output	Meaning
Vendor Id	Manufacturer of disk drive
Product Id	2.1 GB disk: ST32550N 4.2 GB disk: ST15150N
Lun	Logical unit number to which this disk is bound

**Table B-5 (continued)**      Output of *raid5 getdisk*

<b>Output</b>	<b>Meaning</b>
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 Enabling: 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. To display the entire log, use

```
raid5 -d device getlog
```

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

```
raid5 -d device getlog +n
```

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

```
raid5 -d device getlog -n
```

Following is a possible output of the command `raid5 getlog +5`.

```
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 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

**Table B-6 (continued)** getlog Error Codes

Error Code	Explanation
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 <i>bind</i> 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 <i>chglun</i> 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
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

**Table B-6 (continued)**      getlog Error Codes

Error Code	Explanation
27	LUN does not exist
28	LUN already exists
29	Cannot get current working directory for <i>firmware</i> 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 in verbose mode (-v).

## **getlun**

To display information on a logical unit and the components in it, use

```
raid5 -d device getlun lun-number
```

The following example displays information about LUN 3.

```
raid5 -d sc4d210 getlun 3
```

**Note:** If numbers in the output look unreasonable, particularly for numerical statistics such as Queue Max, use *clearstats* to reset all statistics.

The following 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 *raid5 getcontrol*. See “getcontrol” earlier in this appendix.

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 *raid5 getlun* output.

**Table B-7** Output of *raid5 getlun*

Entry	Meaning
Type	RAID0, RAID1, RAID10, RAID5, or 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 <i>chglun</i> ; see “Dual Interfaces, Load Balancing, and Device Names,” in Chapter 4
Auto-trespass	Always Disabled
Auto-assign	Always Enabled
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 <i>chglun</i>
Idle Delay Time	Amount of time in 100-ms intervals that unit is below idle threshold; set with <i>chglun</i>
Write Aside Size	Smallest write-request size in blocks that can bypass the cache and go directly to the disk; set with <i>chglun</i>
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 that a rebuild should be performed in. 0 means rebuild as fast as possible, but means a degradation in host I/O performance

---

**Table B-7 (continued)**      Output of *raid5 getlun*

<b>Entry</b>	<b>Meaning</b>
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 <i>getdisk</i> )
Diskname Read	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.

To set up caching, use the *raid5 setcache* command:

```
raid5 -d device setcache enable | disable [-u usable] [-p page] [-l low] [-h high]
```

In this syntax, variables mean:

*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 *raid5 getagent* (see “getagent,” earlier in this appendix). Valid values are 0, 8, and 64:

- 0: user has selected to disable the cache
- other values: 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.

**Note:** For 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.

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:

```
raid5 -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:

```
raid5 -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 sc4d210 1 -u 8 -p 8 -l 50 -h 75
```

## ***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). For complete messages, it is recommended that you use the *-v* option.

**Note:** You must be root to use this parameter. For complete messages, it is recommended that you use the *-v* option.

To unbind disks, use

```
raid5 -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 *raid5 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 sc4d210 3 -o
```

---

# Index

## A

agent, *see getagent*  
auto-reassign, 82-84

## B

basic configuration, 25  
battery backup unit, 81  
    replacing, 81  
    status, 43  
BBU, *see* battery backup unit  
*bind*, 49-52, 99-102  
    legal configurations, 50, 100

## C

caching, 6, 12-13, 85-91  
    changing parameters, 91  
    disabling, 86, 120  
    enabling, 86, 120  
    failure, 87  
    flag in *bind*, 51, 101  
    flag in *chglun*, 56, 103  
    page size, 86, 120  
    requirements, 85  
    setting parameters, 86-87  
    size, 86, 120  
    statistics, 87-89  
    upgrading for, 90

CHALLENGE server system, xi, 4  
chassis, 8  
    assemblies, on one SCSI bus, 4  
    front view, 2-3  
*chglun*, 56-57, 102-104  
*clearlog*, 44, 104  
*clearstats*, 104  
CLI, *see* command line interface  
command line interface, 21-22, 97-122  
component, 5-10  
    getting information, 42-43  
    identifying failed, 79-82  
    replacement part numbers, 79  
configuration  
    disk, 49-57  
    storage system, 23-33  
        basic, 25  
        dual-bus/dual-initiator, 31-33  
        dual-interface/dual-processor, 26-27  
        features, 24  
        split-bus, 28-30  
conventions, xii

## D

data  
    rebuilding, 13  
    reconstruction, 13  
    redundancy, 13  
        RAID-5, 18-19

device, 37, 49, 98  
  name, 57  
  node, creating, 77

disk module, 10  
  A0, A3, B0, C0, D0, E0  
    moving, 14, 90  
  A0, B0, C0, D0, E0  
    failed, 61  
    failed, and caching, 87  
    removing, 73  
    required for caching, 85

add-on array, 72-77  
  installing, 73-77  
    order, 73  
  ordering, 72-73

busy light, 10

fault light, 10

getting information, 39-42

ID, 8

identifying failed, 59-62

installing replacement, 69-70

location, 8, 40, 61, 75

maintaining, 59-77

order, 8

ordering replacement, 63

ready light, 10

removing failed, 65-68

replacing, 14, 62-71

status lights, 10, 60

swapping, 14

unbinding, 64

unbound, replacing and updating firmware, 71

disk striping, *see* striping

dual-bus/dual-initiator configuration, 31-33, 57

dual-interface/dual-processor configuration, 26-27, 57

## E

electrostatic discharge damage (ESD), avoiding, 62

environment variable, 99

error codes, 114-116

event log  
  clearing, 44  
  displaying, 43-44

## F

fan module  
  closing, 47  
  opening, 46, 82-83  
  replacing, 80  
  status, 43  
  system operation when open, 82

field-replaceable unit, *see* component

*firmware*, 71, 104-105

## G

*getagent*, 38-39, 105-106

*getcache*, 87-89, 106-108

*getcontrol*, 39

*getcontrol*, 109

*getcrus*, 39-42, 109-110

*getdisk*, 39-42, 110-113

*getlog*, 43-44, 114-116

*getlun*, 53-56, 116-119

grouping disk modules, *see* *bind*

- 
- H**
- hotline, 60, 63, 72, 79
  - hot spare, 20-21
    - binding, 50, 100
- I**
- idle
    - delay time, 57, 103
    - threshold, 57, 103
  - IRIX version required, xi
- L**
- log, *see* event log
  - logical unit number, 49
    - see also* LUN
  - LUN, 49
    - changing parameters, 56-57
    - getting information, 53-56
    - number, assigning, 50
- M**
- model number for disk module, 10
  - multiple CHALLENGE RAID storage systems, 4
- O**
- operation, 35-47
- P**
- parity data, 13, 19
  - power supply, 80
    - replacing, 80
  - power supply status, 43
- R**
- RAID
    - defined, 1
    - level 0, 14
    - level 1, 15
    - level 1\_0, 16-17
    - level 5, 18-19
    - levels, binding, 50, 99
    - paper, 1
  - raid5*, 37, 97-122
    - parameters summarized, 97-98
    - syntax, 98
  - RaidAgentDevice environment variable, 99
  - rebuild time, 71, 100
    - bind*, 51
    - chglun*, 56, 103
  - restarting system, 45-47
- S**
- SCSI-2
    - bus, 18
      - and disk modules, 9
      - length limit, 4
    - interface, 5
  - service light, 36
  - setcache*, 86-87, 91, 120-121
  - shutting down system, 44
  - Silicon Graphics hotline, 60, 63, 72, 79

SP, *see* storage-control processor  
specifications, 93-95  
split-bus configuration, 28-30, 57  
status  
  checking, 36  
  light, 36  
storage-control processor, 6-7  
  failed  
    and caching, 87  
  powering off, 84  
  replacing, 81-84  
  status, 43  
striping, 12  
  default size, 12  
  RAID-1\_0, 16-17  
  RAID-5, 18-19  
  size, 51, 101  
system information, 39

**U**

*unbind*, 64, 121

**V**

VSC, *see* power supply, 43



---

## 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 007-2532-004.

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](mailto: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