IBM Parallel System Support Programs for AIX

Managing Shared Disks

SA22-7279-01

IBM Parallel System Support Programs for AIX

Managing Shared Disks

SA22-7279-01

Note!

Before using this information and the product it supports, be sure to read the general information under “Notices” on page xiii.

Second Edition (February 1998)

This is a major revision of SA22–7279–00.

This edition applies to

- Version 2 Release 4 of the IBM Parallel Systems Support Programs for AIX (PSSP) Licensed Program (5765–529)
- Version 2 Release 1.1 of the IBM Recoverable Virtual Shared Disk Licensed Program (5765–646)

and to all subsequent releases and modifications until otherwise indicated in new editions. Significant changes or additions to the text and illustrations are indicated by a vertical line (|) to the left of the change.

Order publications through your IBM representative or the IBM branch office serving your locality. Publications are not stocked at the address below.

IBM welcomes your comments. A form for your comments may be provided at the back of this publication., or you may address your comments to the following address:

International Business Machines Corporation
Department 55JA, Mail Station P384
522 South Road
Poughkeepsie, NY 12601-5400
United States of America

FAX (United States & Canada): 1+914+432-9405
FAX (Other Countries):
Your International Access Code +1+914+432-9405

IBMLink (United States customers only): KGNVMC(MHVRCFS)
IBM Mail Exchange: USIB6TC9 at IBMMAIL
Internet e-mail: mhvrcfs@vnet.ibm.com
World Wide Web: http://www.rs6000.ibm.com

If you would like a reply, be sure to include your name, address, telephone number, or FAX number.

Make sure to include the following in your comment or note:

- Title and order number of this book
- Page number or topic related to your comment

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© **Copyright International Business Machines Corporation 1997, 1998. All rights reserved.**
Note to U.S. Government Users — Documentation related to restricted rights — Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract with IBM Corp.

Contents

Notices . . . xiii
Trademarks . . . xiii
Publicly Available Software . . . xiv

About This Book . . . xv
Who Should Use This Book . . . xv
What You Need to Know First . . . xv
How This Book is Organized . . . xvi
Typography and Terminology . . . xvii
Accessing Online Information . . . xviii
Obtaining Documentation . . . xviii
Related Publications . . . xviii
RS/6000 SP Publications . . . xviii
Parallel System Support Programs for AIX Publications . . . xviii
IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Publication . . . xix
General Parallel File System for AIX Publication . . . xix
LoadLeveler Publications . . . xix
Performance Toolbox (PTX) Publication . . . xix
Performance Monitor Publication . . . xix
Parallel Environment for AIX Publications . . . xix
Client Input Output/Sockets (CLIO/S) Publications . . . xix
Parallel I/O File System Publications . . . xx
Network Tape Access and Control System for AIX (NetTAPE) Publications . . . xx
Other IBM Publications . . . xx
International Technical Support Organization Publications (Red Books) . . . xx
AIX and RS/6000 Publications . . . xx
Service . . . xxi
Network Queueing System/MVS (NQS/MVS) . . . xxii
Network Connectivity . . . xxii
RS/6000 SP Switch Router . . . xxii
Adapters . . . xxii
Non-IBM Publications . . . xxiii
Parallel Computing . . . xxiii
Tcl . . . xxiii
Manual Pages for Public Code . . . xxiii

Chapter 1. An Overview of IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk Licensed Program 1
What's New in This Edition? . . . 1
IBM Virtual Shared Disk Overview . . . 1
IBM Virtual Shared Disk Restrictions . . . 2
Data Striping Device (HSD) Overview . . . 2
Comparing HSD and LVM Striping . . . 3
When to Use the HSD . . . 4
Data Striping Device Restrictions . . . 4
IBM Recoverable Virtual Shared Disk Licensed Program Overview . . . 4
IBM Virtual Shared Disk Recovery with Twin-tailed Disks . . . 5
IBM Recoverable Virtual Shared Disk Restrictions . . . 6

© Copyright IBM Corp. 1997, 1998

Chapter 2. Installing IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk Licensed Program 7
Planning to Use the Shared Disk Components for the First Time 7
Planning for IBM Virtual Shared Disks 7
Planning for HSDs 7
Planning for the IBM Recoverable Virtual Shared Disk Licensed Program . . 8
Migrating from AIX 3.2.5 to AIX 4.1 and Later Releases 8
Limits on the Number of Physical Partitions per Disk 8
Buddy Buffer Size 9
Migrating from PSSP 2.2 to PSSP 2.3 and Later Releases 9
Before PSSP Migration 9
Perform PSSP Migration 10
When PSSP Migration is Complete 10
Before You Install the Current Releases 11
PSSP Installation Considerations 11
Installation Packages 11
Detailed Space Requirements 12
Hardware Requirements for the IBM Recoverable Virtual Shared Disk Licensed Program 13
An Overview of the Installation Process 13
Step 1: Unconfigure the Existing IBM Virtual Shared Disks 13
Step 2: Install PSSP Version 2 Release 4 13
Step 3: Install the Desired Software at the Control Workstation 14
Step 4: Install the Desired Software on all the Nodes in the Configuration . 14
Installing the Shared Disk Software Components 14
Installation Notes for IBM Recoverable Virtual Shared Disk 15
What to Do after Installation 15
New IBM Virtual Shared Disk Users 15
New HSD Users 15
New IBM Recoverable Virtual Shared Disk Users 16
Adding IBM Recoverable Virtual Shared Disk to a System Where IBM Virtual Shared Disk Is Already in Use 16
When You Are Setting up New Physical Disks 17
When You Are Recabling without Adding New Physical Disks 17

Chapter 3. Introducing the IBM Virtual Shared Disk Perspective 19
The IBM Virtual Shared Disk Perspective Primary Window 19
The Menu Bar 19
The Tool Bar 20
Pane Controls 21
The Panes 21
The Information Area 23

Chapter 4. Creating and Managing Shared Disks 25
Setting up New IBM Virtual Shared Disks 25
Preparing to Create IBM Virtual Shared Disks and Data Striping Devices . 26
Creating IBM Virtual Shared Disks with the createvsd Command 29
Creating HSDs with the createhsd Command 34
Configuring and Unconfiguring IBM Virtual Shared Disks and HSDs 39
Verifying an IBM Virtual Shared Disk 40
Starting IBM Virtual Shared Disks and HSDs 41
Displaying Information about IBM Virtual Shared Disk and HSD Nodes, Clients, and Definitions 41
Displaying Information about IBM Virtual Shared Disk and HSD Nodes . . . 41

Displaying Information about IBM Virtual Shared Disk Clients 41
Displaying Information about HSD Definitions and Clients 42
Displaying IBM Virtual Shared Disk and HSD Statistics and Diagnostics 43
Displaying IBM Virtual Shared Disk Node Statistics 43
Displaying HSD Node Statistics 44
Resetting IBM Virtual Shared Disk and HSD Statistics 44
Running IBM Virtual Shared Disk Diagnostics 45
Changing the States of IBM Virtual Shared Disks 45
Starting an IBM Virtual Shared Disk 45
Preparing an IBM Virtual Shared Disk 45
Resuming an IBM Virtual Shared Disk 46
Suspending an IBM Virtual Shared Disk 46
Stopping an IBM Virtual Shared Disk 46
Making Changes to IBM Virtual Shared Disks on Twin-Tailed Volume Groups 47
Making a Change to Only One Side of a Twin-Tailed Volume Group 48
Making a Change to Both Sides of a Twin-Tailed Volume Group 48
Making a Change to Both Sides of a Twin-Tailed Volume Group with the Volume Group Varied Off 49
Removing an IBM Virtual Shared Disk or Data Striping Device 49
Adding a Server Node to an Existing IBM Virtual Shared Disk Configuration . 49
Removing a Node from an Existing IBM Virtual Shared Disk Configuration . . 50

Chapter 5. Performance and Tuning Considerations for IBM Virtual Shared Disks and Data Striping Devices 53
Tuning IBM Virtual Shared Disk Performance 53
IBM Virtual Shared Disk Tunable Parameters 54
Logical Volume Manager (LVM) Tuning Considerations 54
The High Performance Switch/Scalable POWERparallel Switch 55
Buffer Allocation 57
The Cache Buffer 57
Maximum I/O Request Size 58
Request Blocks 59
pbufs 59
Tuning HSD Performance 60

Chapter 6. Application Programming Considerations 63
Application Programming Considerations for IBM Virtual Shared Disks 63
Data Integrity 63
IBM Virtual Shared Disk Transmission Protocol 63
Getting IBM Virtual Shared Disk Information with C Interfaces 65
Application Programming Considerations for HSDs 65
Allocating I/O Buffers 65
Getting HSD Information with C Interfaces 65
Making Your Application Recoverable 66
Preserving Data Integrity During Application Recovery 67

Chapter 7. What You Need to Know about the IBM Recoverable Virtual Shared Disk Software 69
The IBM Recoverable Virtual Shared Disk Recovery Subsystems 69
The rvsd Subsystem 69
Disk Cable and Disk Adapter Failures 71
Communication Adapter Failures 71
The hc Subsystem 71
Startup Time 72

IBM HSD Recovery Considerations 73

Chapter 8. IBM Recoverable Virtual Shared Disk Recovery Scenarios . . 75
Recognizing Recovery 75
Planning For Recovery 75
IBM Virtual Shared Disk Client or Server Node Failure 76
Switch Failure Scenarios 77
When Adapter Recovery is Enabled 77
When Adapter Recovery is Disabled 77
HA Topology Services or Recovery Service Daemon Failure 78
Hardware (EIO) Failures 78

Chapter 9. Diagnosis and Messages Information 81
Diagnosis Information 81
Messages 81

Chapter 10. IOCTL Subroutines 83
fence subroutine 83
Purpose 83
Syntax 83
Parameters 83
Description 84
Return Values 84
Error Values 84
Examples 85

Appendix A. IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands 87
cfghsd 87
cfghsdvsd 87
cfgvsd 87
cksumvsd 87
createhsd 88
createvsd 88
ctlhsd 88
ctlvsd 89
defhsd 89
defvsd 89
fencevsd 89
ha_vsd 90
ha.vsd 90
hc.vsd 90
hsdatalst 90
hsdvts 91
lshsd 91
lsvsd 91
preparevsd 91
removehsd 92
removevsd 92
resumevsd 92
spvsd 92
startvsd 93
statvsd 93
stopvsd 93

suspendvsd . . . 93
ucfghsd . . . 94
ucfghsdvsd . . . 94
ucfgvsd . . . 94
undefhsd . . . 94
undefvsd . . . 95
unfencevsd . . . 95
updatehsd . . . 95
updatevsdnode . . . 95
updatevsdtab . . . 96
verparvsd . . . 96
vsdatalst . . . 96
vsdchgserver . . . 96
vsddiag . . . 97
vsdelnode . . . 97
vsdelvg . . . 97
vsdnode . . . 97
vsdsklst . . . 98
vsdvg . . . 98
vsdvgts . . . 98
vsdvts . . . 98

Appendix B. Single-Node Command and SMIT Interfaces for IBM Virtual Shared Disks, Data Striping Devices and the IBM Recoverable Virtual Shared Disk Licensed Product . . . 99
Command and SMIT Interfaces for IBM Virtual Shared Disks . . . 99
Entering Node Information in the SDR for One Node . . . 99
Entering Global Volume Group Information in the SDR for One Node . . . 99
Entering IBM Virtual Shared Disk Information in the SDR for One Node . . . 100
An Example of Defining an IBM Virtual Shared Disk on One Node at a Time . . . 101
Displaying IBM Virtual Shared Disk Information Stored in the SDR . . . 103
Listing Node Information . . . 104
Listing Global Volume Group Information . . . 104
Listing Defined IBM Virtual Shared Disks . . . 104
Deleting IBM Virtual Shared Disk Information from the SDR . . . 104
Managing IBM Virtual Shared Disks . . . 105
Configuring an IBM Virtual Shared Disk . . . 107
Starting an IBM Virtual Shared Disk . . . 107
Preparing an IBM Virtual Shared Disk . . . 108
Resuming an IBM Virtual Shared Disk . . . 108
Suspending an IBM Virtual Shared Disk . . . 108
Stopping an IBM Virtual Shared Disk . . . 109
Unconfiguring an IBM Virtual Shared Disk . . . 109
Displaying Managed IBM Virtual Shared Disk Characteristics . . . 110
Displaying Managed IBM Virtual Shared Disk Statistics . . . 110
Displaying IBM Virtual Shared Disk Device Driver Statistics . . . 110
Setting and Displaying IBM Virtual Shared Disk Device Driver Operational Parameters . . . 111
Testing to Verify an IBM Virtual Shared Disk (the Verification Test Suite) . . . 112
Command and SMIT Interfaces for Data Striping Devices . . . 112
Defining HSD Device Driver Information in the SDR . . . 112
An Example of Creating an HSD . . . 113
Displaying Data Striping Device Information in the SDR . . . 114

Undefining HSD Information in the SDR 115
Managing Data Striping Devices 115
Verifying Data Striping Device Installation 117
Management and Usage Notes 118
Command and SMIT Interfaces for the IBM Recoverable Virtual Shared Disk Licensed Program 118
Defining Twin-tailed LVM Components in the IBM Recoverable Virtual Shared Disk Environment 118
Coordinating the IBM Recoverable Virtual Shared Disk Environment . . . 119
Overview of Shared Volume Groups Administrative Tasks 119
Prerequisite Tasks 119
Maintaining Twin-Tailed Volume Groups 120
Creating a Twin-Tailed Volume Group 120
Adding Physical Volumes To a Twin-Tailed Volume Group 124
Adding or Removing Logical Volumes to or from a Twin-Tailed Volume Group 128
Reducing a Twin-Tailed Volume Group 130
Removing a Twin-Tailed Volume Group 136

Appendix C. Setting Up 7134 Disks as Twin-tailed Volume Groups . . . 139

Appendix D. Glossary 141

Index 143

Figures

1. An IBM Virtual Shared Disk IP Network Implementation 2
2. HSD Stripes Data Across Two or More VSDs 3
3. Simplified View of a Twin-tailed Disk 5
4. The Secondary Node Serves after a Primary Node Failure 5
5. The Primary Node Is the Server Again after Recovery 6
6. The IBM Virtual Shared Disk Perspective Menu Bar 19
7. The IBM Virtual Shared Disk Perspective Tool Bar 20
8. The Syspars Pane 21
9. The IBM VSD Nodes/Clients Pane 22
10. The HSD Definitions Pane 23
11. A Node-Pair Configuration 39
12. IBM Virtual Shared Disk States and Associated Commands 47
13. An Example of an IBM Virtual Shared Disk Configuration 101
14. IBM Virtual Shared Disk States and Associated Commands 106

© Copyright IBM Corp. 1997, 1998

Tables

1. Possibilities for Interaction in the Same System Partition: PSSP and IBM Recoverable Virtual Shared Disk Levels 8
2. Space Used by Installation Process 12
3. Additional Space Used during Execution on the Control Workstation . . . 12
4. Additional Space Used during Execution on the Nodes 13
5. Recovery Actions When Nodes Fail 70
6. IBM Virtual Shared Disk Information 101
7. IBM Virtual Shared Disk Node Information 102
8. IBM Virtual Shared Disk Volume Group Information 102
9. IBM Virtual Shared Disk Definition Information 103
10. General Procedure for Changing a Twin-tailed LVM Component 119
11. Steps to Create a Twin-tailed Volume Group 120
12. Procedure for Extending a Twin-tailed LVM Component 124
13. Procedure for Modifying a Twin-tailed LVM Component 128
14. Procedure for Reducing a Twin-Tailed Volume Group 131
15. Steps to Remove a Shared Volume Group 136

© Copyright IBM Corp. 1997, 1998

Notices

References in this publication to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program, or service. Evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM, are the user's responsibility.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
500 Columbus Avenue
Thornwood, NY 10594
USA

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
Mail Station P300
522 South Road
Poughkeepsie, NY 12601-5400
USA
Attention: Information Request

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

Trademarks

The following terms are trademarks of the IBM Corporation in the United States or other countries or both:

AIX

IBM

RS/6000

RS/6000 Scalable POWERparallel Systems

Microsoft, Windows, and the Windows 95 logo are trademarks or registered trademarks of Microsoft Corporation.

© Copyright IBM Corp. 1997, 1998

UNIX is a registered trademark in the United States and other countries licensed exclusively through X/Open Company Limited.

Other company, product, and service names, which may be denoted by a double asterisk (**), may be trademarks or service marks of others.

Publicly Available Software

This product includes software that is publicly available:

expect Programmed dialogue with interactive programs

Kerberos Provides authentication of the execution of remote commands

NTP Network Time Protocol

Perl Practical Extraction and Report Language

SUP Software Update Protocol

Tcl Tool Command Language

TclX Tool Command Language Extended

Tk Tcl-based Tool Kit for X-windows

This book discusses the use of these products only as they apply specifically to the RS/6000 SP system. The distribution for these products includes the source code and associated documentation. (Kerberos does not ship source code.) **/usr/lpp/ssp/public** contains the compressed **tar** files of the publicly available software. (IBM has made minor modifications to the versions of Tcl and Tk used in the SP system to improve their security characteristics. Therefore, the IBM-supplied versions do not match exactly the versions you may build from the compressed **tar** files.) All copyright notices in the documentation must be respected. You can find version and distribution information for each of these products that are part of your selected install options in the **/usr/lpp/ssp/README/ssp.public.README** file.

About This Book

This book describes how to install, manage, and use the disk management facilities of the SP system-the IBM Virtual Shared Disk and Data Striping Device (HSD) components of IBM Parallel System Support Programs (PSSP) and the IBM Recoverable Virtual Shared Disk licensed program. It also provides an overview of these components and explains how they can add reliability to your data storage.

This book applies to IBM Parallel System Support Programs for AIX, Version 2 Release 4, and to IBM Recoverable Virtual Shared Disk Version 2 Release 1.1.

To find out what version of PSSP is running on your control workstation, enter the following:

```
SDRGetObjects SP code_version
```

In response, the system displays something similar to:

```
code_version
PSSP-2.4
```

To find out what version of PSSP is running on the nodes of your system, enter the following from your control workstation:

```
splst_versions -G -t
```

in response, the system displays something similar to:

```
1 PSSP-2.4
2 PSSP-2.4
7 PSSP-2.4
8 PSSP-2.4
```

If the response indicates **PSSP-2.4**, this book applies to the version of PSSP that is running on your system.

If you are running mixed levels of PSSP in a system partition, be sure to maintain and refer to the appropriate documentation for whatever versions of PSSP you are running.

Who Should Use This Book

Use this book if you are considering using IBM Virtual Shared Disks, HSDs, or the IBM Recoverable Virtual Shared Disk software.

You can use the information in this book to install, manage, and use these shared disk management components and to write application programs that use their facilities.

What You Need to Know First

It is particularly important for you to understand the AIX operating system, including the Logical Volume Manager subsystem, and to refer to *AIX System Management Guide: Operating Systems and Devices*, SC23-2525.

© Copyright IBM Corp. 1997, 1998

You will probably need to be familiar with and have the documentation for the following:

- IBM Parallel System Support Programs for AIX.
- Installation procedures for this release in *IBM Parallel System Support Programs for AIX: Installation and Migration Guide*, GC23-3897
- The IBM Virtual Shared Disk commands in *IBM Parallel System Support Programs for AIX: Command and Technical Reference*, GC23-3900
- The installation and configuration information that came with your twin-tailed disk hardware
- The Logical Volume Manager information in *IBM AIX Version 4 System Management Guide: Operating System and Devices*, SC23-2525, and *IBM AIX Version 4.3 System Management Guide: Operating System and Devices*, SC23-4126
- *AIX Kernel Extensions and Device Support Programming Concepts*, SC23-2207
- *AIX Performance and Tuning Guide*, SC23-2365
- Planning information in *IBM Parallel System Support Programs for AIX: System Planning Guide*, GC23-3902
- Availability services information in *IBM Parallel System Support Programs for AIX: Group Services Programming Guide and Reference*, SC28-1675
- Diagnosis information in *IBM Parallel System Support Programs for AIX: Diagnosis and Messages*

How This Book is Organized

This book is divided into the following chapters:

- Chapter 1, " An Overview of IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk Licensed Program" on page 1 provides an overview of the IBM Virtual Shared Disk and Data Striping Device components of the SP system and of the IBM Recoverable Virtual Shared Disk licensed program.
- Chapter 2, " Installing IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk Licensed Program" on page 7 lists the installation requirements and describes the procedure for installing, configuring, and starting up IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk software.
- Chapter 3, " Introducing the IBM Virtual Shared Disk Perspective" on page 19 describes the graphical user interface for managing IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk software.
- Chapter 4, " Creating and Managing Shared Disks" on page 25 tells how to create and manage IBM Virtual Shared Disks, Data Striping Devices, and IBM Recoverable Virtual Shared Disks.
- Chapter 5, "Performance and Tuning Considerations for IBM Virtual Shared Disks and Data Striping Devices" on page 53 describes performance and tuning considerations for IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk program product.

- Chapter 6, “Application Programming Considerations” on page 63 contains programming considerations for IBM Virtual Shared Disks, Data Striping Devices, and recoverability.
- Chapter 7, “What You Need to Know about the IBM Recoverable Virtual Shared Disk Software” on page 69 tells you how the IBM Recoverable Virtual Shared Disk licensed program works.
- Chapter 8, “IBM Recoverable Virtual Shared Disk Recovery Scenarios” on page 75 describes IBM Recoverable Virtual Shared Disk recovery procedures.
- Chapter 9, “Diagnosis and Messages Information” on page 81 contains IBM Recoverable Virtual Shared Disk error messages.
- Appendix A, “IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands” on page 87 shows the syntax for the IBM Virtual Shared Disk, Data Striping Device, and IBM Recoverable Virtual Shared Disk commands.
- Appendix C, “Setting Up 7134 Disks as Twin-tailed Volume Groups” on page 139 contains procedures for setting up IBM 7134 disks as twin-tailed volume groups.
- Appendix B, “ Single-Node Command and SMIT Interfaces for IBM Virtual Shared Disks, Data Striping Devices and the IBM Recoverable Virtual Shared Disk Licensed Product” on page 99 tells you how to manage IBM Virtual Shared Disks by using the commands that operate on one node at a time.

The back of the book includes a glossary and an index.

| Vertical bars (|) to the left of the text in this book indicate changes or additions.

Typography and Terminology

This book uses the following typographical conventions:

Typographic	Usage
Bold	• **Bold** words or characters represent system elements that you must use literally, such as commands, flags, and path names. • **Bold** words also indicate the first use of a term included in the glossary.
Italic	• *Italic* words or characters represent variable values that you must supply. • *Italics* are also used for book titles and for general emphasis in text.
`Constant width`	Examples and information that the system displays appear in `constant width` typeface.
[ ]	Brackets enclose optional items in format and syntax descriptions.
{ }	Braces enclose a list from which you must choose an item in format and syntax descriptions.
|	A vertical bar separates items in a list of choices. (In other words, it means “or.”)
< >	Angle brackets (less-than and greater-than) enclose the name of a key on the keyboard. For example, <**Enter**> refers to the key on your terminal or workstation that is labeled with the word Enter.
...	An ellipsis indicates that you can repeat the preceding item one or more times.
<**Ctrl-***x***>**	The notation <**Ctrl-***x***>** indicates a control character sequence. For example, <**Ctrl-c>** means that you hold down the control key while pressing <**c>**.

The term *SP* or *SP system* refers to the current system partition, if you have partitioned your system. Otherwise, the term refers to the physical SP system.

Accessing Online Information

In order to use the PSSP man pages or access the PSSP online (HTML) publications, the **ssp.docs** file set must first be installed. To view the PSSP online publications, you also need access to an HTML document browser such as Netscape. An index to the HTML files that are provided with the **ssp.docs** file set is installed in the **/usr/lpp/ssp/html** directory.

Obtaining Documentation

You can view this book or download a PostScript version of it from the IBM RS/6000 web site at **http://www.rs6000.ibm.com**. At the time this manual was published, the full path was **http://www.rs6000.ibm.com/resource/aix_resource/sp_books**. However, the structure of the RS/6000 web site can change over time.

Related Publications

Here are some related publications.

RS/6000 SP Publications

- *IBM RS/6000 SP: Planning, Volume 1, Hardware and Physical Environment*, GA22-7280
- *IBM RS/6000 SP: Planning, Volume 2, Control Workstation and Software Environment*, GA22-7281
- *IBM RS/6000 SP: Maintenance Information, Volume 1, Installation and CE Operations*, GC23-3903
- *IBM RS/6000 SP: Maintenance Information, Volume 2, Maintenance Analysis Procedures and Parts Catalog*, GC23-3904

Parallel System Support Programs for AIX Publications

- *IBM Parallel System Support Programs for AIX: Administration Guide*, GC23-3897
- *IBM Parallel System Support Programs for AIX: Installation and Migration Guide*, GC23-3898
- *IBM Parallel System Support Programs for AIX: Diagnosis and Messages Guide*, GC23-3899
- *IBM Parallel System Support Programs for AIX: Command and Technical Reference*, GC23-3900
- *IBM Parallel System Support Programs for AIX: Event Management Programming Guide and Reference*, SC23-3996
- *IBM Parallel System Support Programs for AIX: Group Services Programming Guide and Reference*, SC28-1675
- *IBM Parallel System Support Programs for AIX: Licensed Program Specification*, GC23-3901

As an alternative to ordering the individual books, you can use SBOF-8587 to order the entire SP software library.

IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Publication

- *IBM Parallel System Support Programs for AIX: Managing Shared Disks*, SA22-7279

General Parallel File System for AIX Publication

- *IBM General Parallel File System for AIX: Installation and Administration Guide*, SA22-7278

LoadLeveler Publications

- *Using and Administering LoadLeveler*, SC23-3989
- *IBM LoadLeveler: Licensed Program Specification*, GH23-0040

Performance Toolbox (PTX) Publication

- *IBM Performance Toolbox 1.2 and 2 for AIX: Guide and Reference*, SC23-2625

Performance Monitor Publication

- *IBM Performance Toolbox Parallel Extensions for AIX: Guide and Reference*, SC23-3997

Parallel Environment for AIX Publications

- *IBM Parallel Environment for AIX: Hitchhiker's Guide*, GC23-3895
- *IBM Parallel Environment for AIX: Operation and Use, Volume 1*, SC28-1979
- *IBM Parallel Environment for AIX: Operation and Use, Volume 2*, SC28-1980
- *IBM Parallel Environment for AIX: Installation Guide*, GC28-1981
- *IBM Parallel Environment for AIX: Messages*, GC28-1982
- *IBM Parallel Environment for AIX: MPL Programming and Subroutine Reference*, GC23-3893
- *IBM Parallel Environment for AIX: MPI Programming and Subroutine Reference*, GC23-3894
- *IBM Parallel Environment for AIX: Licensed Program Specification*, GC23-3896

As an alternative to ordering the individual books, you can use SBOF-8588 to order the entire Parallel Environment for AIX library.

Client Input Output/Sockets (CLIO/S) Publications

- *IBM CLIO/S General Information*, GC23-3879
- *IBM CLIO/S Licensed Program Specification*, GC23-3789
- *IBM CLIO/S User's Guide and Reference*, GC28-1676

Parallel I/O File System Publications

- *IBM AIX Parallel I/O File System: Fact Sheet*, G325-0649
- *IBM AIX Parallel I/O File System: Installation, Administration, and Use*, SH34-6065
- *IBM AIX Parallel I/O File System: Licensed Program Specification*, GH34-6066

Network Tape Access and Control System for AIX (NetTAPE) Publications

- *IBM NetTAPE General Information*, GC23-3990
- *IBM NetTAPE User's Guide and Reference*, available from your IBM representative

---

Other IBM Publications

Here are some other IBM publications that you may find helpful.

International Technical Support Organization Publications (Red Books)

- *IBM International Technical Support Centers Implementing High Availability on RS/6000 SP*, SG24-4742
- *IBM International Technical Support Centers RS/6000 SP High Availability Infrastructure*, SG24-4838
- *IBM International Technical Support Centers RS/6000 SP PSSP 2.2 Technical Presentation*, SG24-4868
- *IBM International Technical Support Centers RS/6000 SP PSSP 2.3 Technical Presentation*, SG24-2080

AIX and RS/6000 Publications

- *IBM AIX Version 4 Getting Started*, SC23-2527
- *IBM AIX Version 4.3 Quick Beginnings*, SC23-4114
- *IBM AIX General Concepts and Procedures for RS/6000*, GC23-2202
- *IBM AIX Version 4 Files Reference*, SC23-2512
- *IBM AIX Version 4.3 Files Reference*, SC23-4168
- *IBM AIX Version 4 System Management Guide: Communications and Networks*, SC23-2526
- *IBM AIX Version 4.3 System Management Guide: Communications and Networks*, SC23-4127
- *IBM AIX Version 4.1 Installation Guide*, SC23-2550
- *IBM AIX Version 4.2 Installation Guide*, SC23-1924
- *IBM AIX Version 4.3 Installation Guide*, SC23-4112
- *IBM AIX Version 4 Commands Reference*, SBOF-1851 (all volumes)
- *IBM AIX Version 4.3 Commands Reference*, SBOF-1877 (all volumes)
- *IBM AIX Versions 3.2 and 4 Performance Tuning Guide*, SC23-2365

- *IBM AIX Version 4 Messages Guide and Reference*, SC23-2641
- *IBM AIX Version 4.3 Messages Guide and Reference*, SC23-4129
- *IBM AIX Version 4.1 Network Installation Management Guide and Reference*, SC23-2627
- *IBM AIX Version 4.2 Network Installation Management Guide and Reference*, SC23-1926
- *IBM AIX Version 4.3 Network Installation Management Guide and Reference*, SC23-4113
- *IBM AIX Version 4 System Management Guide: Operating System and Devices*, SC23-2525
- *IBM AIX Version 4.3 System Management Guide: Operating System and Devices*, SC23-4126
- *IBM AIX Version 4 General Programming Concepts: Writing and Debugging Programs*, SC23-2533
- *IBM AIX Version 4.3 General Programming Concepts: Writing and Debugging Programs*, SC23-4128
- *IBM AIX Version 4 Communications Programming Concepts*, SC23-2610
- *IBM AIX Version 4.3 Communications Programming Concepts*, SC23-4124
- *Diskless Workstation Management Guide*, SC23-2433
- *C++ for AIX/6000: Language Reference*, SC09-1606
- *C++ for AIX/6000: Standard Class Library Reference*, SC09-1604
- *C++ for AIX/6000: User's Guide*, SC09-1605

Service

- *IBM AIX Version 3.2 Problem Solving Guide and Reference*, SC23-2204
- *IBM AIX Version 4 Problem Solving Guide and Reference*, SC23-2606
- *IBM AIX Version 4.3 Problem Solving Guide and Reference*, SC23-4123
- *Information Manual for Diagnostics Micro Channel Bus*, SA23-2765
- *7012 300 Series Operator Guide*, SA23-2623
- *7012 300 Series Installation and Service Guide*, SA23-2624
- *7013 500 Series Operator Guide*, SA38-0530
- *7013 500 Series Installation and Service Guide*, SA38-0531
- *Electrical Safety for IBM Customer Engineers*, S229-8124
- *Service Information Binder*, SX33-6060
- *7015 Model R30 CPU Enclosure Operator Guide*, SA23-2742
- *7015 Model R30 CPU Enclosure Installation and Service Guide*, SA23-2743
- *Supplemental Information for 7012 G Series Models, 7013 J Series Models, 7015 Models R30 and R50* (No order number — this book comes as part of the ship group with the SP hardware)

Network Queueing System/MVS (NQS/MVS)

- *IBM NQS/MVS Installing and Administering*, SC23-0232
- *IBM NQS/MVS Client User's Guide*, SC23-0231
- *IBM NQS/MVS General Information*, GC23-3684
- *IBM NQS/MVS Licensed Program Specifications*, GC23-0233

Network Connectivity

- *IBM LAN Cabling System Planning and Installation Guide*, GA27-3361
- *IBM Cabling System Optical Fiber Planning and Installation Guide*, GA27-3943
- *IBM 8250/8260/8285 Planning and Site Preparation Guide*, GA33-0285
- *IBM 6611 Network Processor: Introduction and Planning Guide*, GK2T-0334

RS/6000 SP Switch Router

Order according to RS/6000 SP Switch Router model:

- *Getting Started*
- *Configuration Guide*
- *Reference Guide*

You can order the RS/6000 SP Switch Router as the IBM 9077.

Adapters

- *SP Switch Router Adapter Guide*, GA22-7310
- *FDDI Introduction and Planning Guide*, GA27-3892
- *FDDI User's Guide and Programming Reference*, SC28-2823
- *Planning for Fiber Optic Channel Links*, GA23-0367
- *IBM Token-Ring Network Introduction and Planning Guide*, GA27-3677
- *RS/6000 Token Ring Adapter Card*, G511-1681
- *HiPPI User's Guide and Programmer's Reference*, SA23-0369 and SA23-2488
- *AIX Parallel and ESCON Channel Tape Attachment/6000 Installation and User's Guide*, GA32-0311
- *9334 SCSI Expansion Units Operator Guide*, GA33-3232
- *9334 Models 010 and 011 SCSI Expansion Units: Installation and Service Guide*, SY33-0165
- *9334 Models 500 and 501 SCSI Expansion Units: Installation and Service Guide*, SY33-0167
- *SCSI-2 Fast/Wide Adapter*, SC23-2646
- *IBM SCSI-2 Fast/Wide Adapter/A Technical Reference*, S83G-7545
- *Turboways 100 User's Guide ATM*, GA27-4057
- *9333 Model 010 and 011 High-Performance Disk-Drive Subsystem Operator Guide*, GA33-3208

- *9333 Model 010 and 011 High-Performance Disk-Drive Subsystem Installation and Service Guide*, SY33-0161
- *9333 Model 010 and 011 High-Performance Disk-Drive Subsystem Hardware Technical Information*, SA33-3209
- *9333 Model 500 and 501 High-Performance Disk-Drive Subsystem Operator Guide*, GA33-3234
- *9333 Model 500 and 501 High-Performance Disk-Drive Subsystem Installation and Service Guide*, SY33-0168
- *9333 Model 500 and 501 High-Performance Disk-Drive Subsystem Hardware Technical Information*, SA33-3235
- *IBM SCSI Tape Drive, Medium Changer, and Library Device Drivers Installation and User's Guide*, GC35-0154

Non-IBM Publications

Here are some non-IBM publications that you may find helpful.

Parallel Computing

- Almasi, G., Gottlieb, A., *Highly Parallel Computing*, Benjamin-Cummings Publishing Company, Inc., 1989.
- Gropp, W., Lusk, E., Skjellum, A., *Using MPI*, The MIT Press, 1994.
- Message Passing Interface Forum, MPI: *A Message-Passing Interface Standard*, Version 1.1, University of Tennessee, Knoxville, Tennessee, June 6, 1995.
- Foster, I., *Designing and Building Parallel Programs*, Addison-Wesley, 1995.
- Phister, Gregory, F., *In Search of Clusters*, Prentice Hall, 1995.

Tcl

- Ousterhout, John K., *Tcl and the Tk Toolkit*, Addison-Wesley, Reading, MA, 1994, ISBN 0-201-63337-X.

Manual Pages for Public Code

The following manual pages for public code are available in this product:

SUP	/usr/lpp/ssp/man/man1/sup.1
NTP	/usr/lpp/ssp/man/man8/xntpd.8
	/usr/lpp/ssp/man/man8/xntpdc.8
Perl (Version 4.036)	/usr/lpp/ssp/perl/man/perl.man
	/usr/lpp/ssp/perl/man/h2ph.man
	/usr/lpp/ssp/perl/man/s2p.man
	/usr/lpp/ssp/perl/man/a2p.man
Perl (Version 5.003)	Man pages are in the /usr/lpp/ssp/perl5/man/man1 directory

Manual pages and other documentation for Tcl, TclX, Tk, and expect can be found in the compressed **tar** files located in the **/usr/lpp/ssp/public** directory.

Chapter 1. An Overview of IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk Licensed Program

This chapter provides a basic overview of the IBM Virtual Shared Disk and Data Striping Device components of the SP system and of the IBM Recoverable Virtual Shared Disk licensed program. Later chapters give more detailed information about these software features and how to manage them.

What's New in This Edition?

Communication adapter recovery has been added to support the SP switch and Ethernet adapters, allowing IBM Recoverable Virtual Shared Disk to respond to their failures as though a node failure occurred, and initiating a failover to the secondary server.

Revisions were made and new illustrations were added to enhance existing topics.

IBM Virtual Shared Disk Overview

An IBM Virtual Shared Disk is a subsystem that lets application programs executing on different nodes of a system partition access a raw logical volume as if it were local at each of the nodes. See Figure 1 on page 2 for an illustration of a simplified IBM Virtual Shared Disk implementation. Each IBM Virtual Shared Disk corresponds to a logical volume that is actually local at *one* of the nodes, called the *server* node. The IBM Virtual Shared Disk subsystem routes I/O requests from the other nodes, called *client* nodes, to the server node and returns the results to the client nodes.

The I/O routing is done by the IBM Virtual Shared Disk device driver layer that sits on top of the Logical Volume Manager (LVM). The device driver is loaded as a kernel extension on each node. Thus, raw logical volumes can be made globally accessible.

The application program interface to an IBM Virtual Shared Disk is the raw device (or device special file). This means application programs must issue requests to IBM Virtual Shared Disks in the block size specified by the AIX Logical Volume Manager (currently requests that are multiples of 512 bytes on 512-byte block boundaries).

You can find more information on logical volumes in *AIX System Management Guide: Operating Systems and Devices*. See “Related Publications” on page xviii for order numbers.

© Copyright IBM Corp. 1997, 1998

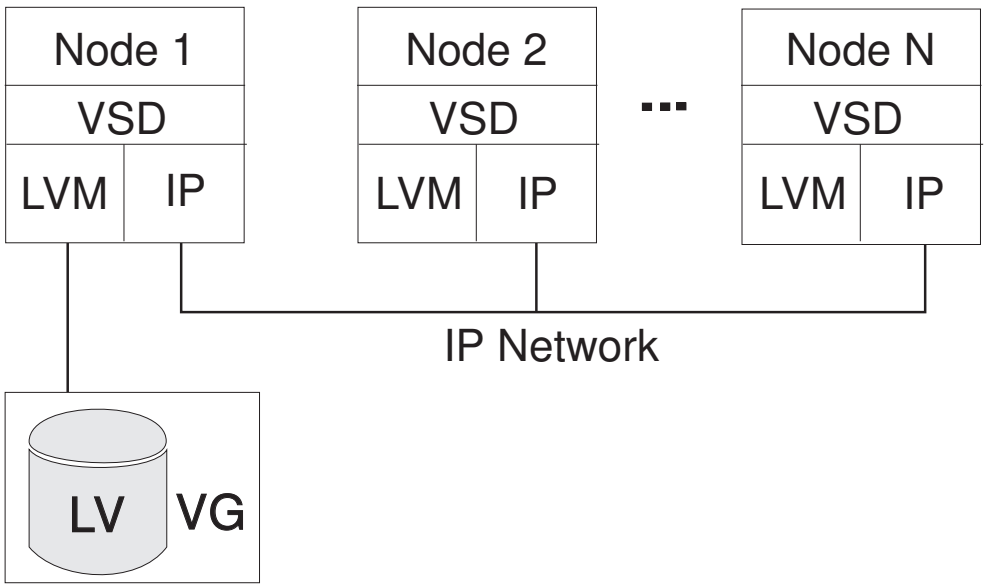


Figure 1. An IBM Virtual Shared Disk IP Network Implementation

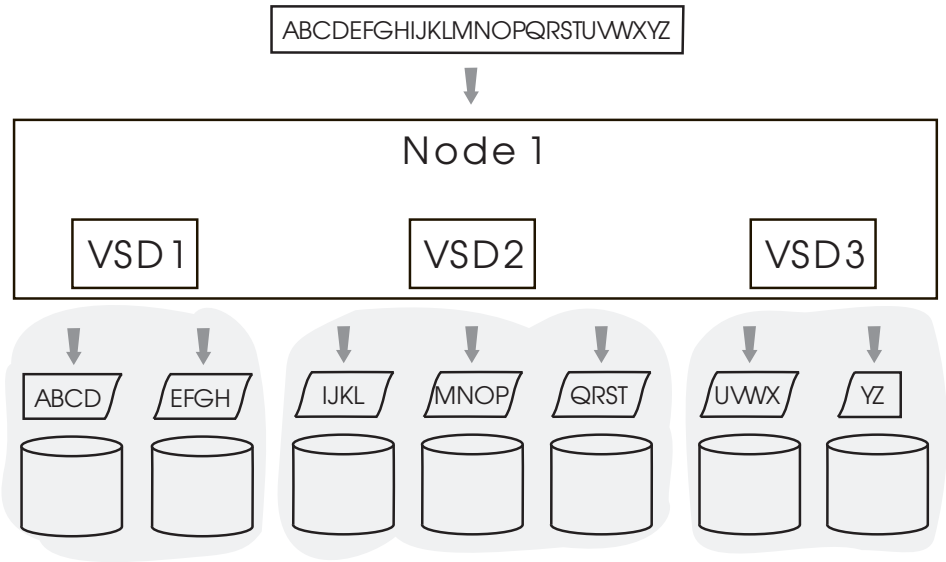
IBM Virtual Shared Disk Restrictions

- In systems with IBM Parallel System Support Programs levels earlier than Version 2.2, all IBM Virtual Shared Disk nodes within a system partition must run at the same software levels; no intermixing is allowed. Later levels of IBM Parallel System Support Programs allow more than one system level in the same system partition. For more information, see 8.
- You should not create AIX Journaled File Systems (JFSs) on volume groups that contain IBM Virtual Shared Disks.
- The maximum number of IBM Virtual Shared Disks that can be defined in a system partition is 32767 (32K).
- The maximum size of an IBM Virtual Shared Disk is 1024 GB (one terabyte).
- You cannot have recoverable and nonrecoverable IBM Virtual Shared Disk nodes in a single system partition.
- Restrictions on AIX logical volumes also affect IBM Virtual Shared Disks; for example:
  - The maximum number of physical volumes in a volume group is 32.
  - The maximum number of physical partitions in a disk is 1016.

Data Striping Device (HSD) Overview

The Data Striping Device (HSD), also known as the Hashed Shared Disk, is a device driver that lets application programs distribute data across multiple IBM Virtual Shared Disks, thus reducing I/O bottlenecks (hot spots).

Instead of writing all the data from one I/O request on one IBM Virtual Shared Disk at a specific location, data striping writes blocks of the data on each of several separate IBM Virtual Shared Disks.



| *Figure 2. HSD Stripes Data Across Two or More VSDs*

The principal benefit of HSD striping is that it provides automatic distribution of data across physical disks and nodes while requiring minimal application specification. When the data is well distributed across the IBM Virtual Shared Disks, the application can get better performance from the underlying physical disks.

Your application can conduct concurrent I/O processing on the different stripes. This is done by the striping software and is transparent to the application.

Comparing HSD and LVM Striping

AIX 4.1 introduced striping capabilities for the Logical Volume Manager. A logical volume can be striped across two or more physical disks. HSD can stripe data across two or more IBM Virtual Shared Disks (an IBM Virtual Shared Disk corresponds to a single logical volume).

The HSD software can stripe data across multiple nodes, while LVM striping is limited to local physical disks. HSD allows you to use mirroring, which LVM striping does not.

If you do not use mirroring, we recommend that you use LVM for local striping and HSD for global striping. For example, if you have two physical disks on 10 nodes and want to stripe the data across all 20 physical disks, you should create a local logical volume that spans both physical disks on each node. After creating IBM Virtual Shared Disks that point to these 10 logical volumes, create one HSD that spans all 10 IBM Virtual Shared Disks. You now have one HSD that spans 10 IBM Virtual Shared Disks (one on each node) and each IBM Virtual Shared Disk points to a logical volume that is striped across two physical disks.

If you do use mirroring, you can use HSD to stripe a local mirrored disk (see the **createhsd** command description in *IBM Parallel System Support Programs for AIX: Command and Technical Reference*.

Refer to *AIX Performance and Tuning Guide*, SC23-2365, for more information on how to tune striped logical volumes.

When to Use the HSD

Whether or not to use the HSD depends on your configuration of IBM Virtual Shared Disks and the I/O characteristics of your application programs.

If the I/O load to a specific IBM Virtual Shared Disk is too heavy, you can use HSD to distribute the load to other IBM Virtual Shared Disks and nodes.

When you plan your application, you should be able to determine whether the bandwidth required for particular IBM Virtual Shared Disks is higher than the system supports. If so, you should consider using HSD. Experience with the IBM Virtual Shared Disk subsystem may also indicate the need for HSD.

Data Striping Device Restrictions

- Striping is supported across IBM Virtual Shared Disks only.
- The function **lseek (SEEK_END)** is not supported because the file may be composed of several IBM Virtual Shared Disks.
- All the underlying IBM Virtual Shared Disks in one HSD must be the same size.
- The size of each IBM Virtual Shared Disk in one HSD must be a multiple of the stripe size in the HSD. For example, if the stripe size is 32KB, the IBM Virtual Shared Disk size must be a multiple of 32KB.
- The size of an HSD (*number_of_vsds* × *size_of_vsd*) cannot be greater than one terabyte (1024 GB).
- After you have written data to an HSD, do not write data to any of the underlying IBM Virtual Shared Disks.
- The address of your I/O buffer must be aligned on a 4KB boundary. If it is not, the HSD read and write routines will **not** parallelize I/O requests to underlying IBM Virtual Shared Disks.

IBM Recoverable Virtual Shared Disk Licensed Program Overview

In order to explain the benefits of the IBM Recoverable Virtual Shared Disk licensed program, let's look at a system *without* the IBM Recoverable Virtual Shared Disk software. The IBM Virtual Shared Disk function lets all nodes in the system partition access a given disk, even though that specific disk is physically attached to only one node. If the server node should fail, access to the disk is lost until the server node is rebooted by the administrator.

By using the IBM Recoverable Virtual Shared Disk software and twin-tailed disks or disk arrays, you can allow a secondary node to take over the server function from the primary node when certain types of failures occur.

A *twin-tailed disk* is a disk or group of disks that are attached to two nodes of an SP. For IBM Recoverable Virtual Shared Disk purposes, only one of these nodes serves the disks at any time. The secondary or backup node provides access to the disks if the primary node fails or is powered off, or if you need to change the server node temporarily for administrative reasons.

A *twin-tailed volume group* is a volume group that contains disks that are accessible to two nodes. Both nodes must be in the same system or system partition.

The IBM Recoverable Virtual Shared Disk software automatically manages your IBM Virtual Shared Disks by detecting conditions such as node failures and EIO errors and switching access to the disk from the primary node to the secondary node, so your application can continue to operate normally. The IBM Recoverable Virtual Shared Disk software also allows you to cut off access to IBM Virtual Shared Disks from certain nodes, using the **fencevsd** command, and to dynamically change the server node with the **vsdchgserver** command.

Users with applications that exploit the IBM Recoverable Virtual Shared Disk software can recover more easily from node failures and have continuous access to the data on the twin-tailed disks.

IBM Virtual Shared Disk Recovery with Twin-tailed Disks

The following three figures show a simple system with one twin-tailed IBM Recoverable Virtual Shared Disk. Figure 3 shows the basic configuration. In Figure 4, the secondary node is active following a failure of the primary node.

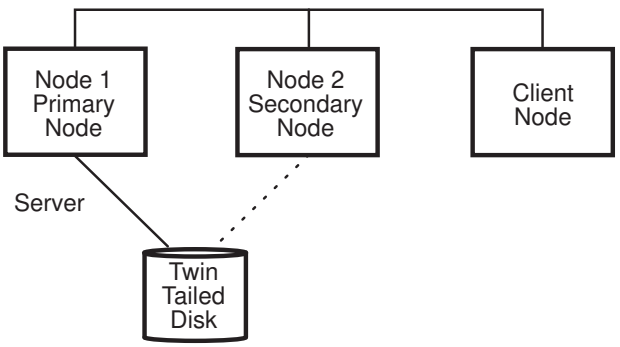


Figure 3. Simplified View of a Twin-tailed Disk

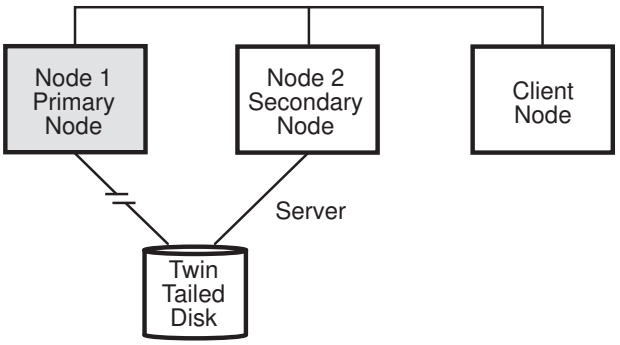


Figure 4. The Secondary Node Serves after a Primary Node Failure

In Figure 5 on page 6, the primary node is active again and has automatically taken over the disk from the secondary node.

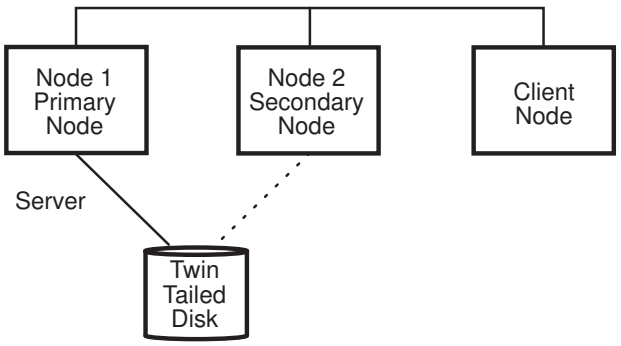


Figure 5. The Primary Node Is the Server Again after Recovery

Recovery is transparent to the application – there is no disruption of service, only a slight delay while takeover occurs. When recovery is complete at the primary server node, the IBM Recoverable Virtual Shared Disk software switches the I/O load back to the server node. The IBM Recoverable Virtual Shared Disk software provides application programming interfaces, including recovery scripts and C programming functions, that enable your application to be recoverable.

IBM Recoverable Virtual Shared Disk Restrictions

Notes:

1. Do not put the root volume group (**rootvg**) or any other volume group that contains bootable logical volumes on a twin-tailed disk.
2. Do not use twin-tailed volume groups with non-IBM Virtual Shared Disk applications. Applications that access data on non-IBM Virtual Shared Disk logical volumes (a file system, for example) will lose access to the logical volume when the IBM Virtual Shared Disk server node changes.
3. You cannot have recoverable and nonrecoverable IBM Virtual Shared Disk nodes in a single system partition.
4. Do not put a tape drive and a twin-tailed disk on the same SCSI bus.

Chapter 2. Installing IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk Licensed Program

This section discusses planning issues you should consider before installation. Migration and installation procedures are also provided.

Planning to Use the Shared Disk Components for the First Time

This section briefly describes some decisions you must make about your applications, and provides some information about the IBM Virtual Shared Disk and HSD components and the IBM Recoverable Virtual Shared Disk licensed program you should know before you install them.

Planning for IBM Virtual Shared Disks

Before you install and use the IBM Virtual Shared Disk software:

1. Refer to the *IBM Parallel System Support Programs for AIX: System Planning Guide* for planning considerations.
2. Check "IBM Virtual Shared Disk Restrictions" on page 2.
3. Determine which applications are to have access to the IBM Virtual Shared Disks you plan to create. See "Tuning IBM Virtual Shared Disk Performance" on page 53 and "Application Programming Considerations for IBM Virtual Shared Disks" on page 63 for considerations in selecting applications to exploit your IBM Virtual Shared Disks and planning your IBM Virtual Shared Disk configuration.
4. Consider the size of the applications you will run.
5. Consider also how you want to spread data across the IBM Virtual Shared Disk nodes.
6. Plan your volume groups and logical volumes, using *AIX System Management Guide: Operating Systems and Devices*.

Planning for HSDs

Before you install and use the HSD component:

1. Check
   "IBM Virtual Shared Disk Restrictions" on page 2
   and "Data Striping Device Restrictions" on page 4.
2. Read "Tuning HSD Performance" on page 60 to understand tuning considerations.
3. Determine where the I/O bottlenecks are on the IBM Virtual Shared Disks.
4. Determine which nodes and disks will be used to eliminate these bottlenecks.
5. Determine the stripe size.
6. Install the HSD software on all nodes with IBM Virtual Shared Disks.

© Copyright IBM Corp. 1997, 1998

Planning for the IBM Recoverable Virtual Shared Disk Licensed Program

If you are using the IBM Recoverable Virtual Shared Disk licensed program, you'll need to consider both primary and secondary nodes for your volume groups and plan your hardware configuration.

1. Plan your twin-tailed volume groups and decide which twin-tailed disks will be in each group.
2. Ensure that the supported disks are properly installed and connected to both nodes.

Note: The following table shows how nodes interoperate and coexist based on the levels of IBM Recoverable Virtual Shared Disk and PSSP.

Table 1. Possibilities for Interaction in the Same System Partition: PSSP and IBM Recoverable Virtual Shared Disk Levels

	PSSP 2.1, IBM Recoverable Virtual Shared Disk 1.1	**PSSP 2.2, IBM Recoverable Virtual Shared Disk 1.2**	**PSSP 2.3, IBM Recoverable Virtual Shared Disk 2.1**	**PSSP 2.4, IBM Recoverable Virtual Shared Disk 2.1**
PSSP 2.1, IBM Recoverable Virtual Shared Disk 1.1	Interoperate	Coexist	Coexist	Coexist
PSSP 2.2, IBM Recoverable Virtual Shared Disk 1.2	Coexist	Interoperate	Interoperate	Interoperate
PSSP 2.3, IBM Recoverable Virtual Shared Disk 2.1	Coexist	Interoperate	Interoperate	Interoperate

Interoperate means that nodes in the same system partition but at different PSSP levels can access each other's IBM Virtual Shared Disks at the lowest functional level. *Coexist* means that nodes at different PSSP levels can reside in the same system or system partition but cannot access each other's IBM Virtual Shared Disks.

Migrating from AIX 3.2.5 to AIX 4.1 and Later Releases

Read this section if you currently are using IBM Virtual Shared Disks and plan to upgrade from AIX Release 3.2.5 to AIX Release 4.1 or later with IBM Parallel System Support Programs for AIX, Version 2 Release 3 or later.

Limits on the Number of Physical Partitions per Disk

The AIX Logical Volume Manager limits the number of physical partitions per disk to 1016. If a disk larger than 4GB is added to a volume group that uses the default physical partition size of 4MB, the disk addition fails.

If you are using disks larger than 4GB in size, define a larger physical partition size. If you already have disks that are over the 1016 limit and intend to move to AIX 4.1

or later AIX releases, you can redefine your volume groups with a larger partition size.

Buddy Buffer Size

If you had large numbers of buddy buffers defined on your AIX 3.2.5 system and your IBM Virtual Shared Disk request size is greater than 60KB when you migrate to AIX 4.1 or later, you should increase your switch pool buffer by the amount of space represented by the current number of buddy buffers, and define two new buddy buffers. See the recommendations in "Buddy Buffers" on page 56 and "mbufs and the Switch Pool" on page 56 for more information.

Migrating from PSSP 2.2 to PSSP 2.3 and Later Releases

Follow these procedures to prepare a node for migration from PSSP 2.2 to PSSP 2.3, and to bring the IBM Virtual Shared Disks back online after PSSP migration is complete. Perform these procedures from the control workstation unless otherwise noted.

Before PSSP Migration

Perform the following steps to prepare for migration.

1. Issue the **vsdatalst** command with the **–n** option and save the output, for example:

   ```
   vsdatalst –n > /tmp/vsdatalst.out
   ```

   The **vsdatalst –n** command returns parameter values similar to the following for each node. You will have to re-enter this data in a later step.

```
        VSD Node Information
                                            Initial Maximum     VSD      rw      Buddy Buffer
  node                  VSD      IP packet   cache   cache request request minimum maximum size: #
number host_name       adapter    size    buffers buffers   count   count    size    size maxbufs
------ --------------- -------- --------- ------- ------- ------- ------- ------- ------- -------
     1 k13n01.ppd.pok. css0         61440      64    4096     256      48    4096  262144      66
```

2. Shut down any applications running on the node that use IBM Virtual Shared Disks.
3. Shut down the IBM Virtual Shared Disk connection manager on the node by issuing:

   ```
   dsh –w node_name /usr/lpp/csd/bin/hc.vsd stop
   ```

   where *node_name* specifies the target hostname.

   Check the connection manager status by issuing:

   ```
   dsh –w node_name /usr/lpp/csd/bin/hc.vsd query
   ```

   Continue if all show inoperative.
4. Shut down the IBM Recoverable Virtual Shared Disk daemon on the node by issuing:

   ```
   dsh –w node_name /usr/lpp/csd/bin/ha.vsd stop
   ```

   where *node_name* specifies the target hostname.

   Check the daemon status by issuing:

```
dsh –w node_name /usr/lpp/csd/bin/ha.vsd query
```

Continue if all show inoperative.

5. Disable the high availability environment on the node to prevent automatic restart of database instances after shut down:

```
dsh –w node_name mv /usr/lpp/csd/bin/hc.activate \
/usr/lpp/csd/bin/hc.activate.off

dsh –w node_name mv /usr/lpp/csd/bin/hc.deactivate \
/usr/lpp/csd/bin/hc.deactivate.off
```

where *node_name* specifies the target node.

6. Disable the node from the IBM Virtual Shared Disk configuration. This step sets the adapter value to `none` and is unnecessary for nodes without IBM Virtual Shared Disks.

```
/usr/lpp/csd/bin/updatevsdnode –n node_number –a none
```

where *node_number* identifies the target node.

7. Verify that the node is removed from the configuration by reissuing the **vsdatalst** command with the **–n** option. A null string should be returned for the *VSD Adapter* value.

Perform PSSP Migration

Install the new PSSP software following the procedures in *IBM Parallel Support for AIX: Installation and Migration Guide*. Then install the filesets listed in “Installation Packages” on page 11 of this document using the procedures that follow.

When PSSP Migration is Complete

Perform these steps to bring your IBM Virtual Shared Disks back online after rebooting the node.

1. If necessary, bring the node back onto the switch by issuing the following command:

```
Eunfence node_number
```

where *node_number* identifies the target node.

2. Update IBM Virtual Shared Disk options on the node using **vsdnode** command. For example, to set the options returned by **vsdatalst** in step 1 of “Before PSSP Migration” on page 9, enter:

```
/usr/lpp/csd/bin/vsdnode 1 css0 64 4096 256 48 4096 262144 61440
```

Note that the positional parameters are entered in a different sequence from the way **vsdatalst** returned them. See Appendix A, “IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands” on page 87, or display the man page for complete syntax.

3. Reactivate high availability environment:

```
dsh –w node_name mv /usr/lpp/csd/bin/hc.activate.off \
/usr/lpp/csd/bin/hc.activate

dsh –w node_name mv /usr/lpp/csd/bin/hc.deactivate.off \
/usr/lpp/csd/bin/hc.deactivate
```

where *node_name* specifies the target hostname.

4. Restart the IBM Virtual Shared Disk on the node by issuing:

   ```
   dsh –w node_name /usr/lpp/csd/bin/ha_vsd reset
   ```

   where *node_name* specifies the target hostname.

Before You Install the Current Releases

Before you install the IBM Virtual Shared Disk and HSD components of IBM Parallel System Support Programs (PSSP) or the IBM Recoverable Virtual Shared Disk licensed program, you need to:

- Install certain components of PSSP
- Identify the software packages you intend to install
- Understand the space requirements for each component
- Understand the hardware requirements for IBM Recoverable Virtual Shared Disk.

PSSP Installation Considerations

Before you can set up IBM Virtual Shared Disk, the SP system software, (IBM Parallel System Support Programs (PSSP), Program Number 5765-529) must have been installed and must be operational, **ssp.basic** and **ssp.sysctl** in particular. Sysctl must also be configured. See *IBM Parallel System Support Programs for AIX: Installation and Migration Guide* for installation information and *IBM Parallel System Support Programs for AIX: Administration* for information about sysctl. PSSP and the IBM Virtual Shared Disk component must be installed before the HSD component or the IBM Recoverable Virtual Shared Disk licensed program can be installed. In particular, PSSP Version 2 Release 3 (or later) and AIX Version 4.2.1 (or later) are prerequisites for IBM Recoverable Virtual Shared Disk Version 2 Release 1.1.

Read the documentation that accompanies the installation media to make sure you have the appropriate version and level of all software. That documentation will list any required APARs (service updates for the PSSP or AIX products) that must be installed along with the IBM Virtual Shared Disk, HSD, or IBM Recoverable Virtual Shared Disk software.

Installation Packages

The installation packages for the shared disk software components are:

- IBM Virtual Shared Disk
  - **ssp.csd.vsd** (on CWS and nodes)
  - **ssp.csd.cmi** (only on the CWS)
- HSD
  - **ssp.csd.hsd**
- IBM Recoverable Virtual Shared Disk
  - **rcsd.vsd**
  - **rcsd.rvsd**
  - **rcsd.hahc**

– **rcsd.docs** (optional)

Note: The **rcsd.hc** installation package no longer exists. **rcsd.docs** now contains only the IBM Recoverable Virtual Shared Disk man pages. The PostScript file for this book is contained in **ssp.docs.**

Detailed Space Requirements

Space Required for the IBM Virtual Shared Disk and HSD Components and the IBM Recoverable Virtual Shared Disk Licensed Program

The individual components of the IBM Virtual Shared Disk and HSD components and the IBM Recoverable Virtual Shared Disk software require the following amounts of space in the specified directories:

Table 2. Space Used by Installation Process

Option Name	root file system	var file system	usr file system
ssp.csd.cml	208KB	0	112KB
ssp.csd.vsd	0	0	374KB
ssp.csd.hsd	0	0	140KB
ssp.csd.sysctl	0	0	360KB
ssp.csd.gui	0	0	3548KB
rcsd.vsd	0	0	110KB
rcsd.rvsd	0	0	231KB
rcsd.hahc	0	0	280KB
rcsd.docs	0	0	36MB

Table 3. Additional Space Used during Execution on the Control Workstation

Option Name	root file system	var file system	usr file system
ssp.csd.cml	depends on how many devices are defined in the SDR	0	0
ssp.csd.vsd	0	0	0
ssp.csd.hsd	0	0	0
ssp.csd.sysctl	0	0	0
ssp.csd.gui	0	0	0
rcsd.vsd	0	0	0
rcsd.rvsd	0	0	0
rcsd.hahc	0	0	0
rcsd.docs	0	0	0

Table 4. Additional Space Used during Execution on the Nodes			
Option Name	**root file system**	**var file system**	**usr file system**
ssp.csd.cml	0	0	0
ssp.csd.vsd	0	depends on how many devices are defined in the system	0
ssp.csd.hsd	0	depends on how many devices are defined in the system	0
ssp.csd.sysctl	0	0	0
ssp.csd.gui	0	0	0
rcsd.vsd	0	0	0
rcsd.rvsd	625KB	0	0
rcsd.hahc	0	0	0
rcsd.docs	0	0	0

The additional space used during execution is for log entries.

/usr/lpp/rcsd is 1980KB in size.

/usr/lpp/ssp.csd is 248KB in size.

Hardware Requirements for the IBM Recoverable Virtual Shared Disk Licensed Program

IBM Recoverable Virtual Shared Disk requires:

- Any twin-tailed disk supported by the Logical Volume Manager running on AIX 4.x.

An Overview of the Installation Process

This section describes the general installation process for the IBM Virtual Shared Disk and HSD components of PSSP and the IBM Recoverable Virtual Shared Disk licensed program.

Step 1: Unconfigure the Existing IBM Virtual Shared Disks

Unconfigure all existing IBM Virtual Shared Disks before installing the new IBM Virtual Shared Disk software. See “Configuring and Unconfiguring IBM Virtual Shared Disks and HSDs” on page 39 and the **ucfgvsd** command reference page for more information.

Step 2: Install PSSP Version 2 Release 4

Install PSSP, program number 5765-529. The **ssp.basic** and **ssp.sysctl** components in particular must be installed and operational and sysctl must be configured. See *IBM Parallel System Support Programs for AIX: Installation and Migration Guide* for installation information

Step 3: Install the Desired Software at the Control Workstation

Install the IBM Virtual Shared Disk software at the control workstation for configuration management. The control workstation is neither a server nor a client.

Use the **ssp.csd installp** package to install both options:

ssp.csd.vsd The IBM Virtual Shared Disk's device driver and commands.

ssp.csd.cmi The IBM Virtual Shared Disk's configuration management support, including the SMIT panels.

You must install both options.

Step 4: Install the Desired Software on all the Nodes in the Configuration

Install the IBM Virtual Shared Disk software on the client and server nodes. The only modules you need at the client and server nodes are in **ssp.csd**.

Installing the Shared Disk Software Components

Follow these steps to install, define, and activate the IBM Virtual Shared Disk software.

If using:	Do this:
SMIT	**TYPE** **smit install_latest** • The Install New Software Products at Latest Level window appears. **ENTER** **/spdata/sys1/install/pssplpp/<code_version>** for Input Device **PRESS** **Do** to display the default install parameters. **PRESS** **List** to show options. **SELECT** One or more program options, or select the header file (called **ssp** with **ALL** on the far right side) to do the full installation. **PRESS** **Do** to complete option selection and to begin installation. When the installation is complete, check the SMIT log file for the installation status. If errors occur, see the *IBM RISC System/6000 Problem Solving Guide*

installp	You can use **installp** to install multiple filesets. For example, to install all of the filesets in PSSP, enter:
	`installp -a -d /spdata/sys1/install/pssplpp/<code_version> -X ssp.*`
	Note - In AIX 4.2, **installp** automatically commits the packaging file set when you specify the **-a** option.
	To list all of the options for **ssp**, enter:
	`installp -l -d /spdata/sys1/install/pssplpp/<code_version>/pssp.installp`

Installation Notes for IBM Recoverable Virtual Shared Disk

All the options in the install package, **rcsd.vsd**, **rcsd.rvsd**, and **rcsd.hahc** must be installed.

The IBM Parallel System Support Programs for AIX (PSSP) is a prerequisite for all the options in the **rcsd installp** package. If you are running PSSP Version 2 Release 4 on AIX 4.2 or later, **ssp.csd.vsd**, **ssp.csd.cmi**, and **ssp.basic** must all be installed.

Note: **ssp.csd.cmi** or **csd.cmi** are only installed on the control workstation.

The **rcsd** package must be installed on all SP nodes that are to be used as either IBM Virtual Shared Disk server nodes or IBM Virtual Shared Disk client nodes. Such nodes must be reset as described in 6 on page 16.

What to Do after Installation

After you install the shared disk components you've decided to use for the first time, what next? New users of each component have slightly different tasks.

New IBM Virtual Shared Disk Users

Follow these general steps:

1. Define your IBM Virtual Shared Disk node information in the System Data Repository (SDR). Chapter 4, " Creating and Managing Shared Disks" on page 25 describes how to do this with the IBM Virtual Shared Disk Perspective. You can also use the **vsdnode** command, described in *IBM Parallel System Support Programs for AIX: Command and Technical Reference*.
2. Create IBM Virtual Shared Disks, using the **createvsd** command.
3. Configure the new IBM Virtual Shared Disks, using the IBM Virtual Shared Disk Perspective or the **cfgvsd** command.
4. Verify the new IBM Virtual Shared Disks, using the IBM Virtual Shared Disk Perspective or the **vsdvts** command.
5. Start the new IBM Virtual Shared Disks, using the IBM Virtual Shared Disk Perspective or the **startvsd** command.

Chapter 4, " Creating and Managing Shared Disks" on page 25 describes how to perform each of these steps in more detail.

New HSD Users

Follow these general steps:

1. Define your IBM Virtual Shared Disk node information in the System Data Repository (SDR). Chapter 4, " Creating and Managing Shared Disks" on page 25 describes how to do this with the IBM Virtual Shared Disk Perspective. You can also use the **vsdnode** command, described in *IBM Parallel System Support Programs for AIX: Command and Technical Reference*.
2. Create HSDs, using the **createhsd** command. The **createhsd** command creates both the new HSDs and their underlying IBM Virtual Shared Disks.

3. Configure the new HSDs, using the IBM Virtual Shared Disk Perspective or the **cfghsdvsd** command.
4. Verify the IBM Virtual Shared Disks that make up the new HSDs, using the IBM Virtual Shared Disk Perspective or the **vsdvts** command.
5. Start the IBM Virtual Shared Disks that make up the new HSDs, using the IBM Virtual Shared Disk Perspective or the **startvsd** command.

Chapter 4, " Creating and Managing Shared Disks" on page 25 describes how to perform each of these steps in more detail.

New IBM Recoverable Virtual Shared Disk Users

Follow these general steps:

1. Define your IBM Virtual Shared Disk node information in the System Data Repository (SDR). Chapter 4, " Creating and Managing Shared Disks" on page 25 describes how to do this with the IBM Virtual Shared Disk Perspective. You can also use the **vsdnode** command, described in *IBM Parallel System Support Programs for AIX: Command and Technical Reference*.
2. Create IBM Virtual Shared Disks, using the **createvsd** command. Identify both primary and secondary nodes for each IBM Virtual Shared Disk. This step automatically creates twin-tailed volume groups and logical volumes.
3. Configure the new IBM Virtual Shared Disks, using the IBM Virtual Shared Disk Perspective or the **cfgvsd** command.
4. Verify the new IBM Virtual Shared Disks, using the IBM Virtual Shared Disk Perspective or the **vsdvts** command.
5. If you have been using the IBM Virtual Shared Disk component without IBM Recoverable Virtual Shared Disk, disable any scripts that issue commands that change the states of IBM Virtual Shared Disks. IBM Recoverable Virtual Shared Disk changes the states of IBM Virtual Shared Disks automatically. Your scripts could disrupt recovery processes.
6. To start IBM Recoverable Virtual Shared Disk and your IBM Virtual Shared Disks, do one of the following:
   - Reboot all the IBM Virtual Shared Disk nodes.

   or

   - Issue the **ha_vsd reset** command on all IBM Virtual Shared Disk nodes. See *IBM Parallel System Support Programs for AIX: Command and Technical Reference* for the complete syntax of the command.

Chapter 4, " Creating and Managing Shared Disks" on page 25 describes in more detail how to perform each of these steps.

Adding IBM Recoverable Virtual Shared Disk to a System Where IBM Virtual Shared Disk Is Already in Use

This section describes how to add IBM Recoverable Virtual Shared Disk to a system where IBM Virtual Shared Disk has already been in use, either with the addition of new disk hardware or without new hardware.

When You Are Setting up New Physical Disks

Follow the steps in this section to migrate IBM Virtual Shared Disk data to a new twin-tailed logical volume.

1. Bring down any application that uses the old IBM Virtual Shared Disks for the duration of this migration.
2. Suspend, stop, and unconfigure all IBM Virtual Shared Disks involved.
3. Add the new physical disks to your hardware configuration, using the instructions that came with them.
4. Use the **createvsd** command to create the new volume groups, logical volumes, and IBM Virtual Shared Disks.
5. Perform test reads and writes to make sure you've successfully created your new IBM Virtual Shared Disks.
6. Copy your data from your existing (old) IBM Virtual Shared Disks to the newly-defined IBM Virtual Shared Disks by copying the data directly from one logical volume to the other, using the new volume group name from the **createvsd** command. For example:

   ```
   cplv [-v new_volume_group]-y new_logical_volume \
   old_logical_volume
   ```

7. After making sure that the data is copied to the new logical volume, use the **removevsd** command to remove the old IBM Virtual Shared Disk.
8. Issue **ha_vsd reset** to make all the IBM Virtual Shared Disks in the configuration active and start the IBM Recoverable Virtual Shared Disk daemons.

Note: If you are using the IBM Virtual Shared Disk in an IBM Virtual Shared Disk striping device (HSD), you should use the **createhsd** command to create the new HSD as well as the underlying IBM Virtual Shared Disks.

When You Are Recabling without Adding New Physical Disks

Follow these steps to add IBM Recoverable Virtual Shared Disk to your system when you are recabling to add the connection to a secondary node but do not need to add new physical disks.

- Stop any applications that use the affected volume groups.
- Suspend, stop, and unconfigure all affected IBM Virtual Shared Disks.
- Add a physical connection from each of the physical disks in the IBM Recoverable Virtual Shared Disk configuration to a secondary node.
- Import the volume groups from the primary node to the secondary, using **smit importvg**.
- If the secondary nodes have not previously been defined as IBM Virtual Shared Disk nodes, enter the SDR information for them, using the notebook pane for the node in the IBM Virtual Shared Disk Perspective or the **vsdnode** command.
- Use the **createvsd** command to redefine the IBM Virtual Shared Disks with the secondary nodes identified.
- Issue **ha_vsd reset** to make all the IBM Virtual Shared Disks in the configuration active and start the IBM Recoverable Virtual Shared Disk daemons.

Chapter 3. Introducing the IBM Virtual Shared Disk Perspective

This chapter introduces you to the IBM Virtual Shared Disk Perspective, a graphical user interface that lets you perform IBM Virtual Shared Disk tasks. If you have the IBM Recoverable Virtual Shared Disk licensed program installed, you can view and change IBM Recoverable Virtual Shared Disk options with the Perspective, too.

Note: Current limitations prevent the IBM Virtual Shared Disk Perspective from starting when there are more than 750 objects (VSD clients and nodes).

To begin working with the IBM Virtual Shared Disk Perspective, use the **perspectives** command to bring up the SP Perspectives launch pad and then double-click on the IBM Virtual Shared Disk Perspective icon. Or type `spvsd` to bring up the IBM Virtual Shared Disk Perspective without starting the launch pad. The full path for this command is **/usr/lpp/ssp/bin/spvsd**.

You can click on **Help→Tasks** at the top right-hand corner of the main panel to see an online help system that explains what you can do with the IBM Virtual Shared Disk Perspective.

The IBM Virtual Shared Disk Perspective Primary Window

Like other windows, Perspectives primary windows have panes. From the top of the primary window, the different areas and panes are:

- The menu bar
- The tool bar
- The pane controls
- Three panes:
  - The Syspars pane
  - The VSD Nodes/Clients pane
  - The VSD/HSD Definitions pane
- The information area

The Menu Bar

This area contains four menu choices: **Window**, **Actions**, **View**, and **Options**.

Figure 6. The IBM Virtual Shared Disk Perspective Menu Bar

- From **Window**, you can close the IBM Virtual Shared Disk Perspective primary window and exit from the Perspective.
- From **Actions**, you can select the actions that apply to any of the three panes (Syspars, IBM Virtual Shared Disk Nodes/Clients, and IBM Virtual Shared Disk/HSD Definitions).

© Copyright IBM Corp. 1997, 1998

- From **View**, you can change how the objects in the IBM Virtual Shared Disk Perspective are displayed; for example, one option lets you select all the objects in a pane.
- From **Options**, you can change how the IBM Virtual Shared Disk Perspective itself is displayed; for example, you can set fonts and colors and change window sizes. The options you set can be saved in a *preference file*, and the perspective is formatted according to your preferences the next time you start it by specifying the preference file.

For more information on the View and Options menu choices, see the Perspectives chapter in *IBM Parallel System Support Programs for AIX: Administration*.

The Tool Bar

The tool bar is displayed right below the menu bar in the primary window. The tool bar consists of icons that represent the most frequently used actions from the Actions and View selections on the menu bar.

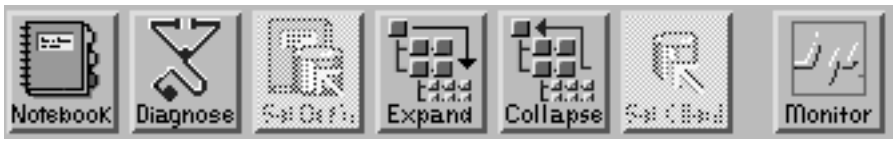


Figure 7. The IBM Virtual Shared Disk Perspective Tool Bar

For the IBM Virtual Shared Disk Perspective, these are:

- The **Notebook** icon, which lets you view and modify attributes of each type of object shown in the panes: syspars (system partitions), IBM Virtual Shared Disk nodes, non-IBM Virtual Shared Disk nodes, IBM Virtual Shared Disk and HSD clients, and IBM Virtual Shared Disk and HSD definitions.
- The **Diagnose** icon, which lets you run IBM Virtual Shared Disk diagnostics for an IBM Virtual Shared Disk node.
- The **Sel. Def'n** icon, which lets you select the IBM Virtual Shared Disk or HSD definition that is associated with a particular IBM Virtual Shared Disk or HSD client or clients.
- The **Expand** icon, which lets you fully expand the tree of objects that are related to the selected object. For example, you can click on this icon to show all the clients of a selected IBM Virtual Shared Disk node or nodes.
- The **Collapse** icon, which lets you fully collapse the tree of objects that are related to the selected object (hide them from the work area).
- The **Sel. Client** icon, which lets you select the client or clients that are associated with a particular IBM Virtual Shared Disk or HSD definition.
- The **Monitor** icon, which lets you monitor conditions associated with IBM Virtual Shared Disk nodes and syspars.

Pane Controls

Just below the tool bar, the pane controls let you add and delete displayed panes. For example, if you want to concentrate on your IBM Virtual Shared Disk nodes, clients, and definitions, you could delete the Syspars pane to make more room for the nodes/clients and definitions panes. You can always add deleted panes again if you need them later.

The Panes

The panes display the objects that you can work with. There are three panes in the IBM Virtual Shared Disk Perspective:

- The **Syspars** pane, where you can set the current system partition, which is indicated by a lightening bolt. The other panes will display the set of nodes, clients, and definitions associated with the current partition. You can set a different current partition by selecting another one and then selecting **Actions→Syspars→Set Current System Partition**.

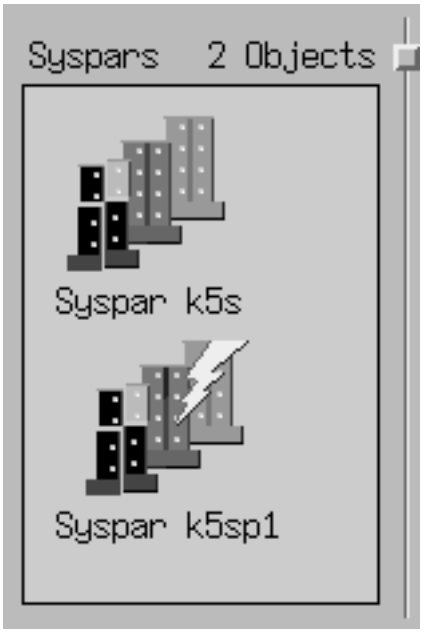


Figure 8. The Syspars Pane

- The **IBM VSD Nodes/Clients** pane, which contains IBM Virtual Shared Disk node icons, non-IBM Virtual Shared Disk node icons, IBM Virtual Shared Disk client icons, and HSD client icons. Non-IBM Virtual Shared Disk node icons represent nodes that have not been initialized for IBM Virtual Shared Disks. The IBM Virtual Shared Disk and HSD client icons are hidden until you do one of the following:
  - click on the arrow to the left of node icons that have clients configured,
  - select Expand from the **View→Expand**, or
  - click on the Expand icon in the tool bar.

  If no clients are configured on a node, no client icons can be displayed. Clients on nodes that are cabled to the physical disks that make up an IBM Virtual Shared Disk (*direct clients*) have a different appearance than clients at remote nodes.

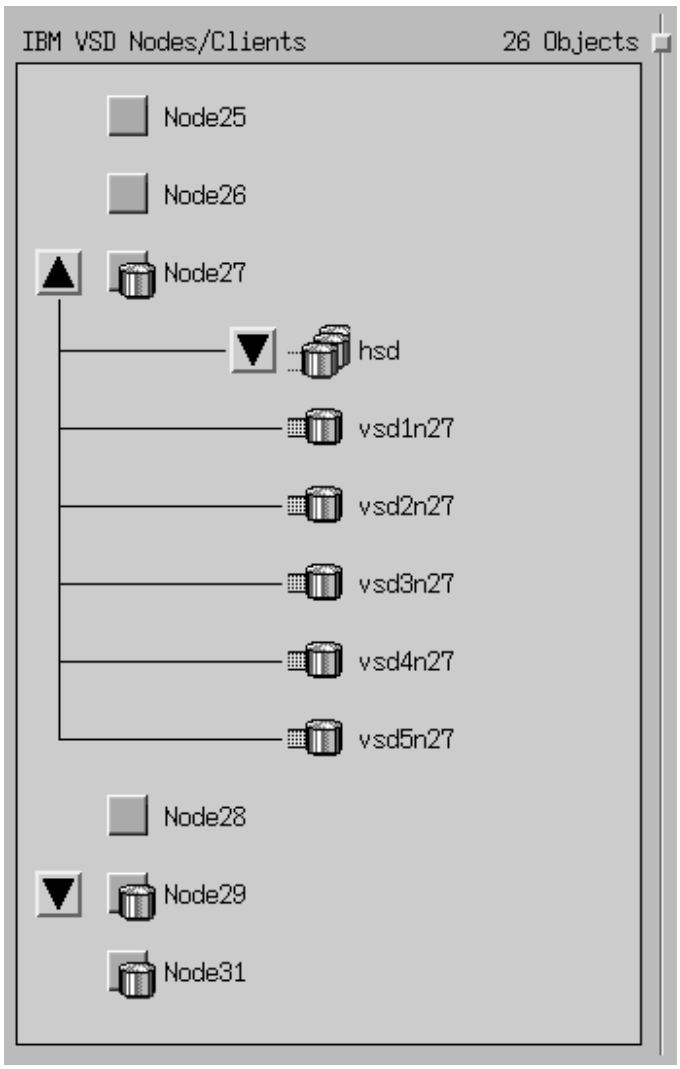


Figure 9. The IBM VSD Nodes/Clients Pane

- The **IBM VSD/HSD Definitions** pane, where you can select IBM Virtual Shared Disk definitions and HSD definitions. HSD definitions are made up of IBM Virtual Shared Disk definitions and can be expanded to show the underlying IBM Virtual Shared Disk definitions.

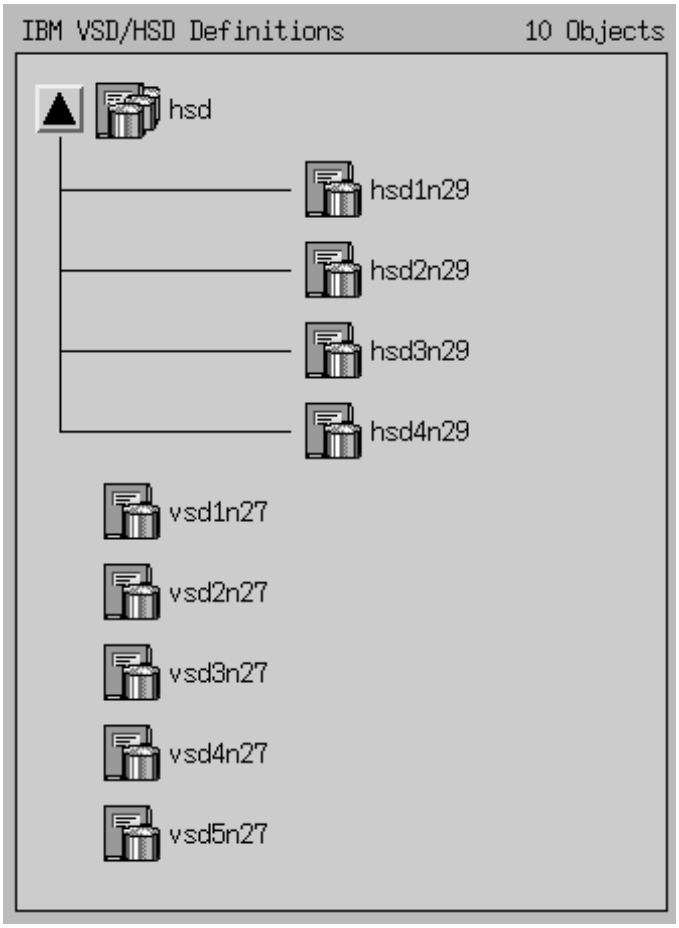


Figure 10. The HSD Definitions Pane

The Information Area

The information area at the bottom of the Perspective gives you a line of help information about each pane and each icon on the tool bar. As you move your cursor over an area or an icon, a line of text that describes it appears in the information area.

Chapter 4. Creating and Managing Shared Disks

This chapter tells you how to:

- Set up and verify new IBM Virtual Shared Disks, using the following steps:
  - Enter information about IBM Virtual Shared Disks in the System Data Repository (SDR), using the IBM Virtual Shared Disk Perspective
  - Create IBM Virtual Shared Disks, using the **createvsd** command
  - Assign backup nodes to IBM Virtual Shared Disks, using the **createvsd** command, if you are using the IBM Recoverable Virtual Shared Disk licensed program
  - Create Data Striping Devices (HSDs), using the **createhsd** command
  - Configure and unconfigure IBM Virtual Shared Disks, using the IBM Virtual Shared Disk Perspective (this action also starts the IBM Virtual Shared Disks and HSDs
  - Verify IBM Virtual Shared Disks, using the IBM Virtual Shared Disk Perspective
  - Start IBM Recoverable Virtual Shared Disks.
- Display information about IBM Virtual Shared Disks and Data Striping Devices, using the IBM Virtual Shared Disk Perspective
- Display and reset IBM Virtual Shared Disk and Data Striping Device node and client statistics and run IBM Virtual Shared Disk diagnostics, using the IBM Virtual Shared Disk Perspective
- Change the states of IBM Virtual Shared Disks, when you do **not** have the IBM Recoverable Virtual Shared Disk licensed program installed, using the IBM Virtual Shared Disk Perspective
- Make changes to IBM Virtual Shared Disks on twin-tailed disks and set timestamps for the volume group
- Remove IBM Virtual Shared Disks and Data Striping Devices, using the **removevsd** and **removehsd** commands
- Add a server node to an existing IBM Virtual Shared Disk configuration
- Remove a node from an existing IBM Virtual Shared Disk configuration

Setting up New IBM Virtual Shared Disks

To set up new IBM Virtual Shared Disks, you need to:

- Make sure your environment is correct by checking your authorization and making sure you have no leftover rollback files
- Enter information about IBM Virtual Shared Disk nodes in the SDR
- Create the IBM Virtual Shared Disks or HSDs
- Configure the IBM Virtual Shared Disks or HSDs
- Verify the IBM Virtual Shared Disks.

© Copyright IBM Corp. 1997, 1998

Preparing to Create IBM Virtual Shared Disks and Data Striping Devices

Before you use the **createvsd** and **createhsd** commands, you must have the correct authorization and you must set up the nodes in your IBM Virtual Shared Disk and Data Striping Device configuration by putting some control information into the System Data Repository (SDR) on each node.

Authorization

In order to issue commands that operate on multiple nodes, such as **createvsd**, you must have Kerberos authority and sysctl authorization. Your system administrator should run **/usr/kerberos/bin/add_principal** to put your ID into the authorized list. The administrator must also add your principal ID to the **/etc/sysctl.acl** and **/etc/sysctl.vsd.acl** files. Then the administrator must run **sysctl svcrestart** to complete the database update for the new entries. The administrator will have to perform these authorization steps on the control workstation and on all nodes that will be IBM Virtual Shared Disk clients or servers.

When you next logon, you may need to run **kinit** to get the correct Kerberos ticket. You can verify your ticket with the **klist** command and run **sysctl whoami** to check your sysctl authorization. To check whether you can run the IBM Virtual Shared Disk and Data Striping Device commands that operate on multiple nodes, use **sysctl sysctl_vsdcheck**. Issue **vsdsklst** as a check before you use any other commands that operate on multiple nodes. If your authorization is correct, it will return information about the IBM Virtual Shared Disks that are defined on the system partition.

Deleting Old Rollback Files

If you issue an IBM Virtual Shared Disk command that operates on multiple nodes, such as **createvsd** or **createhsd**, and the command fails, a rollback file will be created so that a second invocation of that command can start at the last successful operation. (A command that is issued against multiple nodes fails if any of the nodes cannot execute it.) If you later change your IBM Virtual Shared Disk configuration or run a different command, **createvsd** attempts to complete processing using the rollback file and will fail. Be sure there are no rollback files from failed invocations of commands on your system. If there are any, they can be found in **/usr/lpp/csd/vsdfiles/vsd_rollback**. You can delete them with the **rm** command. Old rollback files can interfere with the processing of new invocations of these commands.

Any time a IBM Virtual Shared Disk or HSD command that operates on multiple nodes fails, check for old rollback files.

Displaying and Entering Node Information in the SDR

To enter, display, or change the setup information in the SDR required for each IBM Virtual Shared Disk node:

- Start the IBM Virtual Shared Disk Perspective by issuing **/usr/lpp/spp/bin/spvsd** or by double-clicking on the IBM Virtual Shared Disk icon on the Perspectives initial panel
- Click on a node icon in the IBM Virtual Shared Disk Nodes work area to select it

- Click on the Actions menu item at the top of the IBM Virtual Shared Disk primary window
- Click on IBM Virtual Shared Disk Nodes
- Select the Initialize Node for IBM Virtual Shared Disk action.

The following information must be entered into the SDR before you can create IBM Virtual Shared Disks or Data Striping Devices:

- Node number

  Each node is identified by a *node_number* that represents its position in the SP rack.

- IBM Virtual Shared Disk Adapter Name

  The name of the adapter to be used for IBM Virtual Shared Disk communications with this node.

  The IBM Virtual Shared Disk component uses this name and the related node number to get the IP address associated with each node from the SDR **Adapter** object. Either the High Performance Switch or the Scalable Parallel (SP) Switch (both are defined as **css0**) is recommended as the IBM Virtual Shared Disk adapter for best performance.

  The set of nodes and addresses in the IBM Virtual Shared Disk configuration must define a fully-connected network, that is, all nodes can send messages directly to each other. For the best performance, however, all the nodes should be directly connected to each other over the switch network.

- Initial cache buffer count

  The IBM Virtual Shared Disk device driver implements an optional write-through cache of pinned kernel memory with a block size of 4KB. When the first cached IBM Virtual Shared Disk is configured on a node, the cache is created, containing the specified number of blocks. The minimum value is 1. The recommended value is 256, which results in a 1MB (1024KB) cache.

- Maximum cache buffer count

  The number of buffers in the cache can be increased (using **ctlvsd**) up to this value. You cannot decrease the number of cache blocks. You can start over by unconfiguring all the IBM Virtual Shared Disks, changing the value of this number in the SDR, and then, when you configure the IBM Virtual Shared Disks, the initial number of IBM Virtual Shared Disk buffer cache blocks will be created. The recommended value is 256, resulting in a 1MB cache.

- Number of outstanding logical volume read/write requests

  Specifies the maximum number of outstanding requests the IBM Virtual Shared Disk device driver will allow for each underlying logical volume. Each such request uses a pbuf. A pbuf is a little bigger than a buf structure, which is approximately 104 bytes. The IBM Virtual Shared Disk device driver allocates this number of pbufs in pinned kernel memory whenever an IBM Virtual Shared Disk that has this node as its primary is configured. The minimum value is 1. The recommended value for a primary or secondary node is 48. If the node is not going to be used as a primary or secondary, this value is irrelevant.

- Minimum buddy buffer size

  The smallest buddy buffer a server will use to satisfy a request to an IBM Virtual Shared Disk. This value must be a power of 2 and greater than or equal

to 4096. The recommended value is 4096 (4KB). For a 512-byte request, 4KB is clearly overkill but a buddy buffer is used only for the short period of time a remote request is being processed at the server node.

For more information on buddy buffers, see “Buddy Buffers” on page 56.

- Maximum buddy buffer size

  The largest buddy buffer a server will use to satisfy a request. This value must be a power of 2 and greater than or equal to the *min_buddy_buffer_size*. The maximum value, and the recommended value, is 262144 (256KB). This value must be the same on all nodes within a system partition.

- Number of buddy buffers

  The buddy buffer is pinned kernel memory allocated when the IBM Virtual Shared Disk device driver is loaded. Loading occurs the first time an IBM Virtual Shared Disk is configured. Buddy buffers are freed when the last IBM Virtual Shared Disk is unconfigured. If you do not use the switch as your IP adapter, the size of the buddy buffer affects the number of remote requests the IBM Virtual Shared Disk server node can handle at one time. Remote requests can queue waiting for a buddy buffer. The **statvsd** command and the IBM Virtual Shared Disk Statistics notebook page of the IBM Virtual Shared Disk Perspective report this queuing as buddy buffer shortages. See “Buddy Buffers” on page 56 for more information about buddy buffers and how to determine how many to define. The recommended value when the switch is used as the IP adapter is 2, which results in a 512KB buddy buffer size in combination with the recommended maximum buddy buffer size of 256KB. Tuning may be required on the switch; see “The High Performance Switch/Scalable POWERparallel Switch” on page 55.

- Request block count

  Specifies the maximum number of outstanding IBM Virtual Shared Disk requests originating on the node. The IBM Virtual Shared Disk device driver allocates this number of request blocks in pinned kernel memory the first time an IBM Virtual Shared Disk is configured. If the number is too small, local requests will queue up waiting for a request block to become available. The recommended value is 256. The size of the block is approximately 76 bytes.

- Maximum IP message size

  The maximum IP message size for IBM Virtual Shared Disks, in bytes. The default value is 24,576 (24KB). If you are using the switch as your IBM Virtual Shared Disk adapter, use a value of 61,440 (60KB).

- Level of parallelism

  The IBM Virtual Shared Disk component divides I/O requests into smaller units. This allows one unit to be transmitted, while another is being placed in the buffer. The default value is 9 and it should always be used.

You can display and change the IBM Virtual Shared Disk node attributes in the SDR from this notebook page at any time. You must reconfigure the IBM Virtual Shared Disk definitions on the node before the changes take effect, however.

For more information about these options, see the full description of the **vsdnode** command in *IBM Parallel System Support Programs for AIX: Command and Technical Reference*. The command syntax can be found in Appendix A, “IBM

Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands" on page 87.

Creating IBM Virtual Shared Disks with the createvsd Command

You can create several IBM Virtual Shared Disks with a single command, including secondary disks if you have the IBM Recoverable Virtual Shared Disk licensed program installed. You must first have used the IBM Virtual Shared Disk Perspective or the **vsdnode** command to set up information in the SDR about each node involved in this IBM Virtual Shared Disk configuration.

The syntax of the **createvsd** command is:

createvsd –**n** {*node_list* | **ALL**}
–**s** *size_in_MB* –**g** *volume_group_name*
[{–**c** *number_of_vsds_per_node* | –**L**}]
[–**o cache** | **nocache**] [–**m** *mirror_count*]
[–**p** *lvm_stripe_size*] [–**v** *vsd_name_prefix*]
[–**l** *lv_name_prefix*] [–**T** *lp_size_in_MB*] [–**x**]

Note: Some examples shown in this list do not contain enough flags to be executable. They are shown in an incomplete form to illustrate specific flags.

–**n** Specifies the nodes on which you are creating IBM Virtual Shared Disks. The backup node cannot be the same as the primary node.

node_list Given in the format **P**/**S:***hdisk_list1*+*hdisk_list2*/, where **P** is the primary node, **S**, if specified, is the backup (secondary) node, *hdisk_list1* is the list of local physical disks in the logical volume on the primary, and *hdisk_list1*+*hdisk_list2* is the list of local physical disks in the volume group on the primary, if you want to have more disks in the volume group than are needed for the logical volume. The sequence in which nodes are listed determines the names given to the IBM Virtual Shared Disks. For example:

```
createvsd -n 1,6,4 -v PRE
```

(with the *vsd_prefix* PRE) creates IBM Virtual Shared Disks PRE1n1 on node 1, PRE2n6 on node 6, and PRE3n4 on node 4.

To create a volume group that spans hdisk2, hdisk3, and hdisk4 on node 1, with a backup on node 3, enter:

```
createvsd -n 1/3:hdisk2,hdisk3,hdisk4/ -v DATA
```

This command creates:

- IBM Virtual Shared Disk DATA1n1 with logical volume lvDATA1n1 on a volume group with the global volume group name DATA1n1b3 on node 1, exported to node 3. The Logical Volume Manager (LVM) volume group name is DATA. The logical volumes span hdisk2, hdisk3, and hdisk4.

To create volume groups just like that one on nodes 1, 2, and 3 of a system with backup on nodes 4, 5, and 6 of the same system, enter:

```
createvsd -n 1/4:hdisk1,hdisk2,hdisk3/,2/5:hdisk5,hdisk6,\
hdisk7/,3/6:hdisk2,hdisk4,hdisk6/ -v DATA
```

This command is shown on two lines here, but you must enter it without any spaces between the items in *node_list*.

The command creates:

- IBM Virtual Shared Disk DATA1n1 with logical volume lvDATA1n1 on a volume group with the local volume group name DATA on node 1, exported to node 4. The global volume group name is DATAn1b4.
- IBM Virtual Shared Disk DATA2n2 with logical volume lvDATA2n2 on a volume group with the local volume group name DATA on node 2, exported to node 5. The global volume group name is DATAn2b5.
- IBM Virtual Shared Disk DATA3n3 with logical volume lvDATA3n3 on a volume group with the local volume group name DATA on node 3, exported to node 6. The global volume group name is DATAn3b6.

To create an IBM Virtual Shared Disk where the logical volume spans only two of the physical disks in the volume group, enter:

```
createvsd -n 1/3:hdisk1,hdisk2+hdisk3/ -v DATA
```

This command creates the IBM Virtual Shared Disk DATA1n1 with logical volume lvDATA1n1 spanning hdisk1 and hdisk2 in the volume group DATA, which includes hdisk1, hdisk2, and hdisk3. It exports the volume group DATA to node 3.

If a volume group is already created and the combined physical hdisk lists contain disks that are not needed for the logical volume, those hdisks are added to the volume group. If the volume group has not already been created, **createvsd** creates a volume group that spans *hdisk_list1*+*hdisk_list2*.

Backup nodes cannot use the same physical disk as the primary does to serve IBM Virtual Shared Disks.

ALL Specifies that you are creating IBM Virtual Shared Disks on all nodes in the system or system partition. No backup nodes are assigned if you use this operand. The IBM Virtual Shared Disks will be created on all the physical disks attached to the nodes in *node_list* (you cannot specify which physical disks to use.)

–s Specifies the size in megabytes of each IBM Virtual Shared Disk

–g Specifies the Logical Volume Manager (LVM) volume group name. This name is concatenated with the node number to produce the global volume group name. For example:

```
createvsd -n 6 -g VSDVG
```

creates a volume group with the local volume group name VSDVG and the global volume group name VSDVG1n6 on node 6. The node number is added to the prefix to avoid name conflicts when a backup node takes over a volume group. If a backup node exists, the global volume group name will be concatenated with the backup node number as well as the primary. For example:

```
createvsd -n 6/3/ -g VSDVG
```

creates a volume group with the local volume group name VSDVG and the global volume group name VSDVGn6b3. The primary node is node 6 and the backup node for this volume group is node 3.

–c Specifies the number of IBM Virtual Shared Disks to be created on each node. If *number_of_vsds_per_node* is not specified, one IBM Virtual Shared Disk is created for each node specified on **createvsd**. If more than one IBM Virtual Shared Disk is to be created for each node, the names will be allocated alternately. For example:

```
createvsd -n 1,6 -c 2 -v DATA
```

creates IBM Virtual Shared Disks DATA1n1 on node 1, DATA2n6 on node 6, DATA3n1 on node 1, and DATA4n6 on node 6.

–L Allows you to create one IBM Virtual Shared Disk on each node without using sequential numbers, for locally-accessed IBM Virtual Shared Disks.

–A Specifies that IBM Virtual Shared Disk names will be allocated to each node in turn, for example:

```
createvsd -n 1,6 -c 2 -A DATA
```

creates DATA1n1 and DATA2n1 on node 1, and DATA3n6 and DATA4n6 on node 6.

–o Specifies either the **cache** or the **nocache** option. The default is **nocache**.

–m Specifies the LVM mirroring count. The mirroring count sets the number of physical partitions allocated to each logical partition. The range is from 1 to 3 and the default value is 1.

–p Specifies the LVM stripe size. If this flag is not specified, the logical volumes are not striped. In order to use striping, the node on which the IBM Virtual Shared Disks are defined must have more than one physical disk.

–v Specifies a prefix to be given to the names of the created IBM Virtual Shared Disks. This prefix will be concatenated with the IBM Virtual Shared Disk number, node number, and backup node number, if a backup disk is specified. For example, if the prefix PRE is given to an IBM Virtual Shared Disk created on node 1 and there are already two IBM Virtual Shared Disks with this prefix across the partition, the new IBM Virtual Shared Disk name will be PRE3n1. The name given to the underlying logical volume will be lvPRE3n1, unless the **–l** flag is used. The **createvsd** command continues to sequence IBM Virtual Shared Disk names from the last PRE-prefixed IBM Virtual Shared Disk

If **-v** is not specified, the prefix `vsd` is used.

Note: The last character of the *vsd_name_prefix* cannot be a digit. Otherwise, the 11th IBM Virtual Shared Disk with the prefix PRE would have the same name as the first IBM Virtual Shared Disk with the prefix PRE1. Nor can the *vsd_name_prefix* contain the character '**.**' because '**.**' can be any character in regular expressions.

–**l** Overrides the prefix *lvx* that is given by default to a logical volume by the **createvsd** command, where *x* is the IBM Virtual Shared Disk name prefix specified by *vsd_name_prefix* or the default (`vsd`). For example:

```
createvsd -n 1 -v DATA
```

creates one IBM Virtual Shared Disk on node 1 named DATA1n1 with an underlying logical volume lvDATA1n1. If the command

```
createvsd -n 1 -v DATA -l new
```

is used, the IBM Virtual Shared Disk on node 1 is still named DATA1n1, but the underlying logical volume is named lvnew1n1.

It is usually more helpful **not** to specify –**l**, so that your lists of IBM Virtual Shared Disk names and logical volume names are easy to associate with each other and you avoid naming conflicts.

–**T** Specifies the size of the physical partition in the Logical Volume Manager logical volume group and also the logical partition size (they will be the same) in megabytes. You must select a power of 2 in the range 2–256. The default is 4MB.

The Logical Volume Manager limits the number of physical partitions to 1016 per disk. If a disk is greater than 4 gigabytes in size, the physical partition size must be greater than 4MB to keep the number of partitions under the limit.

–**x** Specifies that the steps required to synchronize the IBM Virtual Shared Disks on the primary and secondary nodes should **NOT** be performed; that is, the sequence:

- **varyoffvg** on the primary node
- **exportvg** on the primary node
- **importvg** on the secondary node
- **varyoffvg** on the secondary node
- **varyonvg** on the primary nodes

is not done as part of the **createvsd** processing. This speeds the operation of the command and avoids unnecessary processing in the case where several IBM Virtual Shared Disks are being created on the same primary/secondary nodes. In this case, however, you should either specify –**x** on the last **createvsd** in the sequence or issue the volume group commands listed above explicitly.

Information about IBM Virtual Shared Disk definitions is stored in the System Data Repository (SDR) **VSD_Global_Volume_Group** object and the **VSD_Table** object.

Examples of Creating IBM Virtual Shared Disks with createvsd

The following examples range from simple to fairly complex.

- To create identical IBM Virtual Shared Disk definitions on each of three nodes in a system partition, type:

  ```
  createvsd –n 1,3,5 -s 4 -g SYS2VG -v SYS2VSD
  ```

  This creates the following IBM Virtual Shared Disk definitions:

  – SYS2VSD1n1 on node 1. The local volume group name on node 1 is SYS2VG. The global volume group name is SYS2VGn1. The logical volume is lvSYS2VSD1n1.

- SYS2VSD2n3 on node 3. The local volume group name on node 3 is SYS2VG. The global volume group name is SYS2VGn3. The logical volume is lvSYS2VSD2n3.
- SYS2VSD3n5 on node 5. The local volume group name on node 5 is SYS2VG. The global volume group name is SYS2VGn5. The logical volume is lvSYS2VSD3n5.

No secondary disks are defined. The space allocated to the IBM Virtual Shared Disk is spread across all the physical disks (hdisks) connected to each node (1,3, and 5).

- To assign each disk in the previous example a secondary node, (with the IBM Recoverable Virtual Shared Disk licensed product installed) type:

  ```
  createvsd –n 1/2/,3/4/,5/6/ -s 4 -g SYS2VG -v SYS2VSD
  ```

  This creates the following IBM Virtual Shared Disk definitions:

  - SYS2VSD1n1 on node 1 with a twin-tailed connection to node 2. The local volume group name on node 1 is SYS2VG. The global volume group name is SYS2VGn1. The logical volume is lvSYS2VSD1n1.
  - SYS2VSD2n3 on node 3 with a twin-tailed connection to node 4. The local volume group name on node 3 is SYS2VG. The global volume group name is SYS2VGn3. The logical volume is lvSYS2VSD2n3.
  - SYS2VSD3n5 on node 5 with a twin-tailed connection to node 6. The local volume group name on node 5 is SYS2VG. The global volume group name is SYS2VGn5. The logical volume is lvSYS2VSD3n5.

  The volume groups in this example are imported to the secondary node.

- To create three IBM Virtual Shared Disk definitions on primary and secondary nodes (with the IBM Recoverable Virtual Shared Disk licensed product installed), where the logical volume created on nodes 3 and 4 spans two disks and the volume group spans three disks, type:

  ```
  createvsd -n 3/4:hdisk1,hdisk2+hdisk3/,5/6/,7/8/ -s 12 -g datavg -v USER -x
  ```

  This command creates the following IBM Virtual Shared Disk definitions:

  - USER1n3, with logical volume lvUSER1n3 defined on a volume group with the global volume group name datavgn3 on node 3, imported to node 4. The volume group datavgn3 spans hdisk1, hdisk2, and hdisk3. The logical volume lvUSER2n3 spans hdisk1 and hdisk2.
  - USER2n5, with logical volume lvUSER2n5 defined on a volume group with the global volume group name datavgn5 on node 5, imported to node 6.
  - USER3n7, with logical volume lvUSER3n7 defined on a volume group with the global volume group name datavgn7 on node 7, also imported to node 8.

    The volume groups datavgn5 and datavgn7 are created with one 4MB partition from a single physical disk. The volume group datavgn3 is created with one 4MB partition from the three physical disks hdisk1, hdisk2, and hdisk3.

  The local volume group name on each node is datavg. The volume groups are **not** imported to the secondary node.

Creating HSDs with the createhsd Command

You can create a whole configuration of HSDs with a single command, including secondary disks for the underlying IBM Virtual Shared Disks if you have the IBM Recoverable Virtual Shared Disk licensed program installed. The syntax of the **createhsd** command is:

createhsd – Creates one data striping device (HSD) that encompasses two or more IBM Virtual Shared Disks.

createhsd –**n** {*node_list* | **ALL**} –**s** *size_in_MB* –**g** *vg_name*
–**t** *stripe_size_in_KB* [{–**c** *vsds_per_node* | –**L**} [–**A**]]
[–**m** *mirror_cnt*] [–**d** *hsd_name*] [–**l** *lv_name_prefix*] [–**S**]
[–**o cache** | **nocache**] [–**T** *lp_size_in_MB*] [–**x**]

Note: Some examples shown in this list do not contain enough flags to be executable. They are shown in an incomplete form to illustrate specific flags.

–**n** Specifies the numbers of the nodes on which you are creating a data striping device (HSD). The backup node for the underlying IBM Virtual Shared Disks cannot be the same as the primary node.

node_list Given in the format **P**/**S:***hdisk_list1*+*hdisk_list2*/, where **P** is the primary node, **S**, if specified, is the backup (secondary) node, *hdisk_list1* is the list of local physical disks in the logical volume on the primary, and *hdisk_list1*+*hdisk_list2* is the list of local physical disks in the volume group on the primary, if you want to have more disks in the volume group than are needed for the logical volume.

The sequence in which nodes are listed determines the names given to the IBM Virtual Shared Disks; for example:

```
createhsd -n 1,6,4 -d DATA
```

(with the *hsd_prefix* DATA) creates IBM Virtual Shared Disks DATA1n1 on node 1, DATA2n6 on node 6, and DATA3n4 on node 4, which make up a single HSD named DATA. To create volume groups that span specified disks on nodes 1, 2, and 3 of a system with backup on nodes 4, 5, and 6 of the same system, and that make up a single HSD, enter:

```
createhsd -n 1/4:hdisk1,hdisk2,hdisk3/,2/5:hdisk5,hdisk6,\
hdisk7/,3/6:hdisk2,hdisk4,hdisk6/ -d DATA -s 12 -g OLD\
-t 4096
```

This command is shown on two lines here, but you must enter it without any spaces between the items in *node_list*.

The command creates:

- IBM Virtual Shared Disk DATA1n1 with logical volume lvDATA1n1 on a volume group with the local volume group name OLD on node 1, exported to node 4. The global volume group name is OLDn1b4.
- IBM Virtual Shared Disk DATA2n2 with logical volume lvDATA2n2 on a volume group with the local volume group name OLD on node 2, exported to node 5. The global volume group name is OLDn2b5.

- IBM Virtual Shared Disk DATA3n3 with logical volume lvDATA3n3 on a volume group with the local volume group name OLD on node 3, exported to node 6. The global volume group name is OLDn3b6.
- An HSD named DATA that comprises the three IBM Virtual Shared Disks listed previously. The HSD is 12MB in total usable size and has a 512KB stripe size.

If a volume group is already created and the combined physical hdisk lists contain disks that are not needed for the logical volume, those hdisks are added to the volume group. If the volume group has not already been created, **createhsd** creates a volume group that spans *hdisk_list1*+*hdisk_list2*.

Backup nodes cannot use the same physical disk as the primary does to serve IBM Virtual Shared Disks.

ALL Specifies that you are creating HSDs on all nodes in the system or system partition. If you use **ALL**, you can't assign backup nodes for the disks.

–s Specifies the total usable size of the HSD in MB. Unless **–S** is specified, **createhsd** adds at least a stripe size to each IBM Virtual Shared Disk's size for each HSD.

–g Specifies the Logical Volume Manager (LVM) volume group name, or local volume group name. This name is concatenated with the node number to form the global volume group name (VSD_GVG). For example:

```
createhsd -n 6 -g VSDVG
```

creates a volume group with the IBM Virtual Shared Disk global volume group name VSDVGn6 and the local AIX volume group name VSDVG. The node number is added to the prefix for the global volume group name to avoid name conflicts with the name used for this volume group on other nodes. If a backup node exists, the global volume group name will be concatenated with the backup node number as well as the primary. For example:

```
createhsd -n 6/3/ -g VSDVG
```

creates VSDVGn6b3, where the primary node is node 6 and the backup node for this volume group is node 3.

–c Specifies the number of IBM Virtual Shared Disks to be created on each node. If *number_of_vsds_per_node* is not specified, one IBM Virtual Shared Disk is created for each node specified on **createvsd**. If more than one IBM Virtual Shared Disk is to be created for each node, the names will be allocated cyclically. For example:

```
createhsd -n 1,6 -c 2 -d DATA
```

creates IBM Virtual Shared Disks DATA1n1 on node 1, DATA2n6 on node 6, DATA3n1 on node 1, and DATA4n6 on node 6 and uses them to make up the HSD DATA.

–L Allows you to create one IBM Virtual Shared Disk on each node without using sequential numbers for locally-accessed IBM Virtual Shared Disks.

–A Specifies that IBM Virtual Shared Disk names will be allocated to each node in turn. For example:

```
createhsd -n 1,6 -c 2 -A DATA
```

creates DATA1n1 and DATA2n1 on node 1, and DATA3n6 and DATA4n6 on node 6.

–o Specifies either the **cache** or **nocache** option for the underlying IBM Virtual Shared Disks. The default is **nocache**.

–m Specifies the LVM mirroring count. The mirroring count sets the number of physical partitions allocated to each logical partition. The range is from 1 to 3. If **–m** is not specified, the count is set to 1.

–t Specifies the stripe size in kilobytes that an HSD will use. The stripe size must be a multiple of 4KB and less than or equal to 1GB.

–d Specifies the name assigned to the created HSD. It is used as the IBM Virtual Shared Disk prefix name (the **–v** in **createvsd**). If an HSD name is not specified, a default name, *x*HsD is used, where x denotes a sequence number.

The command:

```
createhsd -n 1,2 -d DATA
```

creates two IBM Virtual Shared Disks, DATA1n1 and DATA2n2. These IBM Virtual Shared Disks make up one HSD named DATA.

–l Overrides the prefix *lvx* that is given by default to a logical volume by the **createvsd** command, where *x* is the IBM Virtual Shared Disk name prefix specified by *vsd_name_prefix* or the default (`vsd`). For example:

```
createhsd -n 1 -v DATA
```

creates one IBM Virtual Shared Disk on node 1 named DATA1n1 with an underlying logical volume lvDATA1n1. If the command

```
createhsd -n 1 -v DATA -l new
```

is used, the IBM Virtual Shared Disk on node 1 is still named DATA1n1, but the underlying logical volume is named lvnew1n1.

It is usually more helpful **not** to specify **–l**, so that your lists of IBM Virtual Shared Disk names and logical volume names are easy to associate with each other and you avoid naming conflicts.

–S Specifies that the HSD overrides the default skip option and does not skip the first stripe to protect the first LVM Control Block (LVCB).

–T Specifies the size of the physical partition in the Logical Volume Manager logical volume group and also the logical partition size (they will be the same) in megabytes. You must select a power of 2 in the range 2–256. The default is 4MB.

The Logical Volume Manager limits the number of physical partitions to 1016 per disk. If a disk is greater than 4 gigabytes in size, the physical partition size must be greater than 4MB to keep the number of partitions under the limit.

–**x** Specifies that the steps required to synchronize the IBM Virtual Shared Disks on the primary and secondary nodes should **NOT** be performed; that is, the sequence:

- **varyoffvg** on the primary node
- **exportvg** on the primary node
- **importvg** on the secondary node
- **varyoffvg** on the secondary node
- **varyonvg** on the primary nodes

is not done as part of the **createvsd** processing. This speeds the operation of the command and avoids unnecessary processing in the case where several IBM Virtual Shared Disks are being created on the same primary/secondary nodes. In this case, however, you should either specify –**x** on the last **createvsd** in the sequence or issue the volume group commands listed above explicitly.

Examples of Creating Data Striping Devices with createhsd

The following examples range from simple to fairly complex.

- To create an HSD that stripes data across three identical IBM Virtual Shared Disk definitions on each of three disks in a system partition, type:

  ```
  createhsd –n 1,3,5 -s 12 -g SYS2VG -t 4 -d SYS2HSD
  ```

  This creates the Data Striping Device definition SYS2HSD and its underlying IBM Virtual Shared Disk definitions:

  – SYS2HSD1n1 on node 1. The local volume group name on node 1 is SYS2VG. The global volume group name is SYS2VGn1. The logical volume is lvSYS2HSD1n1.

  – SYS2HSD2n3 on node 3. The local volume group name on node 3 is SYS2VG. The global volume group name is SYS2VGn3. The logical volume is lvSYS2HSD2n3.

  – SYS2HSD3n5 on node 5. The local volume group name on node 5 is SYS2VG. The global volume group name is SYS2VGn5. The logical volume is lvSYS2HSD3n5.

    The usable Data Striping Device size is 12MB. The stripe size is 4KB. The first stripe on each disk is skipped to allow space for the LVCB.

  No secondary disks are defined. The space allocated to the Data Striping Device is spread across all the physical disks (hdisks) connected to each node (1,3, and 5).

- With the IBM Recoverable Virtual Shared Disk licensed product installed, to create an HSD definition whose underlying IBM Virtual Shared Disks are twin-tailed and have secondary direct client nodes assigned, type:

  ```
  createhsd –n 1/2/,3/4/,5/6/ -s 48 -g SYS2VG -t 8 -d SYS2HSD
  ```

  This adds the backup node number to the naming convention and creates Data Striping Device SYS2HSD with the following IBM Virtual Shared Disk definitions:

  – SYS2HSD1n1 on node 1 with a twin-tailed connection to node 2. The local volume group name on node 1 is SYS2VG. The global volume group name is SYS2VGn1. The logical volume is lvSYS2HSD1n1.

– SYS2HSD2n3 on node 3 with a twin-tailed connection to a second direct client on node 4. The local volume group name on node 3 is SYS2VG. The global volume group name is SYS2VGn3. The logical volume is lvSYS2HSD2n3.

– SYS2HSD3n5 on node 5 with a twin-tailed connection to a second direct client on node 6. The local volume group name on node 5 is SYS2VG. The global volume group name is SYS2VGn5. The logical volume is lvSYS2HSD3n5.

The usable Data Striping Device size is 48MB. The stripe size is 8KB. The first stripe on each disk is skipped to allow space for the LVCB.

- With the IBM Recoverable Virtual Shared Disk licensed product installed, to create an Data Striping Device with three underlying IBM Virtual Shared Disk definitions on primary and secondary nodes, where the logical volumes and volume groups span two physical disks, type:

```
createhsd -n 3/4:hdisk1,hdisk2/,5/6:hdisk1,hdisk2/,7/8:hdisk1,hdisk2/
          -s 48 -t 12 -S -g datavg -d USER
```

This command creates the Data Striping Device USER, with the following underlying IBM Virtual Shared Disk definitions:

– USER1n3, with logical volume lvUSER1n3 defined on a volume group with the global volume group name datavgn3 on node 3, imported to node 4. The volume group datavgn3 and the logical volume lvUSER2n3 span hdisk1 and hdisk2.
– USER2n5, with logical volume lvUSER2n5 defined on a volume group with the global volume group name datavgn5 on node 5, imported to node 6. The volume group datavgn5 and the logical volume lvUSER2n5 span hdisk1 and hdisk2.
– USER3n7, with logical volume lvUSER3n7 defined on a volume group with the global volume group name datavgn7 on node 7, imported to node 8. The volume group datavgn7 and the logical volume lvUSER2n7 span hdisk1 and hdisk2.

The usable size of the Data Striping Deviceis 48MB. The stripe size is 12KB. The first stripe on each disk is **not** skipped to allow space for the LVCB.

Defining Data Striping Devices in a Node-Pair Configuration

If your system partition is configured in a node-pair arrangement like that shown in Figure 11 on page 39, where sets of physical disks are twin-tailed on nodes 1 and 2, nodes 3 and 4, nodes 5 and 6, and so forth, you will need to use the following process to create a single HSD for the system partition. In this configuration, odd-numbered nodes are the backups for IBM Virtual Shared Disks defined on even-numbered nodes and even-numbered nodes are the backups for IBM Virtual Shared Disks defined on odd-numbered nodes.

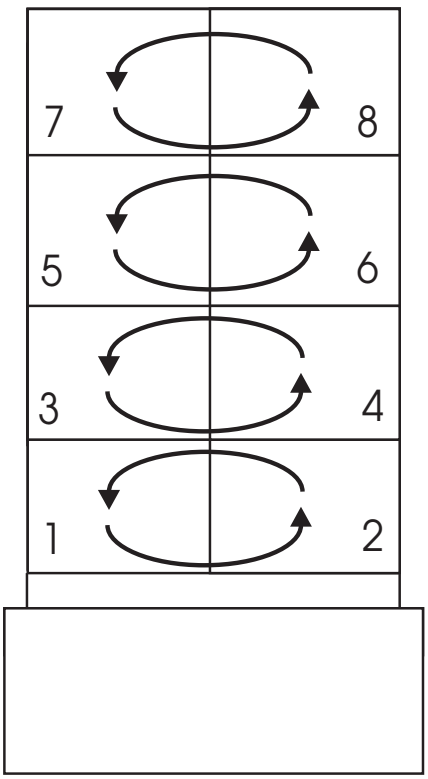


Figure 11. A Node-Pair Configuration

To set up the HSD for this configuration, you use **createvsd** to define one or more IBM Virtual Shared Disks on the odd-numbered nodes, and then to define one or more IBM Virtual Shared Disks on the even-numbered nodes. IBM Virtual Shared Disks on odd-numbered nodes must be defined separately from those on even-numbered nodes in this configuration if they are to be part of a single HSD. You then use **defhsd** to define the HSD.

For example, to define an HSD for the configuration in Figure 11, you could use the following sequence of commands:

```
createvsd 1/2/,3/4/,5/6/,7/8/ –s 8192 –g s21ovg –v ovsd

createvsd 2/1/,4/3/,6/5/,8/7/ –s 8192 –g s21evg –v evsd

defhsd protect_lvcb s21hsd 8192 ovsdn1 ovsdn3 ovsdn5 ovsdn7
evsdn2 evsdn4 evsdn6 evsdn8
```

Configuring and Unconfiguring IBM Virtual Shared Disks and HSDs

After you've created your IBM Virtual Shared Disks or HSDs, you must configure them so that all the clients can read from and write to them. To configure or unconfigure all the IBM Virtual Shared Disks or HSDs on a node:

- Start the IBM Virtual Shared Disk Perspective by issuing **/usr/lpp/spp/bin/spvsd** or by double-clicking on the IBM Virtual Shared Disk icon on the Perspectives initial panel
- Select the node in the IBM Virtual Shared Disk Nodes work area
- Click on Actions in the menu bar
- Click on IBM Virtual Shared Disk Nodes

- From the Actions list, click on the appropriate configure/unconfigure action.

To unconfigure and reconfigure the IBM Virtual Shared Disks or HSDs in a single step, click on Reconfigure all IBM Virtual Shared Disks or Reconfigure all HSDs.

You can also unconfigure or reconfigure an IBM Virtual Shared Disk or HSD client. To do so, select the client icon in the IBM Virtual Shared Disk Nodes work area and then follow the same steps as in the previous paragraph. The highlighted options in the Actions list will be Unconfigure and Reconfigure.

You can also use the **cfgvsd**, **cfghsdvsd**, **ucfgvsd**, or **ucfghsdvsd** commands or SMIT.

Verifying an IBM Virtual Shared Disk

CAUTION:
The verification procedure writes data to the IBM Virtual Shared Disks. Do not select this action after you have put real data on an IBM Virtual Shared Disk.

To test that you have successfully installed the IBM Virtual Shared Disk software and that you can successfully define, activate, read from, and write to an IBM Virtual Shared Disk, follow these steps after creating your IBM Virtual Shared Disk definition with **createvsd**:

1. Bring up the IBM Virtual Shared Disk Perspective by typing **spvsd** or by selecting the IBM Virtual Shared Disk icon from the IBM Virtual Shared Disk Perspectives main panel.
2. Look for the icon that represents the primary node for your new IBM Virtual Shared Disk definition in the IBM Virtual Shared Disk Nodes work area. Click on the icon to select it.
3. Configure the IBM Virtual Shared Disk definition on your system by clicking on Actions in the menu bar and then double-clicking on Configure All IBM Virtual Shared Disks.
4. Change the state of all clients on this IBM Virtual Shared Disk to Active by selecting Actions again and choosing Change IBM Virtual Shared Disk States. Use the Change States dialog box to select Start. You'll see a warning message telling you not to use this command if you have the IBM Recoverable Virtual Shared Disk licensed program installed. This is because IBM Virtual Shared Disks are usually activated through Group Services when the IBM Recoverable Virtual Shared Disk product is in the system. You may use this Start action right after you create your IBM Virtual Shared Disks, however, in order to verify the configuration, even if you have the IBM Recoverable Virtual Shared Disk product installed.
5. Display configuration information about the IBM Virtual Shared Disk by clicking on the notebook icon in the tool bar. Select the notebook page labelled IBM Virtual Shared Disk Node Attributes. The display should show that your IBM Virtual Shared Disk is in the active state.
6. Confirm that your IBM Virtual Shared Disks are working properly by exiting from the notebook page and double-clicking on Run Verification Tests under Actions. You'll see a couple of warning messages here, too, since the verification tests write data to the IBM Virtual Shared Disks and would overwrite any data you had on them. The operation should complete quickly, and you should get the message:

```
VSD verification test successful!
```

You will need to use this procedure from each IBM Virtual Shared Disk client node.

You can also use the **vsdvts** command to verify all of your IBM Virtual Shared Disk clients.

Starting IBM Virtual Shared Disks and HSDs

If you have the IBM Recoverable Virtual Shared Disk licensed program installed, issue the **ha_vsd reset** command to start all your IBM Virtual Shared Disks and HSDs.

If you do not have the IBM Recoverable Virtual Shared Disk licensed program installed, use the IBM Virtual Shared Disk Perspective to start your IBM Virtual Shared Disks and HSDs. See “Changing the States of IBM Virtual Shared Disks” on page 45.

Displaying Information about IBM Virtual Shared Disk and HSD Nodes, Clients, and Definitions

You can use the notebook pages in the IBM Virtual Shared Disk Perspective to display a variety of information about your IBM Virtual Shared Disks.

Displaying Information about IBM Virtual Shared Disk and HSD Nodes

To display information about IBM Virtual Shared Disk and HSD nodes, select the node in the IBM Virtual Shared Disk Nodes work area. Click on the notebook icon in the tool bar. The IBM Virtual Shared Disk Node Attributes notebook page reflects the options set in the SDR for this node when it was defined (see “Displaying and Entering Node Information in the SDR” on page 26).

The HSD Node Attributes notebook page shows you the level of parallelism for this HSD.

Displaying Information about IBM Virtual Shared Disk Clients

To view client attributes for IBM Virtual Shared Disks, select the client and click on the notebook icon in the tool bar. The Client Attributes page shows the following information:

- Name of the IBM Virtual Shared Disk
- Node number on which this client is configured
- Primary node for this IBM Virtual Shared Disk
- Secondary node for this IBM Virtual Shared Disk, if one is defined
- Physical disks spanned by this IBM Virtual Shared Disk
- Local volume group name
- Global volume group name
- Minor number of this IBM Virtual Shared Disk (the device type)
- Logical volume name
- Logical volume major number (the device ID)
- Logical volume minor number (a unique number that helps identify the device)
- Cache option setting (*cache* or *nocache*)
- IBM Virtual Shared Disk Client reads (count of 512–byte blocks)
- IBM Virtual Shared Disk Client writes (count of 512–byte blocks)

- State of the IBM Virtual Shared Disk
- Owner of this IBM Virtual Shared Disk
- Group to which the owner of this IBM Virtual Shared Disk belongs

See *AIX System Management Guide: Operating Systems and Devices* for more information about some of these values.

The Client Values page of this notebook shows you the values of the following information for this client:

- The adapter on which this IBM Virtual Shared Disk is defined
- Initial cache buffer count for this client
- Maximum cache buffer count
- Outstanding logical volume read and write requests
- Minimum buddy buffer size
- Maximum buddy buffer size
- Number of buddy buffers
- Request block count
- Maximum IP message size
- Level of parallelism
- Current setting of the cache option

See "Displaying and Entering Node Information in the SDR" on page 26 for descriptions of these values.

The System Values page of this notebook shows you the values of the information shown above that were set for the node on which this IBM Virtual Shared Disk was defined.

You can use the **vsdatalst**, **statvsd**, or **ctlvsd** commands or SMIT to get the same information.

Displaying Information about HSD Definitions and Clients

To view definition attributes for HSDs, select the definition whose attributes you want to view and click on the notebook icon in the tool bar. The IBM HSD Definition Attributes page shows the following information:

- Name of the HSD
- Minor number of this client
- Stripe size
- Status of the Protect LVCB option (*protect* or *not-protect*)

From the same notebook page, you can click on the HSD Clients tab to see the clients on which this definition has been configured.

To view client attributes for HSDs, select the HSD client icon whose attributes you want to view and click on the notebook icon in the tool bar. The system displays the IBM HSDClient Attributes notebook page, which shows you the following information:

- Name of the HSD
- Node number on which this HSD client is configured
- HSD minor number
- Stripe size
- HSD level of parallelism

- IBM Virtual Shared Disk client reads (512–byte block count)
- IBM Virtual Shared Disk client writes (512–byte block count)
- Owner of this HSD
- Group of this HSD

The Client Values—HSD Client Attributes and System Values—HSD Client Attributes tabs show you the settings of the *protect_lvcb* option.

You can also use the **hsdatalst**, **ctlhsd**, and **lsvsd** commands with no operands, or SMIT, to see this information.

Displaying IBM Virtual Shared Disk and HSD Statistics and Diagnostics

Displaying IBM Virtual Shared Disk Node Statistics

You can use the IBM Virtual Shared Disk Perspective to display the IBM Virtual Shared Disk node statistics and other information about the IBM Virtual Shared Disks defined on a node. The values you see are snapshot values. You will need to check them at different points in time and compare the values to understand what the results mean for your configuration. To interpret the results, see Chapter 5, “Performance and Tuning Considerations for IBM Virtual Shared Disks and Data Striping Devices” on page 53.

The information is displayed on the node attributes notebook page. To open this page from within the IBM Virtual Shared Disk Perspective:

- Click on the appropriate IBM Virtual Shared Disk node icon in the IBM Virtual Shared Disk Nodes work area to select it
- Click on the notebook icon in the tool bar
- When the notebook page is displayed, click on the tab at the bottom of the page that says IBM Virtual Shared Disk Node Statistics. (You may have to page to the right, using the small arrow button at the bottom of the screen, to see the tab for IBM Virtual Shared Disk Node Statistics.)

The following statistics are displayed:

- Status of the IP checksum counters
- SP system product level
- Level of parallelism
- Maximum IP message size
- Requests queued waiting for various resources, such as pbufs
- Average buddy buffer wait queue size
- Different types of rejected operations
- 64-byte unaligned reads
- Communications buffer pool shortages
- Number of timeouts
- Retries
- Information about the status of each IBM Virtual Shared Disk node, including:
  - Node name
  - Volume group name
  - Logical partition size
  - Physical disks and the space used on them
  - IBM Virtual Shared Disks defined on this node

This information is also available from the **statvsd** and **lsvsd** commands.

Displaying HSD Node Statistics

You can use the IBM Virtual Shared Disk Perspective to display HSD node statistics and other information about the HSDs defined on a node. The information is displayed on the node attributes notebook page. To open this page from within the IBM Virtual Shared DiskPerspective:

- Click on the appropriate node icon in the IBM Virtual Shared Disk Nodes work area to select it
- Click on the notebook icon in the tool bar
- When the notebook page is displayed, click on the tab at the bottom of the page that says HSD Node Statistics. (You may have to page to the right, using the small arrow button at the bottom of the screen, to see the tab for HSD Node Statistics.)

The following information is displayed:

- A list of all the HSDs configured on the node
- For each HSD,
  - The stripe size
  - The HSD minor number
  - The number of underlying IBM Virtual Shared Disks
  - The names of the underlying IBM Virtual Shared Disks
  - The HSD level of parallelism
  - The number of read requests not on a page boundary
  - The number of write requests not on a page boundary

You can also get this information from the **lshsd** command.

Resetting IBM Virtual Shared Disk and HSD Statistics

IBM Virtual Shared Disk and HSD statistics are cumulative. If you want to use them to see the results of specific parameter changes, you must reset them, make the change, and then read them after the appropriate time period.

You can reset IBM Virtual Shared Disk and HSD statistics with the IBM Virtual Shared Disk Perspective. In the IBM Virtual Shared Disk Nodes work area of the IBM Virtual Shared Disk Perspective:

- Click on the appropriate node icon to select it
- Click on Actions from the menu bar at the top of the page
- Select IBM Virtual Shared Disk Nodes
- Select Reset All IBM Virtual Shared Disk Statistics or Reset All HSD Statistics.

You can also use the **ctlvsd** or **ctlhsd** commands to reset the statistics.

Running IBM Virtual Shared Disk Diagnostics

If you have problems with your IBM Virtual Shared Disk or HSD configuration, your IBM service representative may ask you to look at the IBM Virtual Shared Disk diagnostics. In the IBM Virtual Shared Disk Nodes work area of the IBM Virtual Shared Disk Perspective:

- Click on the appropriate IBM Virtual Shared Disk node icon to select it
- Click on Actions from the menu bar at the top of the page
- Select IBM Virtual Shared Disk Nodes
- Select Run IBM Virtual Shared Disk Diagnostics.

The diagnostics compare the settings of some IBM Virtual Shared Disk parameters across all the nodes in an IBM Virtual Shared Disk configuration and look for discrepancies that could affect the operation of your IBM Virtual Shared Disks.

You can also use the **vsddiag** command for this purpose.

Changing the States of IBM Virtual Shared Disks

If you haven't installed the IBM Recoverable Virtual Shared Disk licensed program, changes to your configuration or system, application, or network problems may require you to move your IBM Virtual Shared Disks from one state (stopped, suspended, active) to another. (IBM Recoverable Virtual Shared Disk changes the states of your IBM Virtual Shared Disks for you, automatically.) You can do this with the IBM Virtual Shared Disk Perspective. In the IBM Virtual Shared Disk Nodes work area:

- Click on the appropriate IBM Virtual Shared Disk node icon to select it
- Click on Actions from the menu bar at the top of the page
- Select IBM Virtual Shared Disk Nodes
- Select Change IBM Virtual Shared Disk States.

Configuring an IBM Virtual Shared Disk or HSD with the IBM Virtual Shared Disk Perspective starts it. Configuring an IBM Virtual Shared Disk or HSD with the **cfgvsd** or **cfghsdvsd** commands puts a defined IBM Virtual Shared Disk in the stopped state, but does not make it available.

Starting an IBM Virtual Shared Disk

Starting puts a stopped IBM Virtual Shared Disk in the active (and available) state. (This is equivalent to preparing and resuming an IBM Virtual Shared Disk.) Note that for an IBM Virtual Shared Disk to be usable, it must be in the active state on both the server and client nodes.

Preparing an IBM Virtual Shared Disk

Preparing puts a stopped IBM Virtual Shared Disk in the suspended state. In the suspended state, open and close requests are honored. Read and write requests are held until the IBM Virtual Shared Disk is brought to the active state.

Resuming an IBM Virtual Shared Disk

Resuming an IBM Virtual Shared Disk puts a suspended IBM Virtual Shared Disk in the active state. The IBM Virtual Shared Disk remains available and read and write requests that were held are resumed.

Suspending an IBM Virtual Shared Disk

Suspending puts an active IBM Virtual Shared Disk in the suspended state. The IBM Virtual Shared Disk remains available. Read and write requests that were active are suspended and held. All read and write requests subsequent to those that were active are also held.

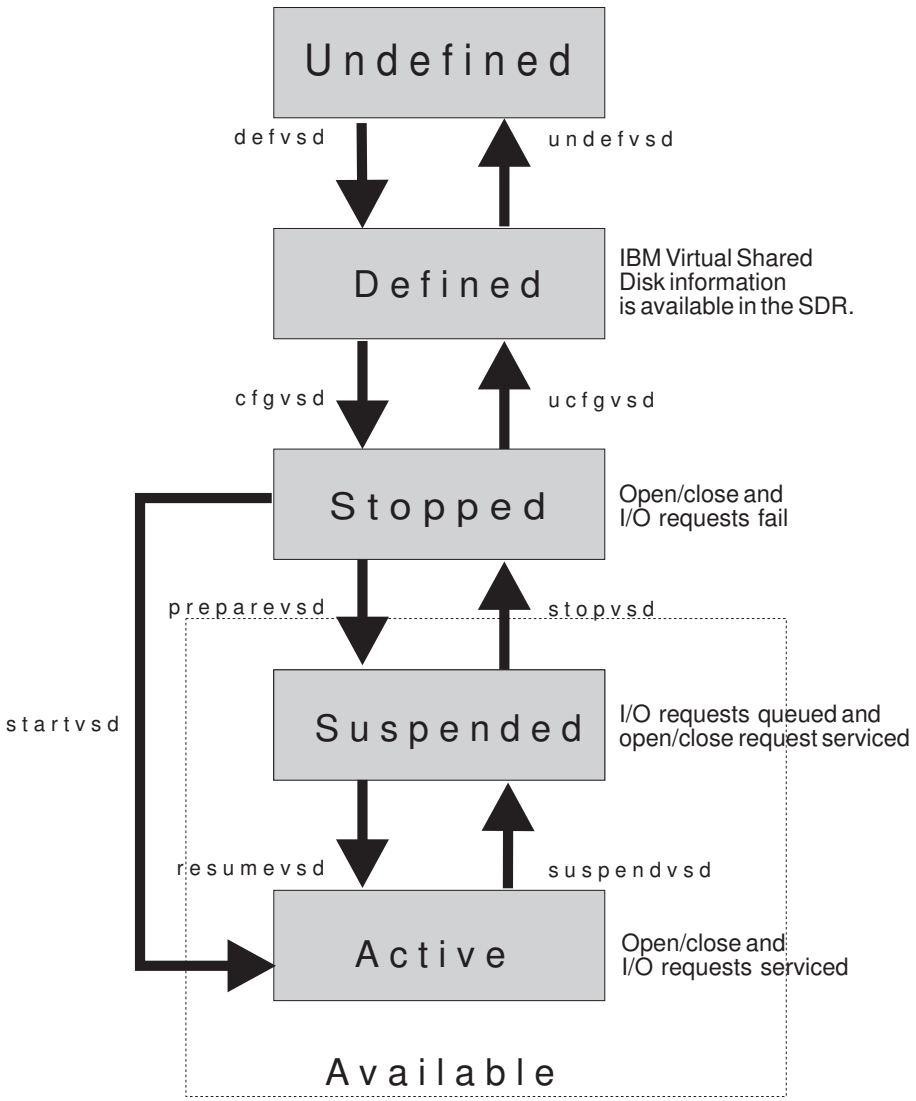
Stopping an IBM Virtual Shared Disk

Stopping puts a suspended IBM Virtual Shared Disk in the stopped state. The IBM Virtual Shared Disk becomes unavailable. All applications that have outstanding requests to a stopped IBM Virtual Shared Disk terminate in error.

Unconfiguring a stopped IBM Virtual Shared Disk makes it inaccessible. It does not, however, undefine or change the definition information for the IBM Virtual Shared Disk in the SDR.

If you use the IBM Recoverable Virtual Shared Disk licensed program, do not issue any commands that change the state of the IBM Virtual Shared Disks (**cfgvsd**, **cfghsdvsd**, **startvsd**, **preparevsd**, **resumevsd**, **suspendvsd**, **stopvsd**, **ucfgvsd**, or **ucfghsdvsd**). IBM Recoverable Virtual Shared Disk handles state changes for you. Issuing any of these commands could cause the recovery processes that IBM Recoverable Virtual Shared Disk manages to fail.

Figure 12 on page 47 shows the commands that move IBM Virtual Shared Disks from one state to another.



Note: Methods on arrows cause transitions

Figure 12. IBM Virtual Shared Disk States and Associated Commands

Making Changes to IBM Virtual Shared Disks on Twin-Tailed Volume Groups

This section describes how to make a logical volume configuration change to your IBM Virtual Shared Disks, such as extending a volume group, changing the size of your IBM Virtual Shared Disks, or adding a new physical disk, and coordinate the timestamps so IBM Recoverable Virtual Shared Disk does not incur unnecessary overhead during recovery processing. If you make a change to your IBM Virtual Shared Disks on twin-tailed volume groups that are managed by IBM Recoverable Virtual Shared Disk without resetting the timestamps, any later recovery processing will cause the volume groups to be exported and then imported to the secondary nodes. This will occur even if you have made the updates to the IBM Virtual Shared Disks at both nodes.

Timestamps are maintained on the primary node, the secondary node, and the IBM Virtual Shared Disk itself.

Making a Change to Only One Side of a Twin-Tailed Volume Group

This procedure describes how to make a change to one side (side A) of a twin-tailed volume group **without** varying off the volume group on that side.

1. Make the change.
2. Issue

   ```
   vsdvgts volume_group_name
   ```

   on side A. This reads the volume group timestamp from the disk and saves the value. The next time the IBM Recoverable Virtual Shared Disk recovery scripts vary on this volume group on this node, the timestamps will be the same and the overhead of importing the volume group will be avoided. If a failover occurs, the timestamps will be different on side B, so the volume group will be exported and then imported.
3. If IBM Virtual Shared Disks have been added, use the IBM Virtual Shared Disk Perspective to configure and start the IBM Virtual Shared Disks on all the nodes that need to be aware of this IBM Virtual Shared Disk, or issue the appropriate IBM Virtual Shared Disk commands.
4. Start IBM Recoverable Virtual Shared Disk if it is not running, using the **ha_vsd reset** command.

Making a Change to Both Sides of a Twin-Tailed Volume Group

This procedure describes how to make a change to both sides (sides A and B) of a twin-tailed volume group **without** varying off the volume group on side A.

1. Make the change, issuing all the appropriate AIX commands on side B to reflect the changes made on side A (the volume group on side A is varied online).
2. Issue

   ```
   vsdvgts -a volume_group_name
   ```

   on side A. This reads the volume group timestamp from the disk and saves the value for both sides A and B, since the same changes have been made on both. The next time the IBM Recoverable Virtual Shared Disk recovery scripts vary on this volume group on either node, the timestamps will be the same and the overhead of importing the volume group will be avoided.
3. If IBM Virtual Shared Disks have been added, use the IBM Virtual Shared Disk Perspective to configure and start the IBM Virtual Shared Disks on all the nodes that need to be aware of this IBM Virtual Shared Disk, or issue the appropriate IBM Virtual Shared Disk commands.
4. Start IBM Recoverable Virtual Shared Disk if it is not running, using the **ha_vsd reset** command.

Making a Change to Both Sides of a Twin-Tailed Volume Group with the Volume Group Varied Off

This procedure describes how to make a change to both sides (sides A and B) of a twin-tailed volume group where both sides are varied off and you intend to explicitly export and import the volume group.

1. Make the change on side A. On side B, export the volume group and import it (issue **exportvg** and **importvg**).
2. Issue

   ```
   vsdvgts -a volume_group_name
   ```

   on either side. This reads the volume group timestamp from the disk and saves the value for both sides A and B, since the changes made on A have been exported to B. The next time the IBM Recoverable Virtual Shared Disk recovery scripts vary on this volume group on either node, the timestamps will be the same and the overhead of importing the volume group will be avoided.
3. If IBM Virtual Shared Disks have been added, use the IBM Virtual Shared Disk Perspective to configure and start the IBM Virtual Shared Disks on all the nodes that need to be aware of this IBM Virtual Shared Disk, or issue the appropriate IBM Virtual Shared Disk commands.
4. Start IBM Recoverable Virtual Shared Disk if it is not running, using the **ha_vsd reset** command.

Removing an IBM Virtual Shared Disk or Data Striping Device

To remove an IBM Virtual Shared Disk, use the **removevsd** command. To remove a Data Striping Device, use the **removehsd** command. The **removehsd** command also removes the underlying IBM Virtual Shared Disks of the HSD that is to be removed.

Adding a Server Node to an Existing IBM Virtual Shared Disk Configuration

To add a new node to an existing IBM Virtual Shared Disk configuration, follow these steps:

1. Enter the necessary information for the new node into the SDR, using the notebook page of the IBM Virtual Shared Disk Perspective. See “Displaying and Entering Node Information in the SDR” on page 26.
2. Shut down any applications that use IBM Virtual Shared Disks.
3. Exit from the Perspectives interface and type in the **createvsd** command for the new configuration, or use the SMIT fastpath, **smit vsd_data**. See “Creating IBM Virtual Shared Disks with the createvsd Command” on page 29
4. Reenter the IBM Virtual Shared Disk Perspective and reconfigure all the IBM Virtual Shared Disks on the system partition, using the Reconfigure All IBM Virtual Shared Disk Clients action from the IBM Virtual Shared Disk Nodes work area. When you do this, the new configuration and the new node information in the SDR take effect.
5. If you are not using the IBM Recoverable Virtual Shared Disk licensed program, start all the IBM Virtual Shared Disks on the system partition using the Start All

IBM Virtual Shared Disk Clients action from the IBM Virtual Shared Disk Nodes work area of the IBM Virtual Shared Disk Perspectives initial panel.

6. If you are using the IBM Recoverable Virtual Shared Disk licensed program, start all the IBM Virtual Shared Disks on the system partition with the **ha_vsd reset** command.
7. You can now restart any application that uses IBM Virtual Shared Disks.

See Appendix A, "IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands" on page 87 for information on the **vsdnode**, **createvsd**, **ucfgvsd**, and **cfgvsd** commands. The commands can be used instead of the IBM Virtual Shared Disk Perspective in the description above.

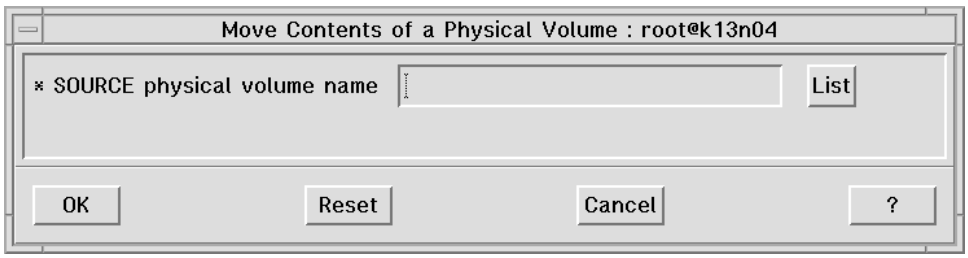
Removing a Node from an Existing IBM Virtual Shared Disk Configuration

To remove a node from an existing IBM Virtual Shared Disk configuration and move the data from the disks on that node to another node in the configuration, follow these steps:

1. Shut down any applications that use IBM Virtual Shared Disks.
2. If you need the data on the IBM Virtual Shared Disks that have been defined on the node that is to be removed, back it up.
3. If the data on the physical disks must be moved, use the **smit migratepv** fastpath to move it.
   a. As the root user on the primary node, enter:

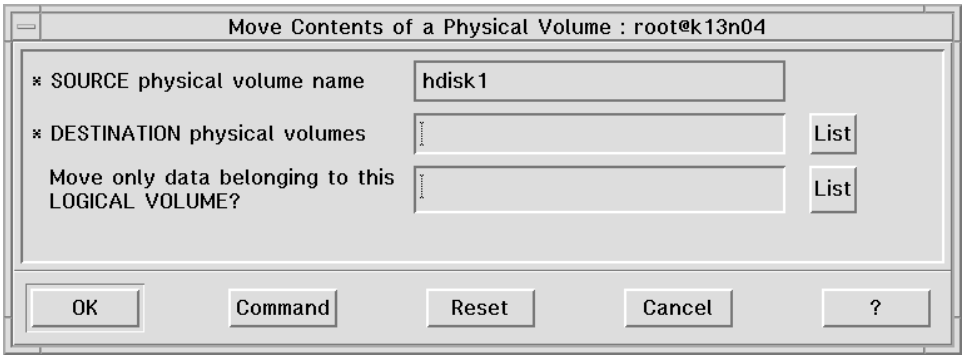
      ```
      smit migratepv
      ```

      SMIT returns a screen similar to the following:



   b. Enter the names of the physical volumes being moved in the SOURCE physical volume name field and press Enter.

      SMIT returns a screen similar to the following. The physical volumes being migrated are entered in the SOURCE physical volume names field.



c. Enter field values as follows:

DESTINATION physical volumes
Enter the names of the physical volumes to which you want to move the data.

Move only data belonging to this LOGICAL VOLUME?
Set this field to **no**.

d. Press Enter to migrate the physical volumes.

e. After the command completes, press F10 to leave SMIT and return to the command line.

4. Using the **removevsd** command, unconfigure and remove all IBM Virtual Shared Disks defined on this node.

   CAUTION:
   The removevsd command removes the logical volumes associated with that IBM Virtual Shared Disk definition and causes you to lose access to all the data on the IBM Virtual Shared Disk.

5. Type in the **createvsd** command for the new configuration, or use the SMIT fastpath, **smit vsd_data**. See Appendix A, "IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands" on page 87 for the command syntax.

6. Reenter the IBM Virtual Shared Disk Perspective and reconfigure all the IBM Virtual Shared Disks on the system or system partition, using the Reconfigure All IBM Virtual Shared Disk Clients action from the IBM Virtual Shared Disk Nodes work area of the IBM Virtual Shared Disk Perspective primary window. When you do this, the new configuration takes effect.

7. If you are not using the IBM Recoverable Virtual Shared Disk licensed program, start all the IBM Virtual Shared Disks on the system partition, using the Start All IBM Virtual Shared Disk Clients action from the IBM Virtual Shared Disk Nodes work area of the IBM Virtual Shared Disk Perspectives primary window.

8. If you are using the IBM Recoverable Virtual Shared Disk licensed program, start all the IBM Virtual Shared Disks on the system partition with the **ha_vsd reset** command.

9. You can now restart any applications that use IBM Virtual Shared Disks.

See Appendix A, "IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands" on page 87 for the syntax of the **removevsd** and **cfgvsd** commands.

Chapter 5. Performance and Tuning Considerations for IBM Virtual Shared Disks and Data Striping Devices

System administrators and application programmers should be aware of the performance characteristics of IBM Virtual Shared Disks and Data Striping Devices and of how they can affect those characteristics.

In reading this section, remember to take the performance characteristics of your database subsystem and applications into account.

Tuning IBM Virtual Shared Disk Performance

The IBM Virtual Shared Disk device driver layer passes all its requests to the underlying Logical Volume Manager. Before you tune the IBM Virtual Shared Disk software, check that the I/O subsystem is not the bottleneck. See *AIX Performance and Tuning Guide* for information on I/O subsystem performance and tuning. If an overloaded I/O subsystem is degrading your system's performance, tuning the IBM Virtual Shared Disk will not help. In the case of I/O subsystem overload, consider spreading the I/O load over more disks or nodes.

We recommend taking the following steps to tune the IBM Virtual Shared Disk for best performance:

1. Use the defaults when defining IBM Virtual Shared Disks (refer to “Displaying and Entering Node Information in the SDR” on page 26)
2. Turn IBM Virtual Shared Disk caching off (if you are not using your system for Online Transaction Processing)
3. Do a performance run to collect statistics on the IBM Virtual Shared Disk, your I/O subsystem, and the CPU on all nodes (or use Performance Monitor to collect information during normal system operation.) Issue **statvsd** several times during the performance run and compare the values for the various statistics. Use **iostat** to check your disk utilization. If you notice increasing numbers of queued requests:
   - If the system is I/O bound (meaning your disks are more than 50% utilized), add disks.
   - If the system is CPU bound, add nodes or spread the workload on the IBM Virtual Shared Disk servers. HSD could be used to spread the workload.
   - If nodes are doing excessive swapping due to insufficient pinned memory, which is used by pbufs, buddy buffers, cache, and the switch pool, reduce cache size.
   - If requests are queuing because of a shortage of buddy buffers or pbufs, you may have disk bottlenecks. Spread the data or add disks to the server nodes.
   - If your application issues requests that are larger than 64KB, set your maximum buddy buffer size to 256KB.
   - If you see too many retries, check for disk bottlenecks. If that's not the problem, consider increasing the switch pool size (see “mbufs and the Switch Pool” on page 56).

© Copyright IBM Corp. 1997, 1998

4. Reset the statistics counter, using the IBM Virtual Shared Disk Perspective or the **ctlvsd** command. (See the *IBM Parallel System Support Programs for AIX: Command and Technical Reference*.)
5. Do another performance run.

You should generally turn IBM Virtual Shared Disk caching off. Memory is better allocated to the operating system itself, for paging, and to the cache belonging to the application using the IBM Virtual Shared Disk. To turn IBM Virtual Shared Disk caching off,

1. Shut down your application
2. Start the IBM Virtual Shared Disk Perspective
3. Select one or more nodes
4. Go to the IBM Virtual Shared Disk Node Attributes notebook page
5. Change the *cache/nocache* option to nocache
6. Close the notebook page and select “Reconfigure all IBM Virtual Shared Disks on a Node” from the Actions list
7. Restart your application.

This changes the *cache/nocache* option for all the IBM Virtual Shared Disks on the selected node or nodes.

If you do use caching, remember that the IBM Virtual Shared Disk component only caches 4KB requests aligned on 4KB boundaries.

IBM Virtual Shared Disk Tunable Parameters

The main IBM Virtual Shared Disk tunable parameters are:

- Logical Volume Manager (striping and other LVM characteristics)
- IP communications adapter (usually the switch)
- IBM Virtual Shared Disk cache buffer
- Buddy buffer
- Maximum I/O request size
- Request blocks
- pbufs
- mbufs

In the following sections we define these parameters and discuss relevant tuning considerations. You should also consider the tunable characteristics of the application that uses IBM Virtual Shared Disks, especially its use of buffers.

Logical Volume Manager (LVM) Tuning Considerations

There is always an associated logical volume for every IBM Virtual Shared Disk device defined and configured in a system. Every IBM Virtual Shared Disk I/O request eventually becomes an I/O request to the associated logical volume (unless you get a cache hit at the server). This mapping of IBM Virtual Shared Disk device I/O requests to the associated logical volume I/O requests is handled transparently by the IBM Virtual Shared Disk software layer. All the performance tuning considerations that apply to a logical volume also apply to IBM Virtual Shared Disk

devices. Refer to *AIX System Management Guide: Operating Systems and Devices* and *AIX Performance and Tuning Guide* for information on the performance tuning of logical volumes.

The High Performance Switch/Scalable POWERparallel Switch

If you configure the IBM Virtual Shared Disk nodes to use the High Performance Switch or the SP Switch (**css0**), the maximum IP message size (utilized by the IBM Virtual Shared Disk driver) should be set to 61440 (60KB). You should ensure the maximum buddy buffer size is 256KB, since the value you assign to *maximum_buddy_buffer_size* in the SDR also limits the maximum size of the request that IBM Virtual Shared Disk sends across the nodes. For example, if you have:

- A request from a client to write 256KB of data to a remote IBM Virtual Shared Disk
- A maximum buddy buffer size of 64KB
- A maximum IP message size of 60KB

the following transmission sequence occurs:

1. The IBM Virtual Shared Disk component divides the 256KB of data into four 64KB requests in four buddy buffers
2. Each 64KB block of data becomes one 60KB packet and one 4KB packet for transmission to the server via IP
3. At the server, the eight packets are reassembled into four 64KB blocks of data, each in a 64KB buddy buffer
4. The server then has to perform four 64KB write operations and return four acknowledgements to the client.

A better scenario for the same write operation would use the maximum buddy buffer size:

- The same 256KB client request to the remote IBM Virtual Shared Disk
- The maximum buddy buffer size of **256KB**
- The maximum IP message size of 60KB

Producing the following transmission sequence:

1. The 256KB request becomes four 60 KB packets and one 16KB packet for transmission to the server via IP
2. At the server, the five packets are reassembled into one 256KB block of data in a single buddy buffer
3. The server then performs one 256 KB write operation and returns an acknowledgement to the client.

The second scenario is preferable to the first because the I/O operations at the server are minimized. A perfect scenario would be one where IBM Virtual Shared Disk does not use buddy buffers at all—when the client request is less than or equal to the maximum IP message size. For example:

- A request from a client to write 60KB of data to a remote IBM Virtual Shared Disk server

- A maximum IP message size of 60KB

When you use the switch, send pool clusters are used instead of buddy buffers as long as the request size is less than the *ip_message_size*, as in the example just cited. Buddy buffers are used only when a shortage in the switch buffer pool occurs or when the size of the request is greater than the *ip_message_size*. If you see buddy buffer shortages, instead of increasing your buddy buffers, you need to increase your switch send pool size. See “mbufs and the Switch Pool.”

mbufs and the Switch Pool

mbufs are used for data transfer between the client and the server nodes by IBM Virtual Shared Disk's own UDP-like protocol and IP protocol (which IBM Virtual Shared Disk uses). If you are using the switch (**css0**) as your communications adapter, IBM Virtual Shared Disk uses mbuf clusters to do I/O directly from the switch's send and receive pools.

If you notice that the indirect I/O statistic (from the IBM Virtual Shared Disk Statistics notebook page or from the output of the **vsdstat** command) is incremented consistently, run **errpt** to check the error log. If you see the line:

```
IFIOCTL_MGET(): send pool shortage
```

you should consider increasing the size of the send and receive pools.

To check the current sizes of the send and receive pools, type:

```
lsattr -l css0 -E
```

The default size for each pool is 524288 bytes (512KB).

To change the sizes of the pools to 4MB, type:

```
/usr/lpp/ssp/css/chgcss -l css0 -a spoolsize=4194304
/usr/lpp/ssp/css/chgcss -l css0 -a rpoolsize=4194304
```

This command increases the send and receive pool size to 4MB.

Note: You must reboot the node for the new sizes to take effect.

We recommend allowing 16MB for mbufs and clusters. You can set this value by issuing:

```
no -o thewall=16384
```

To see what your current system mbuf setting is, type:

```
no -a | grep thewall
```

System performance considerations regarding mbufs and mbuf clusters also apply to IBM Virtual Shared Disk environments. See *AIX Performance and Tuning Guide* for more information.

Buddy Buffers

The IBM Virtual Shared Disk server node uses the buddy buffer to temporarily store data for I/O operations originating at a client node and to handle requests that are greater than the *ip_message_size*. In contrast to the data in the cache buffer, the data in a buddy buffer is purged immediately after the I/O operation completes.

Note: Buddy buffers are used only when a shortage in the switch buffer pool occurs or on certain networks (for example, the Ethernet).

The values associated with the buddy buffer are:

- Minimum buddy buffer size allocated to a single request
- Maximum buddy buffer size allocated to a single request
- Total size of the buddy buffer

These values can be set in the IBM Virtual Shared Disk Perspective or with the **vsdnode** command.

Buddy buffer space is allocated in powers of two. If an I/O request size is not a power of two, the smallest power of two that is larger than the request is allocated. For example, for a request size of 24KB, 32KB are allocated on the server.

If you are using the switch as your adapter for IBM Virtual Shared Disks, we recommend setting 4096 (4KB) and 262144 (256KB), respectively, for minimum and maximum buddy buffer size allocated to a single request.

To define the total size of the buddy buffer, consider the remote I/O throughput for the server and specify the number of maximum-sized buddy buffers in the buffer. For example, if you expect the server to serve 10MB per second on behalf of remote clients and a request spends an average of 60 milliseconds on the server, multiply 10MB per second X 0.06 second and, for safety, double or triple the result for a total buddy buffer size of 1.8MB (eight 256KB maximum buddy buffers).

If the IBM Virtual Shared Disk statistics consistently show requests queued waiting for buddy buffers, do not add more buddy buffers. Instead, increase the size of the switch send pool (see “mbufs and the Switch Pool” on page 56) or spread the data over disks attached to other nodes, to prevent a bottleneck.

Buffer Allocation

Your application should make all new allocated buffers on the page boundary. If your I/O buffer is not aligned on a page boundary, the IBM Virtual Shared Disk device drivers will not parallelize I/O requests to underlying IBM Virtual Shared Disks and performance will be degraded.

The Cache Buffer

Each IBM Virtual Shared Disk device driver, that is, each node, has a single cache buffer, shared by all cacheable IBM Virtual Shared Disks configured on and served by the node. The cache buffer is used to store the most recently accessed data from the cached IBM Virtual Shared Disk devices (associated logical volumes) on the server node. The objective is to minimize physical disk I/O activity. If the requested data is found in the cache, it is read from the cache, rather than the corresponding logical volume.

Data in the cache is stored in 4KB blocks. The content of the cache is a replica of the corresponding data blocks on the physical disks. Write-through cache semantics apply; that is, the write operation is not complete until the data is on the disk.

When you define IBM Virtual Shared Disks with the **createvsd** command, you can specify the *cache* option or the *nocache* option. We recommend that you specify **nocache** (or make the cache buffer small) in most instances (especially in the cases of read-only or other than 4KB applications) for the following reasons:

- Requests that are not exactly 4KB and not aligned on a 4KB boundary will bypass the cache buffer, but will incur the overhead of searching the cache blocks for overlapping pages.
- Every 4KB I/O operation incurs the overhead of copying into or out of the cache buffer, as well as the overhead of moving program data from the processor cache due to the copy.
- There is overhead for maintaining an index on the blocks cached.

If you are running an application that involves heavy writing followed immediately by reading, it may be advantageous to turn the cache buffer on for some IBM Virtual Shared Disks on a particular node. Choose the appropriate size for the cache based on the expected throughput and the expected time lag between writes and reads. For example, if the expected throughput is 100 4 KB-aligned I/O operations per second and reads lag writes by 0.5 seconds, calculate the cache buffer size by multiplying 100 X 0.5 and, as a safety factor, double it for a total of 100 cache blocks.

The IBM Virtual Shared Disk Node Statistics tab of the IBM Virtual Shared Disk Perspective gives detailed statistics on IBM Virtual Shared Disk cache hits and I/O activities. You can also use the **lsvsd -s** command. Both methods will tell you which IBM Virtual Shared Disks are heavily used.

Maximum I/O Request Size

The following factors limit the block size that the IBM Virtual Shared Disk software uses to process each I/O request:

- The largest block size the IBM Virtual Shared Disk software will use is the smaller of *max_buddy_buffer_size* or 256KB.
- If the IBM Virtual Shared Disk uses the switch as its adapter, the *max_IP_msg_size* that could be sent is 65024 bytes (63.5KB). 61440 (60KB) is the recommended value for the IBM Virtual Shared Disk device driver when **css0** is defined as the IBM Virtual Shared Disk adapter in the SDR. The **ctlvsd -M** command can override the default. The *max_IP_msg_size* should be set to a value that is a multiple of 512 bytes and is less than or equal to 63.5KB (when the switch is used) and less than or equal to 24KB (when the switch is not used). The **statvsd** command displays the current value.

  Note: Setting the *max_IP_msg_size* to more than 24KB when using communication adapters with small MTU (maximum transmission unit) could overflow the adapter driver's internal buffers, causing the IP layer to drop packets. This forces the IBM Virtual Shared Disk device driver to retry, sometimes without success, resulting in a timeout.

Every IBM Virtual Shared Disk I/O request is subject to both limits. For example, with a *max_buddy_buffer_size* of 262144 (256KB) and a *max_ip_msg_size* of 61440 (60KB), if an application requests a single 64KB read to an IBM Virtual Shared Disk served by the local node, the request is passed down to the local LVM as one 60KB request and one 4KB request.

The atomicity of an I/O operation is gated by the size of the IBM Virtual Shared Disk request, rather than the size of the application request (if the IBM Virtual Shared Disk request is smaller than the application request the application request would be split down to the size of the IBM Virtual Shared Disk requests).

Request Blocks

The number of request blocks is the total number of physical I/O operations that have been issued by all processes on a node, but have not completed. The number includes requests to both local and remote devices. Because large requests may be broken up into smaller subrequests, the request block number may be several times greater than the actual number of pending read/write requests.

For example, if I/O requests are:

- issued at a rate of 1000 I/O operations per second
- broken on the average into three pieces
- responded to in 50 milliseconds

then the algorithm for calculating the number of request blocks is 1000 X 3 X 0.05 for a total of 150 request blocks. You may want to increase the number somewhat as a safety precaution to account for the possibility of workload surges. Specifying an inordinately large number of request blocks could have a negative performance impact. A large number of request blocks could flood the network, causing servers to run out of mbufs and causing unnecessary retransmissions. What constitutes a large number of request depends on how large the request size is and how many nodes are in the system.

Although the **statvsd** command reports the number of times there is no request block available, queueing for request blocks does not necessarily imply a performance bottleneck. If you increased the number or request blocks infinitely, queueing would occur elsewhere in the operating system.

The number of request blocks can be set with the **vsdnode** command and changed with the **updatevsdnode** command.

pbufs

pbufs are used for actual physical I/O requests that are submitted to the disks. A pbuf shortage degrades overall performance.

pbufs are specified by the read/write request count parameter of the **vsdnode** command as a way of controlling the number of pending device requests for a specific IBM Virtual Shared Disk on its server node. pbufs are specified on a per device basis.

A pbuf shortage reported by the Node Statistics notebook page of the IBM Virtual Shared Disk Perspective or by **statvsd** does not imply that increasing the number of pbufs will improve performance because requests will be queued at other parts of the operating system.

To check on the interactions among request blocks, pbufs, and cache blocks from within the IBM Virtual Shared Disk Perspective:

1. With a node selected, click on the Actions menu item.
2. Select IBM Virtual Shared Disk Nodes
3. From the list of available actions, choose Reset IBM Virtual Shared Disk Statistics
4. Go to the IBM Virtual Shared Disk Node Attributes notebook page
5. Select the IBM Virtual Shared Disk Node Statistics page

6. View the statistics and decide which parameters to tune

The **statvsd** command also gives detailed statistics on request shortages, pbuf shortages and cache block shortages. You can run this command together with **lsvsd -s** before and after an application execution to determine how to tune these parameters to best fit a particular application workload.

Tuning HSD Performance

This section presents information for tuning the HSD.

Note: For more information on tuning the HSD, access the RS/6000 home page at **http://rs6000.ibm.com**. Once you get to the home page, look under **Resources** for information about IBM Virtual Shared Disks and HSDs.

All I/O requests are passed to the underlying IBM Virtual Shared Disks. Thus, the parameters selected for IBM Virtual Shared Disk configuration affect the HSD performance. We discuss HSD tuning in relationship to these additional performance considerations:

- Database (“hot table”) access
- Disk throughput
- CPU overhead
- I/O completion speeds

The parameters that can be tuned to affect performance are:

- Request size
- Stripe size

To take advantage of the parallelism potential, HSD's stripe size should be smaller than the application's request size.

HSD and IBM Virtual Shared Disk request latency is primarily gated by disk I/O. Disk I/O can be separated into three distinct operations:

- Seek delay
- Rotational delay
- Transfer delay

The smaller stripe size, the smaller the transfer delay. HSD striping allows the system to handle IBM Virtual Shared Disk requests in parallel.

If your system is CPU bound, using HSD (with a stripe size greater than the buddy buffer size) can help only with eliminating I/O bottlenecks. Eliminating I/O bottlenecks may improve overall transaction throughput. You should consider spreading the workload by adding more nodes or disks.

If your system is I/O bound, HSD with striping may help. However, you should consider that:

- Processing more information from disks requires more CPU cycles. HSD striping uses some of these CPU cycles. It is important to stay below the point of diminishing returns when you set the stripe size.

- You should set a stripe size large enough to minimize the chances that the same IBM Virtual Shared Disk could be hit more than once by a parallelized application request.
- Disks should be on separate adapters because adapters (such as SCSI) limit the number of concurrent requests they can send to disks that are sharing the same bus.

The only way to fully optimize an HSD to your environment is to try out different configurations. The main obstacle in tuning an HSD is that it is not simple to add or remove IBM Virtual Shared Disks and change the stripe size after the HSD is loaded with data. The only way to do this is to back up the HSD (using **dd**), recreate it, and load it with the new parameters. Therefore, careful consideration should be given to HSD usage planning. If tuned properly, HSD can significantly improve IBM Virtual Shared Disk throughput performance.

Chapter 6. Application Programming Considerations

This chapter tells you how to create or modify your application programs to take advantage of functions in the IBM Virtual Shared Disk and HSD components and in the IBM Recoverable Virtual Shared Disk licensed program.

Application Programming Considerations for IBM Virtual Shared Disks

Application programmers who work with IBM Virtual Shared Disks need to be aware of data integrity considerations, the IBM Virtual Shared Disk transmission protocol, and how to get IBM Virtual Shared Disk information with C language interfaces.

Data Integrity

Use the raw device (also called device special file) to access an IBM Virtual Shared Disk, because the block device user can be presented with stale data from the operating system cache.

An IBM Virtual Shared Disk is a device. On a single node, users must provide and use their own synchronization mechanisms for their reads and writes to insure data integrity. Since the IBM Virtual Shared Disk allows multiple nodes to have simultaneous access to devices, users must provide their own distributed synchronization mechanisms. Neither the IBM Virtual Shared Disk nor the SP system software provides a distributed synchronization mechanism.

Checksum Processing on Unreliable Networks

The switch is a reliable communication medium. The switch adapter performs checksum processing in hardware to provide a reliable network. Some networks do not provide checksum processing (for example, the Ethernet). The IBM Virtual Shared Disk software provides a checksum processing option for use on unreliable networks. You can use the IBM Virtual Shared Disk Perspective or the **cksumvsd** command to turn checksum processing on and off in the IBM Virtual Shared Disk device driver. Because the switch provides checksum processing in hardware, don't use **cksumvsd** if you use **css0** as your IBM Virtual Shared Disk communication adapter. For more information refer to *IBM Parallel System Support Programs for AIX: Command and Technical Reference*.

Note: Checksum processing should be either on or off on all IBM Virtual Shared Disk nodes in a system partition.

IBM Virtual Shared Disk Transmission Protocol

The IBM Virtual Shared Disk implements its own UDP-like IP protocol. The requesting node implements an exponential back-off retransmission strategy. If the remote node does not service the request after about 15 minutes of retransmission, the IBM Virtual Shared Disk fails the application's I/O request. Suspending and resuming an IBM Virtual Shared Disk (done by IBM Recoverable Virtual Shared Disk) gives all requests a fresh start with the full 15 minutes of retransmission time. The time needed for recovery with IBM Recoverable Virtual Shared Disk does not cause application requests to fail.

© Copyright IBM Corp. 1997, 1998

Sequence Numbers

Each node keeps an expected and outgoing sequence number for all other nodes. Each time a node sends a request message to another node, it increments its outgoing sequence number for that node. If a node receives a message, and its sequence number is not within the valid range, based on the expected number for that node, the message is discarded. The valid range is defined as the current value of the expected number for the node, plus a fairly large window size. When a node receives a valid request message from another node, the receiving node resets its expected sequence number for that node to the sequence number in the message.

IBM Recoverable Virtual Shared Disk maintains and resets sequence numbers for its users.

Sequence numbers become significant if a node in the IBM Virtual Shared Disk cluster is rebooted after it has performed IBM Virtual Shared Disk I/O. When the node comes back up, all its sequence numbers will be reset to 0, but all other nodes in the system partition will expect numbers greater than the last sequence number they received from the node before it rebooted, that is, nonzero. Therefore, none of the other existing nodes that received messages from the node before it crashed will now accept requests from the rebooted node because the sequence number is below the expected value. The rebooted node continues to retransmit, increasing the outgoing sequence number each time, until it eventually reaches the expected range on all the other nodes. Some requests, however, may exhaust their retry limits and fail. The rebooted node will probably accept requests from the existing nodes as their sequence number will be within the range of the expected count (0) and the (large) window size.

When one node's sequence numbers are out of synchronization with the rest of the IBM Virtual Shared Disk nodes, it will look like all IBM Virtual Shared Disk I/O to remote servers from the out-of-synch node is hung. After about 15 minutes of retransmissions, the I/O will start to fail if the system administrator does not intervene and reset the sequence numbers.

Checking and Resetting Sequence Numbers: To check sequence numbers from the IBM Virtual Shared Disk Perspective:

- Select the appropriate node in the IBM Virtual Shared Disk Nodes work area
- Click on the notebook icon in the tool bar
- When the IBM Virtual Shared Disk Node Attributes notebook page is displayed, look for the IBM Virtual Shared Disk Node Statistics tab at the bottom of the page. You may have to click on the small arrow at the bottom of the page to see more tabs.

If you are not using IBM Recoverable Virtual Shared Disk and you need to reset the sequence numbers:

- Select the appropriate node
- Click on Actions
- Click on IBM Virtual Shared Disk Nodes
- Click on Reset IBM Virtual Shared Disk Sequence Numbers.

You can use **statvsd** on client and server nodes to check the sequence numbers. Use the **-R** and **-r** options of the **ctlvsd** to reset the sequence number of all or some of the nodes to 0 (both the expected and the outgoing). Before issuing **ctlvsd**, put the IBM Virtual Shared Disk in suspended state. (See the **statvsd** and **ctlvsd** command reference pages for more details.)

Getting IBM Virtual Shared Disk Information with C Interfaces

Note: The header file /**usr/include/vsd_ioctl.h** is installed with the **ssp.csd.vsd** option. Structures defined in **vsd_ioctl.h**may change from release to release.

The IBM Virtual Shared Disk base code has a sample directory: /**usr/lpp/csd/samples**.

This directory contains the following files:

vsdinfo.c When compiled, generates a binary that shows IBM Virtual Shared Disk information for the **vsd_name** given from the command line.

vsdsinfo.c When compiled, generates a binary that lists information for all configured IBM Virtual Shared Disks.

Application Programming Considerations for HSDs

Read this section for information on application programming performance with HSD and on how to use HSD through C language interfaces.

Allocating I/O Buffers

The address of your I/O buffer must be aligned on a 4KB boundary. If it is not, the HSD read and write routines will **not** parallelize I/O requests to underlying IBM Virtual Shared Disks. Performance will be less than optimal.

Set the environment variable MALLOCTYPE to **3.1** to make all new allocated buffers (MALLOC) at the page boundary.

Getting HSD Information with C Interfaces

The HSD base code has a sample directory: /**usr/lpp/csd/samples**.

This directory contains the following files:

hsdinfo.c When compiled, this file generates a binary that shows Data Striping Device information for the **hsd_name** given from the command line.

hsdsinfo.c When compiled, this file generates a binary that lists information for all configured Data Striping Devices.

Note: The header file /**usr/include/hsd_ioctl.h** is installed with the **ssp.csd.hsd** image. Structures defined in **hsd_ioctl.h** may change from release to release. This may require that you recompile your applications that utilize C interfaces to the HSD to run on new releases. If you see IOCTL failures in your application after migrating to a new release, try recompiling the application.

Making Your Application Recoverable

You can code your program to use the **hc** daemon of the IBM Recoverable Virtual Shared Disk software to aid in application recovery. A data management subsystem is the type of program that might use **hc**'s services.

Your program should include the following **hc.h** header file:

```
#include <hc.h>
```

Your program connects to a socket where **hc** is listening. If a node fails or the client application on a node fails, **hc** sends a reconfiguration message according to the protocol defined in the **hc.h** header file. Your program should check for messages from that socket and react accordingly.

The hc subsystem can support mutliple applications by running multiple instances of hc. Each hc is invoked by a different **hc.vsd** script.

To create a new instance of the hc subsystem, make a copy of the **hc.vsd** script (from **usr/lpp/csd/bin**) and rename it; for example, *hc.myappname*. Then edit the newly-created *hc.myappname* and change the *INSTANCE* variable to *myappname*. You should also export the path to the well-known socket for communication with your application, inserting the following line into the script:

```
export CLIENT_PATH=myappath/
```

where PATH is an absolute pathname. The hc subsystem then creates the socket **myappath/myappname** to communicate with your application.

As a convenience, the **hc** program runs an optional script: **/usr/lpp/csd/bin/hc.activate** once when it initializes (note that **/usr/lpp/csd/bin/hc.deactivate** automatically runs first). The **hc.activate** script is passed two parameters; the local node number and the path name of the UNIX domain stream socket that it serves. Your recovery program is started according to the value you gave **INSTANCE**. If `INSTANCE=myappname`, for example, **hc** calls **myappname.activate**

If the script doesn't complete, **hc** won't communicate with the application. An application that checks connectivity through **hc** running on the other clients can't perform that check until the **hc.activate** script completes. If you need **hc** to continue processing, fork a background process from the **hc.activate** script. Then exit the **hc.activate** script with an exit code of 1, which indicates that communication with the application does not have to be complete for **hc** to continue processing. **hc.activate** is executed only once.

The **hc** program also runs the **/usr/lpp/csd/bin/hc.deactivate** script upon detecting failure of its client. The same parameters are passed as are passed to **/usr/lpp/csd/bin/hc.activate**, letting you restart the client if you desire. The client may fail either by exiting or by failure to respond to a **ping** message within 10 minutes.

Note: Ten minutes is the default for the **PING_DELAY**. That is, if a client application fails to respond to a ping sent from the **hc** program within 10 minutes, **hc** will consider the application to have failed and will invoke **hc.deactivate.**

After **hc.deactivate** runs, it sends a **node down** message to the other nodes.

Some application programs running under some system loads may require a longer **PING_DELAY** timeframe. To change the **PING_DELAY**, edit the **hc.vsd** script. For example, if you wish to increase the **PING_DELAY** to 15 minutes, change the line in the **hc.vsd** script that begins **export SCRIPT_PATH** to:

```
export SCRIPT_PATH=...PING_DELAY=900
```

If **hc** fails, the application will receive a zero-length message over the socket and should shut itself down.

Applications are responsible for cleaning up system resources when they complete or fail; **hc.deactivate** should be used for this purpose.

Preserving Data Integrity During Application Recovery

If a recoverable database application is running in a system and one of the nodes in that system fails, any data that had been locked by the failed application instance is returned to a consistent state. If the node that failed is still capable of issuing I/O requests to IBM Virtual Shared Disks, those requests must not be carried out. An application's recovery script can use the **fencevsd** command to prevent I/O operations from failed nodes to IBM Virtual Shared Disks in the system.

When the node that failed is active again, the application's recovery script can issue the **unfencevsd** command to permit it to issue I/O requests to IBM Virtual Shared Disks again.

If a primary (server) node must be fenced from the IBM Virtual Shared Disks whose data requests it handles, you can use the **vsdchgserver** command to switch the serving node to the secondary (backup) node for those IBM Virtual Shared Disks.

You can check the fenced status of all the nodes in the system with the **lsfencevsd** command, which displays a map of all fenced IBM Virtual Shared Disks and the nodes that are fenced from them.

Chapter 7. What You Need to Know about the IBM Recoverable Virtual Shared Disk Software

This chapter explains how the IBM Recoverable Virtual Shared Disk software works.

The IBM Recoverable Virtual Shared Disk Recovery Subsystems

The IBM Recoverable Virtual Shared Disk recovery subsystems, rvsd and hc, respond to changes in the status of the system by running recovery scripts and notifying client applications. The subsystems operate as daemons named **rvsdd** and **hcd**. They use the utilities of the Group Services component of PSSP. For more information about Group Services, see *IBM Parallel System Support Programs for AIX: Group Services Programming Guide and Reference*.

The following sections describe the functions of the rvsd and hc subsystems.

The rvsd Subsystem

The rvsd subsystem controls recovery for the IBM Recoverable Virtual Shared Disk licensed program. It invokes the recovery scripts whenever there is a change in the group membership. The **ha.vsd** command controls the rvsd subsystem. When a node goes down or a disk adapter or cable fails, rvsd notifies all surviving processes in the remaining IBM Virtual Shared Disk nodes, so they can begin recovery. If a node fails, recovery involves switching the ownership of a twin-tailed disk to the secondary node. If a disk adapter or cable fails, recovery involves switching the server node for a volume group to the secondary node. When the failed component comes back up, recovery involves switching disk or volume group ownership back to the primary node.

Communication adapter (**css** or **en**) failures are treated the same as node failures. Recovery for twin-tailed volume groups consists of switching to the secondary server.

The *primary node* is a node that is physically connected to a set of IBM Virtual Shared Disks and will always manage them if it is active. The *secondary node* is a node that is physically connected to a set of IBM Virtual Shared Disks and will manage them only if the primary node becomes inactive. Primary and secondary nodes are defined with the **createvsd** and **createhsd** commands.

A *client* is a node that has access to IBM Virtual Shared Disks but is not physically connected to them. Some nodes in a system partition may not have any access to IBM Virtual Shared Disks defined on that system and are neither servers nor clients.

The rvsd subsystem uses the notion of *quorum*, the majority of the IBM Virtual Shared Disk nodes, to cope with communication failures. If the nodes in a system partition are divided by a network failure, so that the nodes in one group cannot communicate with the nodes in the other group, rvsd uses the quorum to decide which group continues operating and which group is deactivated. In previous releases of the SP system, quorum was defined as a majority of all the nodes in a system partition. With IBM Recoverable Virtual Shared Disk Version 1 Release 2

© Copyright IBM Corp. 1997, 1998

and later, quorum is based on nodes that have been defined as IBM Virtual Shared Disk nodes. You can override the system default value for the quorum with the **ha.vsd quorum** command.

Table 5 shows how the daemons in the nodes in a system partition react as inactive nodes come back up and as active nodes fail. The table shows the changes that affect three of the nodes in a system that has more than three nodes.

Table 5. Recovery Actions When Nodes Fail

Nodes and Clients	Node Is Active	Recovery Scenario (Node Is Inactive)
Node1, primary	• Daemons running on the other nodes in the system partition accept node1 into the group. • All IBM Virtual Shared Disks on node1 become active. • All clients designate node1 as the manager for the node1 IBM Virtual Shared Disks and designate those IBM Virtual Shared Disks as active.	• All IBM Virtual Shared Disks defined on node1 are put into suspended state on all clients. • Daemons running on the other nodes in the system partition remove node1 from the group. • Node2, the secondary node, takes over the management of node1's IBM Virtual Shared Disks. • All clients designate node2 as the server for node1's IBM Virtual Shared Disks and put those IBM Virtual Shared Disks into active state.
Node2, secondary to node1	• Daemons running on the other nodes in the system partition accept node2 into the group. • If node1 is active, there is no change to the status of the node1 IBM Virtual Shared Disks. • If node1 is inactive, node2 takes over the management of the node1 IBM Virtual Shared Disks. The node1 IBM Virtual Shared Disks are designated as active and managed by node2 on all clients.	• If node1 is active, there is no change to the status of its IBM Virtual Shared Disks. • If node1 is inactive, the node1 IBM Virtual Shared Disks are put into stopped state on all clients. They remain in stopped state until node1 or node2 comes back up. • Daemons running on the other nodes in the system partition remove node2 from the group.
Node3, client	• Daemons running on the other nodes in the system partition accept node3 into the group. • All IBM Virtual Shared Disks on active nodes for which node3 is a client are put into active state from node3's point of view.	• All IBM Virtual Shared Disks defined on node3 are put into stopped state from node3's point of view. • Daemons running on the other nodes in the system partition remove node3 from the group.

Disk Cable and Disk Adapter Failures

Hardware interruptions at the disk are known as EIO errors. IBM Recoverable Virtual Shared Disk can recover by volume group from some kinds of EIO errors, for example, disk cable and disk adapter failures. These failures can affect some of the volume groups defined on a node without affecting other volume groups. IBM Recoverable Virtual Shared Disk switches the server function from the primary node to the secondary node for the failed volume groups on the node, without changing the server for those volume groups that have not failed.

When the cause of the failure has been located and repaired, use the **vsdchgserver** command to restore the server function to the primary setting. This does not happen automatically.

Communication Adapter Failures

Communication adapter failure is supported for SP switch and Ethernet adapters. When communication adapter recovery is enabled, an adapter failure is promoted to an IBM Recoverable Virtual Shared Disk node failure so that twin-tailed volume groups can fail over to the secondary server.

Only twin-tailed volume groups on nodes connected by SP switch and Ethernet adapters will recover from adapter failure. Other communication adapters such as FDDI and Token Ring can be used, but volume groups connected by these will not switch to the secondary server in the event of adapter failure.

Communication adapter recovery is enabled by default but can be disabled by issuing the **ha.vsd adapter_recovery off** command. Use this command when you have supported communication adapters and do not want their failures promoted to node failures.

Issue **ha.vsd query** to determine whether adapter recovery is enabled or disabled. The output will be similar to the following example, where *adapter_recovery* can be `on` or `off`, and *adapter_status* can be `up`, `down`, or `unknown`.

```
Subsystem         Group            PID     Status
rvsd              rvsd             13660   active
rvsd(vsd): quorum= 8, active=0, state=idle, isolation=member,
              NoNodes=2, lastProtocol=nodes_joining,
              adapter_recovery=on, adapter_status=down.
```

The hc Subsystem

The hc subsystem is also called the Connection Manager. It supports the development of recoverable applications. Chapter 6, “Application Programming Considerations” on page 63 provides more information on how to write recoverable applications. The hc subsystem maintains a membership list of the nodes that are currently running hc processes and an incarnation number that is changed every time the membership list changes. The hc subsystem shadows the rvsd subsystem; recording the same changes in state and management of IBM Virtual Shared Disks that rvsd records. The difference is that hc only records these changes after rvsd processes them, to assure that IBM Recoverable Virtual Shared Disk recovery activities begin and complete before the recovery of hc client applications takes place. This serialization helps ensure data integrity. It also explains why hc cannot run on a node where rvsd is not running.

You can use the **fencevsd** command to implement application recovery procedures that are independent of IBM Recoverable Virtual Shared Disk recovery. The application can request IBM Recoverable Virtual Shared Disk to fence nodes with failing application instances from nodes where the application is running correctly. See “Preserving Data Integrity During Application Recovery” on page 67 for more information.

Another important characteristic of the hc subsystem is that it waits for its client application to connect before joining the hc group. If the client application loses its connection, the hc subsystem runs the **hc.deactivate** script and leaves the group. This means that the hc group's membership list corresponds to the list of nodes in which the client application is currently running.

Startup Time

When you start up your SP, the IBM Recoverable Virtual Shared Disk software does the following:

- Configures **all** defined IBM Virtual Shared Disks at system startup using **cfgvsd -a**. All the IBM Virtual Shared Disks on the nodes are in stopped state.
- When quorum is reached, the IBM Recoverable Virtual Shared Disk software activates all IBM Virtual Shared Disks with servers in the active group.
  - All IBM Virtual Shared Disks on the active nodes that also have a server in the active group will be in the active state.
  - If an IBM Virtual Shared Disk's primary server node is in the active group, the primary will be the server for that IBM Virtual Shared Disk.
  - If an IBM Virtual Shared Disk's primary server is not in the active group but the secondary server is, the secondary will be the server for that IBM Virtual Shared Disk.
  - All IBM Virtual Shared Disks without a server in the active group will be in the stopped state.
- If quorum is lost, all the IBM Virtual Shared Disks are put into the stopped state. When quorum is active again, the IBM Virtual Shared Disks are put into appropriate states based on the list above.

The IBM Recoverable Virtual Shared Disk software actively manages all configured IBM Virtual Shared Disks. The recovery scripts invoke the **preparevsd**, **resumevsd**, **suspendvsd**, **stopvsd**, **varyonvg**, and **varyoffvg** commands asynchronously and automatically in the background as IBM Virtual Shared Disk server nodes become active or fail.

If you have any script that issues any IBM Virtual Shared Disk command, you should disable it so it does not interfere with IBM Recoverable Virtual Shared Disk recovery. Do not issue any commands that change the state of IBM Virtual Shared Disks.

The commands are:

cfgvsd
cfghsdvsd
preparevsd
resumevsd
startvsd
stopvsd
suspendvsd

Notes:

1. In order for ha to be activated, there must be a quorum.
2. After you install your system, you can start IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk with the **ha_vsd reset** command. See the **ha_vsd** reference page in *IBM Parallel System Support Programs for AIX: Command and Technical Reference* or in Appendix A, "IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands" on page 87, for the command syntax.

IBM HSD Recovery Considerations

When an IBM Virtual Shared Disk node fails and its secondary node takes over, the takeover is transparent to the HSDs. The HSDs work with the IBM Virtual Shared Disks and do not know where the IBM Virtual Shared Disk servers are.

If the HSD driver is working with the IBM Recoverable Virtual Shared Disk software, the HSDs are recovered, too.

Chapter 8. IBM Recoverable Virtual Shared Disk Recovery Scenarios

The recovery scenarios describe what an administrator or operator may see when the IBM Recoverable Virtual Shared Disk software goes into action to recover from system problems.

Recognizing Recovery

You know that the IBM Recoverable Virtual Shared Disk software is performing recovery processing when you see IBM Virtual Shared Disks that were active in the suspended state when you did not put them there. The IBM Virtual Shared Disk Perspective main panel displays the states of your IBM Virtual Shared Disk clients. To see them, select a node in the IBM Virtual Shared Disk Nodes work area and then click on the expansion arrow. All the clients on that node will be displayed and the icons will show what state each client is in.

If you recognize that recovery is not taking place normally, verify that the recovery daemon is active on all nodes. Use **ha.vsd query** and **hc.vsd query** to see if the rvsd and hc subsystems are active. **ha.vsd query** returns output in the following format:

```
Subsystem         Group            PID     Status
rvsd              rvsd             13660   active
rvsd(vsd): quorum= 8, active=0, state=idle, isolation=member,
           NoNodes=2, lastProtocol=nodes_joining,
           adapter_recovery=on, adapter_status=down.
```

hc.vsd query output looks like this:

```
Subsystem         Group            PID     Status
hc.hc              rvsd             16510   active
 hc(hc): active=0, state=waiting for client to connect
```

Planning For Recovery

You should have disabled or removed all user-provided scripts that issue change-of-state commands to IBM Virtual Shared Disks.

Monitoring the IBM Recoverable Virtual Shared Disk activity with the IBM Virtual Shared Disk Perspective can help you to be aware of potential problems earlier.

The format for each of the following recovery scenarios is:

1. Symptom
2. Detection
3. Affected components
4. Recovery steps
5. Restart

© Copyright IBM Corp. 1997, 1998

IBM Virtual Shared Disk Client or Server Node Failure

- **Symptoms**

  A node has either hung or has failed.

- **Detection**

  Node failure was detected by the **ha** subsystem. An operator may have seen the change in state displayed by the IBM Virtual Shared Disk Perspective.

- **Affected Components**

  The affected components can include everything accessing the IBM Virtual Shared Disks, including the IBM Virtual Shared Disks themselves, the application and its components, the **ha** subsystem and its related functions, and the IBM Recoverable Virtual Shared Disk software running on the failed nodes.

- **Recovery Steps**

  The recovery steps are:

  1. The **ha** subsystem running on surviving nodes notice the node failure. Recovery then begins on the surviving IBM Virtual Shared Disk nodes.
  2. The recovery services of surviving nodes suspend IBM Virtual Shared Disks served by the failing node. The IBM Virtual Shared Disk recovery process puts the suspended IBM Virtual Shared Disks into the active state on the secondary node.
  3. Optional subscribers to **hc** on surviving nodes are informed of the membership change.

- **Restarting the Failed Node**

  Some of the following actions must be done manually; others are done automatically by the IBM Recoverable Virtual Shared Disk software. You should have procedures in place to instruct operators when and how to do the manual operations.

  1. An operator reboots the failed node.
  2. An operator may need to issue the **Estart** or **Eunfence** command to enable the rebooted node to access the switch. Nodes can be set up to reboot automatically, using the **Estart -M** command, which starts the monitor function.
  3. The IBM Recoverable Virtual Shared Disk software components are automatically brought up on reboot, once the communications adapter is available.
  4. The **hc.activate** script, if present, is invoked.
  5. The rebooted node rejoins the active group. If it was a primary node, it takes over again as a primary node.
  6. Optional subscribers to **hc** on surviving nodes are informed of the membership change.

Switch Failure Scenarios

The sequence of events triggered by switch failure depends on whether Adapter Recovery is enabled.

When Adapter Recovery is Enabled

- **Failure Symptoms**

  IBM Recoverable Virtual Shared Disk initiates recovery on failing nodes. Errors may be generated by applications running on these nodes. Error reports may be generated as well.

- **Detection**

  Node failure is detected by the **ha** subsystem. An operator may see the change in state displayed by the IBM Virtual Shared Disk Perspective.

- **Affected Components**

  The affected components can include everything accessing the IBM Virtual Shared Disks, including the IBM Virtual Shared Disks themselves, the application and its components, the **ha** subsystem and its related functions, and the IBM Recoverable Virtual Shared Disk software running on the failed nodes.

- **Recovery Steps and Restart**

  Follow problem determination procedures for switch failure. See *IBM Parallel System Support Programs for AIX: Diagnosis and Messages.*

When Adapter Recovery is Disabled

- **Failure Symptoms**

  Remote IBM Virtual Shared Disk I/O requests hang and then fail after about 15 minutes. The IBM Virtual Shared Disk clients see a time out error.

- **Detection**

  The IBM Recoverable Virtual Shared Disk software neither detects nor handles switch failure when adapter recovery is disabled or when a non-supported adapter is used. An operator using the IBM Virtual Shared Disk SP Perspective may recognize that **switch_responds** for the node is off.

- **Affected Components**

  Applications using IBM Virtual Shared Disks will hang and then fail after about 15 minutes.

- **Recovery Steps and Restart**

  An operator issues the **Estart** or **Eunfence** command to restart the switch.

  If **Estart** or **Eunfence** fails, use standard diagnostic methods for handling switch problems (see the switch error logs in *IBM Parallel System Support Programs for AIX: Diagnosis and Messages*).

  The Problem Management interface could be used to run a script that would stop IBM Recoverable Virtual Shared Disk on switch failures and restart IBM Recoverable Virtual Shared Disk when the switch is active again.

HA Topology Services or Recovery Service Daemon Failure

- **Failure Symptoms**
  - At the node where the failure occurred, all the IBM Virtual Shared Disks stop and restart, causing I/O errors to the application.
  - At other nodes, the IBM Virtual Shared Disks served by the problem node are switched to their secondary servers briefly, then returned to the primary server.
- **Detection**

  There is no automatic method to determine that the daemons are failing.
- **Recovery Steps**
  1. In the unlikely event that the recovery service daemons hang, no recovery is performed.
  2. At the node where the failure occurred, the IBM Virtual Shared Disks remain in the active state, but subsequent node failures or reboots can cause I/O requests to remote IBM Virtual Shared Disks to hang and fail after 15 minutes. Issue the **ps** command at that node to check for the **rvsd** and **hc** daemons.
  3. At the other nodes, some IBM Virtual Shared Disks remain indefinitely in the suspended state. All the IBM Recoverable Virtual Shared Disk daemons are present and their presence can be verified with **ps**.
  4. When the daemons are available again, you can manually issue **ha_vsd reset** on the problem node. If this is insufficient, reboot the problem node.

  For more information, see *IBM Parallel System Support Programs for AIX: Group Services Programming Guide and Reference*

Hardware (EIO) Failures

- **Failure Symptoms**

  I/O errors occur on some or all of the IBM Virtual Shared Disks served from a node.
- **Detection**

  There is no automatic method of determining that volume group failures are occurring. A hardware error return code (EIO) is returned to the IBM Virtual Shared Disk device driver.
- **Affected Components**

  The IBM Virtual Shared Disk application subsystem cannot access the data on the failed volume groups.
- **Recovery Steps**
  1. The volume group that contains the failed IBM Virtual Shared Disks is automatically put into the suspended state.
  2. If the secondary node is available, the volume group failure recovery procedure takes place and the secondary node begins handling requests for the IBM Virtual Shared Disks that share the failed volume group.

3. If the secondary node is inactive or if the secondary node was already acting as the server for this volume group, the IBM Virtual Shared Disks that share the failed volume group are put into the stopped state. IBM Recoverable Virtual Shared Disk is not able to recover in this situation.
4. Correct the condition that caused the EIO error. You can now issue the **vsdchgserver** command to change the server function back to the primary node.

Note: If you mirror data from each IBM Virtual Shared Disk to a IBM Virtual Shared Disk on another adapter, you will never need to recover from this type of error.

Chapter 9. Diagnosis and Messages Information

This chapter provides information on debug files and messages that can aid you in diagnosing and resolving IBM Recoverable Virtual Shared Disk software problems.

Diagnosis Information

There are two IBM Recoverable Virtual Shared Disk script debug files that provide information that can help you in diagnosing problems with the IBM Recoverable Virtual Shared Disk software:

/var/adm/SPlogs/csd/vsd.log
: Provides a history of actions taken by the recovery scripts (driven by rvsd)

/var/tmp/vsd.debuglog
: Provides verbose descriptions of actions taken by the recovery scripts (driven by rvsd)

The **errpt** command, which can be used to generate a report of logged errors and any available console logs, may also prove helpful.

Messages

2506-001 Variable text.

Explanation: These messages come from the **rvsdd** daemon, which provides services to the IBM Recoverable Virtual Shared Disk software on the SP The text of the message will reveal either a usage, configuration, resource, or internal problem.

User Response: The **rvsdd** daemon is started automatically by the internal SP heartbeat service on all SP rack nodes. This daemon is only required on nodes defined with a non-null VSD_adapter in the SDR, and will exit otherwise.

If the **hc** daemon is running on a node, it will monitor the condition of the **rvsdd** daemon. If **hc** detects **rvsdd** exiting, if **rvsdd** detects **hc** exiting, or if **init** detects **heartbeatd** exiting, automatic recovery will restart **rvsdd**. Therefore, no action is normally required.

If you believe **rvsdd** is broken on a node, you can issue the **/usr/lpp/csd/bin/ha_vsd reset** command to cause **heartbeatd** and thus **rvsdd** and **hc** recovery to be attempted manually.

If the problem persists, follow your local procedures for obtaining service.

2506-002 Variable text.

Explanation: These messages come from the **hc** daemon, which provides services to a recoverable service on the SP. The text of the message will reveal either a usage, configuration, resource, or internal problem.

User Response: The **hc** daemon is started automatically by the internal SP heartbeat service on those SP rack nodes that have a non-null VSD_adapter value defined in the SDR.

If the **hc** daemon is running on a node, it will monitor the condition of the **rvsdd** daemon. If **hc** detects **rvsdd** exiting, if **rvsdd** detects **hc** exiting, or if **init** detects **heartbeatd** exiting, automatic recovery will restart **hc**. Therefore, no action is normally required.

If you believe **hc** is broken on a node, you can issue the **/usr/lpp/csd/bin/ha_vsd reset** command to cause **heartbeatd** and thus **hc** and **rvsdd** recovery to be attempted manually.

If the problem persists, follow your local procedures for obtaining service.

© Copyright IBM Corp. 1997, 1998

Chapter 10. IOCTL Subroutines

fence subroutine

Purpose

fence – Allows you to request and change the IBM Virtual Shared Disk fence map. The fence map controls whether IBM Virtual Shared Disks can send or satisfy requests from IBM Virtual Shared Disks at remote nodes.

Syntax

```
#include <vsd_ioctl.h>
int ioctl(FileDescriptor, Command, Argument)
int FileDescriptor, Command;
void *Argument;
```

Parameters

FileDescriptor
Specifies the open file descriptor for which the control operation is to be performed.

Command
Specifies the control function to be performed. The value of this parameter is always **GIOCFENCE**.

Argument Specifies a pointer to a **vsd_fence_map** structure.

The flags field of the **vsd_fence_map** structure determines the type of operation that is performed. The flags could be set with one or more options using the OR operator. These options are as follows:

VSD_FENCE_COMPARE
Indicates that all the bits in the **node_FenceMask** bitmap represent the current status for a given VSD minor, and the **node_FenceMask** bitmap is the new status to be applied by this request. The request will succeed only if the current status depicted in the **node_FenceMask** is valid.

VSD_FENCE_FORCE
If this option is specified, a node can unfence itself.

VSD_FENCE_GET
Denotes a query request.

VSD_FENCE_SET
Denotes a fence request.

VSD_FENCE_SWAP
Indicates that the **node_FenceMask** bitmap denotes which bits in the **node_FenceMap** the request applies to. Only nodes whose bit is turned on in the **node_FenceMask** bitmap may change status.

© Copyright IBM Corp. 1997, 1998

Description

The fence-map is an array with the following definition:

```
typedef struct
unsigned int   nvsds; /* number of elements in the vsd_Fence
                         array */
ulong          flags; /* specifies the type of request */
vsd_Fence_t    *Map;  /* array of vsd_Fence structures */
} vsd_FenceMap_t;
```

nvsds represents the number of elements in the **vsd_Fence** array. The flags specify the type of request. The map is an array of the following structure:

```
typedef struct
{
    unsigned int      vsd_minor;       /* The bitmap is associated
                                          with this minor */
    vsd_Fence_Bitmap  node_FenceMap;   /* Nodes status - fenced or
                                          unfenced */
    vsd_Fence_Bitmap  node_FenceMask;  /* Determine how node_FenceMap
                                          is used */
} vsd_Fence_t;
```

The *vsd_minor* represents the IBM Virtual Shared Disk device's minor number. The **node_FenceMap** bitmap denotes which nodes are fenced (bit turned on) or unfenced (bit turned off) from the IBM Virtual Shared Disk. You could use the **MAP_SET** and **MAP_CLR** macros defined in **vsd_ioctl.h** to turn a given bit in the map on or off. The **node_FenceMask** is used to interpret the **node_FenceMap** bitmap, depending on the flags field.

You can submit a fence map that changes the current IBM Virtual Shared Disk map (maintained by the IBM Virtual Shared Disk subsystem) according to the flags set in the **vsd_fence_map** structure. The request is considered a delta to the current IBM Virtual Shared Disk map. You can also utilize this ioctl to query the current IBM Virtual Shared Disk map.

Return Values

If the request succeeds, a 0 will be returned by the ioctl. In the case of an error, a value of -1 is returned with the global variable **errno** set to identify the error.

Error Values

The **fence** ioctl subroutine can return the following error codes:

EACCES Indicates that an unfence was requested from a fenced node without the **VSD_FENCE_FORCE** option.

EINVAL Indicates an invalid request (ambiguous flags or unidentified IBM Virtual Shared Disks).

ENOCONNECT
Indicates that either the primary or the secondary node for an IBM Virtual Shared Disk to be fenced is not a member of the IBM Virtual Shared Disk group, or the IBM Virtual Shared Disk in question is in the **stopped** state.

ENOMEM Indicates, in the case of a query request, that the current IBM Virtual Shared Disk fence map is larger than the one specified in the query request. The query fails and the *nvsds* field in the request is changed to reflect the current size of the fence map.

ENOTREADY Indicates that the group is not active or IBM Recoverable Virtual Shared Disk is not available.

ENXIO Indicates that the IBM Virtual Shared Disk driver is being unloaded.

Examples

The following is an example of how to fence an IBM Virtual Shared Disk with a minor number of 7 from node 4 and 5, and to unfence an IBM Virtual Shared Disk with a minor number of 5 from node 1:

```
int fd;
vsd_FenceMap_t FenceMap;

/* two VSDs to work with */
FenceMap.Map=(vsd_Fence *) malloc (2*sizeof(vsd_Fence));
FenceMap.nvsds = 2;

/* Clear the Map and Mask*/
MAP_ZERO(FenceMap.Map[0].node_FenceMap);
MAP_ZERO(FenceMap.Map[0].node_FenceMask);
MAP_ZERO(FenceMap.Map[1].node_FenceMap);
MAP_ZERO(FenceMap.Map[1].node_FenceMask);
FenceMap.flags = VSD_FENCE_SET | VSD_FENCE_FORCE | VSD_FENCE_SWAP;

/* fence nodes 4,5 from minor 7 */
FenceMap.Map[0].vsd_minor = 7;
MAP_SET(4, FenceMap.Map[0]node_FenceMap);
MAP_SET(4, FenceMap.Map[0].node_FenceMask);
MAP_SET(5, FenceMap.Map[0].node_FenceMap);
MAP_SET(5, FenceMap.Map[0].node_FenceMask)

/* Unfence node 1 from minor 5*/
FenceMap.Map[1].vsd_minor = 5;
MAP_SET(1, FenceMap.Map[1].node_FenceMask);

/* Issue the fence request */
ioctl(fd,GIOCFENCE,&FenceMap);
```

Appendix A. IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands

cfghsd

Purpose

cfghsd – Configures a Data Striping Device (HSD) for an IBM Virtual Shared Disk.

Syntax

cfghsd –**a** *hsd_name* ...

cfghsdvsd

Purpose

cfghsdvsd – Configures a Data Striping Device (HSD) and the IBM Virtual Shared Disks that comprise it on one node.

Syntax

cfghsdvsd –**a** | {*hsd_name*...}

cfgvsd

Purpose

cfgvsd – Configures an IBM Virtual Shared Disk.

Syntax

cfgvsd –**a** | *vsd_name* ...

cksumvsd

Purpose

cksumvsd – Views and manipulates the IBM Virtual Shared Disk checksum parameters.

Syntax

cksumvsd [–**s**] [–**R**] [–**i** | –**l**]

© Copyright IBM Corp. 1997, 1998

createhsd

Purpose

createhsd – Creates one data striping device (HSD) that encompasses two or more IBM Virtual Shared Disks.

Syntax

createhsd –**n** {*node_list* | **ALL**} –**s** *size_in_MB* –**g** *vg_name*
–**t** *stripe_size_in_KB* [{–**c** *vsds_per_node* | –**L**} [–**A**]]
[–**m** *mirror_cnt*] [–**d** *hsd_name*] [–**l** *lv_name_prefix*] [–**S**]
[–**o cache** | **nocache**] [–**T** *lp_size_in_MB*] [–**x**]

createvsd

Purpose

createvsd – Creates a set of IBM Virtual Shared Disks, with their associated logical volumes, and puts information about them into the System Data Repository (SDR).

Syntax

Note: Some examples shown in this list do not contain enough flags to be executable. They are shown in an incomplete form to illustrate specific flags.

createvsd –**n** {*node_list* | **ALL**}
–**s** *size_in_MB* –**g** *volume_group_name*
[{–**c** *number_of_vsds_per_node* | –**L**}]
[–**o cache** | **nocache**] [–**m** *mirror_count*]
[–**p** *lvm_stripe_size*] [–**v** *vsd_name_prefix*]
[–**l** *lv_name_prefix*] [–**T** *lp_size_in_MB*] [–**x**]

ctlhsd

Purpose

ctlhsd – Sets the data striping device (HSD) for the IBM Virtual Shared Disks operational parameters or reset statistics.

Syntax

ctlhsd [–**p** *parallel_level* | –**v** *hsd_name* ... | –**C** | –**V**]

ctlvsd

Purpose

ctlvsd – Sets IBM Virtual Shared Disk operational parameters.

Syntax

ctlvsd [–**t**] | [–**T**] | [–**c** *NewCacheSize* | –**r** *node_number*... | –**R** | –**k** *node_number*... –**K** | –**p** *1—9* | –**M** *max_IP_msg_size* | –**v** *vsd_name* ... | –**C** | –**V**]

defhsd

Purpose

defhsd – Defines a data striping device (HSD).

Syntax

defhsd **protect_LVCB** | **not_protect_LVCB** *hsd_name stripe_size vsd_name...*

defvsd

Purpose

defvsd – Defines an IBM Virtual Shared Disk.

Syntax

defvsd *logical_volume_name global_group_name vsd_name* [**nocache** | **cache**]

fencevsd

Purpose

fencevsd – Prevents an application running on a node or group of nodes from accessing an IBM Virtual Shared Disk or group of IBM Virtual Shared Disks.

Syntax

fencevsd –**v** *vsd_name_list* –**n** *node_list*

ha_vsd

Purpose

ha_vsd – Starts and restarts the rvsd subsystem of IBM Recoverable Virtual Shared Disk (RVSD). This includes configuring IBM Virtual Shared Disks and data striping devices (HSDs) as well as starting the rvsd and hc daemons.

Syntax

ha_vsd [**reset**]

ha.vsd

Purpose

ha.vsd – Queries and controls the rvsd subsystem of IBM Recoverable Virtual Shared Disk.

Syntax

ha.vsd {**adapter_recovery**[**on** | **off**] | **debug**[**off**] | **mksrc** | **query** | **quorum** *n* | **qsrc** | **reset** | **reset_quorum** | **rmsrc** | **start** | **stop** | **trace** [**off**]}

hc.vsd

Purpose

hc.vsd – Queries and controls the hc subsystem of IBM Recoverable Virtual Shared Disk.

Syntax

hc.vsd {**CLIENT_PATH** *socket_path* | **debug** [**off**] | **mksrc** | **PING_DELAY** *delay_in_sec* | **query** | **qsrc** | **reset** | **rmsrc** | **SCRIPT_PATH** *de/activate_path* | **start** | **stop** | **trace [off]**}

hsdatalst

Purpose

hsdatalst – Displays Data Striping Device (HSD) data for the IBM Virtual Shared Disks from the System Data Repository (SDR).

Syntax

hsdatalst [**–G**]

hsdvts

Purpose

hsdvts – Verifies that a Data Striping Device (HSD) for the IBM Virtual Shared Disks has been correctly configured and works.

Syntax

hsdvts *hsd_name*

lshsd

Purpose

lshsd – Displays configured Data Striping Device (HSD) for the IBM Virtual Shared Disks and their characteristics.

Syntax

lshsd [–**l** | –**s**] [*hsd_name* ...]

lsvsd

Purpose

lsvsd – Displays configured IBM Virtual Shared Disks and their characteristics.

Syntax

lsvsd [–**l** | –**s**] [–**i**] [*vsd_name*...]

preparevsd

Purpose

preparevsd – Makes an IBM Virtual Shared Disk available.

Syntax

preparevsd –**a** | *vsd_name*...

removehsd

Purpose

removehsd – Removes one or more data striping devices, the IBM Virtual Shared Disks associated with them, and the System Data Repository (SDR) information for IBM Virtual Shared Disks on the associated nodes. If the only volume groups defined on the associated nodes are the ones used by the data striping devices that are being removed, those volume groups are also removed.

Syntax

removehsd {–**v** {*hsd_names* | –**a**}} [–**f**]

removevsd

Purpose

removevsd – Removes a set of IBM Virtual Shared Disks that are not part of any data striping device (HSD).

Syntax

removevsd –**v** *vsd_names* | –**a** [–**f**]

resumevsd

Purpose

resumevsd – Activates an available IBM Virtual Shared Disk.

Syntax

resumevsd [–**p** | –**b**] –**a** | *vsd_name* ...

spvsd

Purpose

spvsd – Directly launches the IBM Virtual Shared Disk window of the SP Perspectives graphical user interface (GUI).

Syntax

spvsd [–**userProfile** *name*] [–**systemProfile** *name*] [–**noProfile**]
[{–**backgroundColor** | **bg**} *colorName*]
[{–**foregroundColor** | –**fg**} *colorName*] [–**fontFamily** *name*]
[–**fontSize** *size*] [–**fontBold**] [–**fontItalic**] [–**nosplash**]

---

startvsd

Purpose

startvsd – Makes an IBM Virtual Shared Disk available and activates it.

Syntax

startvsd [–p | –**b**] –**a** | *vsd_name* ...

---

statvsd

Purpose

statvsd – Displays IBM Virtual Shared Disk device driver statistics of a node.

Syntax

statvsd

---

stopvsd

Purpose

stopvsd – Makes an IBM Virtual Shared Disk unavailable.

Syntax

stopvsd –**a** | *vsd_name* ...

---

suspendvsd

Purpose

suspendvsd – Deactivates an available IBM Virtual Shared Disk.

Syntax

suspendvsd –**a** | *vsd_name*...

ucfghsd

Purpose

ucfghsd – Makes a Data Striping Device (HSD) for the IBM Virtual Shared Disks unavailable.

Syntax

ucfghsd –**a** | *hsd_name*...

ucfghsdvsd

Purpose

ucfghsdvsd – Makes a Data Striping Device (HSD) and the IBM Virtual Shared Disks that comprise it unavailable on one node.

Syntax

ucfghsdvsd –**a** | {*hsd_name*...}

ucfgvsd

Purpose

ucfgvsd – Makes an IBM Virtual Shared Disk unavailable.

Syntax

ucfgvsd –**a** | *vsd_name* ...

undefhsd

Purpose

undefhsd – Undefines a data striping device (HSD).

Syntax

undefhsd *hsd_name*...

undefvsd

Purpose

undefvsd – Undefines an IBM Virtual Shared Disk.

Syntax

undefvsd *vsd_name* ...

unfencevsd

Purpose

unfencevsd – Gives applications running on a node or group of nodes access to an IBM Virtual Shared Disk or group of IBM Virtual Shared Disks that were previously fenced from applications running on those nodes.

Syntax

unfencevsd [–**v**] *vsd_name_list* {–**n** *node_list* [–**f**] | –**r**}

updatehsd

Purpose

updatehsd – Lets you change the option in the System Data Repository (SDR) that prevents overwriting the Logical Volume Control Block (LVCB) for specified data striping devices (HSDs).

Syntax

updatehsd {–**d** *hsd_names* | –**a}** [–**f**]
–**o** {**protect_lvcb** | **not_protect_lvcb**}

updatevsdnode

Purpose

updatevsdnode – Changes IBM Virtual Shared Disk options in the System Data Repository (SDR).

Syntax

updatevsdnode –**n** {**ALL** | *node_number*[,*node_number* ...]}
{[–**a** {*VSD_adapter*] | **none**}]
[–**i** *init_cache_buffer_count*]
[–**m** *max_cache_buffer_count*] [–**r** *vsd_request_count*]
[–**p** *rw_request_count*] [–**b** *min_buddy_buffer_size*]
[–**x** *max_buddy_buffer_size*] [–**s** *max_buddy_buffers*]
[–**M** *vsd_max_ip_packet_size*]}
[–**f**]

updatevsdtab

Purpose

updatevsdtab – Changes the IBM Virtual Shared Disk **cache**/**nocache** option in the System Data Repository (SDR).

Syntax

updatevsdtab {–**v** *vsd_names* | –**a**} {[–**o** {**cache** | **nocache**}] [–**s**]} [–**f**]

verparvsd

Purpose

verparvsd – Verifies IBM Virtual Shared Disk system partitioning.

Syntax

verparvsd [–**F**] [–**o** *output_file*] *layout_directory* [*new_partition* ...]

vsdatalst

Purpose

vsdatalst – Displays IBM Virtual Shared Disk system definition data from the System Data Repository (SDR).

Syntax

vsdatalst [–**G**] –**g** | –**n** | –**v**

vsdchgserver

Purpose

vsdchgserver – Switches the server function for one or more IBM Virtual Shared Disks from the node that is currently acting as the server node to the other node.

Syntax

vsdchgserver –**g** *vsd_global_volume_group_name* –**p** *primary_node*
–**s** *secondary_node* [–**o** *EIO_recovery*]

vsddiag

Purpose

vsddiag – Displays information about the status of IBM Virtual Shared Disks.

Syntax

vsddiag

vsdelnode

Purpose

vsdelnode – Removes IBM Virtual Shared Disk information for a node or series of nodes from the System Data Repository (SDR).

Syntax

vsdelnode *node_number* ...

vsdelvg

Purpose

vsdelvg – Removes IBM Virtual Shared Disk global volume group information from the System Data Repository (SDR).

Syntax

vsdelvg [–**f**] *global_group_name* ...

vsdnode

Purpose

vsdnode – Enters IBM Virtual Shared Disk information for a node or series of nodes into the System Data Repository (SDR).

Syntax

vsdnode *node_number... adapter_name init_cache_buffer_count max_cache_buffer_count vsd_request_count rw_request_count min_buddy_buffer_size max_buddy_buffer_size max_buddy_buffers vsd_max_ip_msg_size*

vsdsklst

Purpose

vsdsklst – Produces output that shows you the disk resources used by the IBM Virtual Shared Disk subsystem across a system or system partition.

Syntax

vsdsklst {–**d** | –**v** | –**dv**} {[–**a**] | –**n** *node_list*]} [–**G**]

vsdvg

Purpose

vsdvg – Defines an IBM Virtual Shared Disk global volume group.

Syntax

vsdvg [–**g** *global_group_name*] *local_group_name primary_server_node*
[*secondary_server_node*]

vsdvgts

Purpose

vsdvgts – Reads the timestamp from the volume group descriptor area (VGDA) of the physical disks and sets the value in the System Data Repository (SDR).

Syntax

vsdvgts [–**a**] [*volgrp*]

vsdvts

Purpose

vsdvts – Verifies that the IBM Virtual Shared Disk software works.

Syntax

vsdvts [–**b** *block_size*] [–**n** *number_of_blocks*] *vsd_name* [*file*]

Appendix B. Single-Node Command and SMIT Interfaces for IBM Virtual Shared Disks, Data Striping Devices and the IBM Recoverable Virtual Shared Disk Licensed Product

With Version 2 Release 2 of the SP system, commands that operated on multiple nodes at once, such as **createvsd** and **createhsd**, have been introduced to improve the usability of these components. However, some users may have scripts that use single-node commands, such as **defvsd** and **defhsd**. This appendix contains the previous version of the information on defining and managing IBM Virtual Shared Disks, Data Striping Devices, and the IBM Recoverable Virtual Shared Disk software.

Command and SMIT Interfaces for IBM Virtual Shared Disks

Entering Node Information in the SDR for One Node

You can use SMIT or the **vsdnode** command to enter IBM Virtual Shared Disk node information into the SDR.

To enter IBM Virtual Shared Disk node information using SMIT:

SELECT VSD Node Information

▲ The VSD Node Information dialog appears.

For the syntax of the **vsdnode** command, see Appendix A, “IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands” on page 87.

Entering Global Volume Group Information in the SDR for One Node

Global volume group information is stored in the System Data Repository (SDR) **VSD_Global_Volume_Group** object. For each volume group on which IBM Virtual Shared Disks will be defined, the following data is required:

1. Local volume group name

   The LVM volume group's name on the serving node. The length of the name must be less than or equal to 15 characters.

2. Primary server node for the volume group

   The node number of the server node. The primary server node may be identified in four ways. See the **vsdvg** command reference page for specific information.

   If you do not have the IBM Recoverable Virtual Shared Disk product installed, all server nodes are primary nodes.

3. Secondary server node for the volume group

   The secondary server node is only intended for use with the IBM Recoverable Virtual Shared Disk product.

   The node number of the serving node. The secondary server node may be identified in four ways. See the **vsdvg** command reference page for specific information.

© Copyright IBM Corp. 1997, 1998

4. Global volume group name (must be unique across a system partition)

   The global volume group name is the globally unique name for this IBM Virtual Shared Disk volume group. The global volume group name is usually identical to the volume group name. The length of the name must be less than or equal to 31 characters. We recommend that the global volume group names be unique across the physical SP system. It *must* be unique within a system partition.

To enter IBM Virtual Shared Disk global volume group information using SMIT:

SELECT VSD Global Volume Group Information

▲ The VSD Global Volume Group Database Information dialog appears.

You can also enter IBM Virtual Shared Disk volume group information using the **vsdvg** command:

vsdvg [**-g** *global_group_name*] *local_group_name*
primary_server_node [*secondary_server_node*]

Entering IBM Virtual Shared Disk Information in the SDR for One Node

An IBM Virtual Shared Disk is defined by creating a **VSD_Table** object in the SDR. For each IBM Virtual Shared Disk, the following data is required:

1. IBM Virtual Shared Disk name (this name must be unique across the system partition)

   The globally unique name of the IBM Virtual Shared Disk. The name must be less than or equal to 31 characters. We recommend that the name be unique across the physical SP system.

   If you choose a *vsd_name* that is already the name of another device, the **cfgvsd** command will reject the name to insure that the special device files created for the name do not overlay and destroy files of the same name representing another device type (such as a logical volume).

2. Logical volume name

   The name of the logical volume on the serving node. The length of the name must be less than or equal to 15 characters.

3. Global volume group name

   The globally unique name of the volume group on which the IBM Virtual Shared Disk resides. This field is used to access the **VSD_Global_Volume_Group** to determine the node and the volume group that contains the underlying logical volume. The length of the name must be less than or equal to 31 characters.

4. Cache or nocache option

   Use the cache option only if your application does I/O in 4K blocks aligned on 4K disk boundaries, and issues a read immediately following a write.

   Nocache is the default for new IBM Virtual Shared Disks created with the **defvsd** command.

5. Minor number

   This number is automatically allocated.

To enter IBM Virtual Shared Disk information using SMIT:

SELECT Define a Virtual Shared Disk

▲ The Define A Virtual Shared Disk dialog appears.

You can also enter IBM Virtual Shared Disk information using the **defvsd** command:

defvsd *logical_volume_name global_group_name vsd_name [nocache | cache]*

An Example of Defining an IBM Virtual Shared Disk on One Node at a Time

Figure 13 is a simple example of an IBM Virtual Shared Disk. There are three IBM Virtual Shared Disks, one on each node. We have two physical hard disks per node and we defined a logical volume on one disk per node.

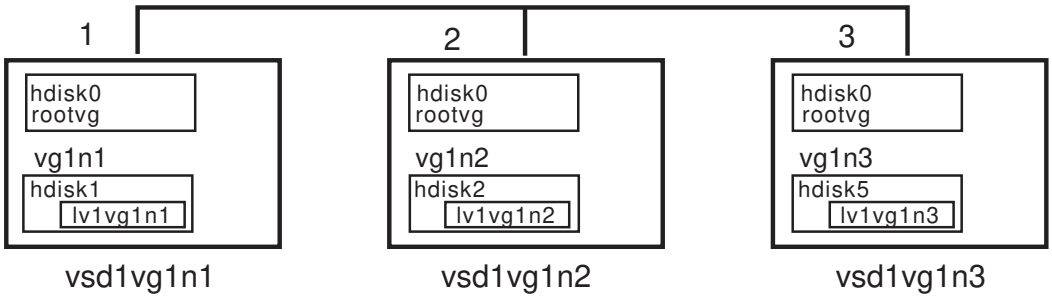


Figure 13. An Example of an IBM Virtual Shared Disk Configuration

Here's how we did it. First, we decided that three IBM Virtual Shared Disks were needed, one on each node on a separate volume group. We created a table of the information for our configuration.

Table 6. IBM Virtual Shared Disk Information

Node Number	hdisks	Volume Groups	Logical Volumes	vsd Name
1	hdisk0			
	hdisk1	vg1n1	lv1vg1n1	vsd1vg1n1
2	hdisk0			
	hdisk2	vg1n2	lv1vg1n2	vsd1vg1n2
3	hdisk0			
	hdisk5	vg1n3	lv1vg1n3	vsdvg1n3

Create Volume Groups and Logical Volumes on One Node

You must create the volume groups and logical volumes before defining the IBM Virtual Shared Disk information. We created the volume groups (**vg1n1**, **vg1n2**, **vg1n3**) on each node for each IBM Virtual Shared Disk using SMIT:

smit mkvg

and entered the volume group information requested. We took the defaults. See *AIX System Management Guide: Operating Systems and Devices*, for more information.

Next we created the logical volume groups (**lv1vg1n1**, **lv1vg1n2**, **lv1vg1n3**) on each node using SMIT:

```
smit mklv
```

and entered the global volume group information. See *AIX System Management Guide: Operating Systems and Devices*, for more information on tuning logical volumes.

Next we used SMIT to check the volume group and logical volume information we created:

```
smit lsvg
```

smit lslv

Define IBM Virtual Shared Disks

Now we start to define our IBM Virtual Shared Disks. We entered the IBM Virtual Shared Disk node information using SMIT. This information was entered separately for node **1**, **2**, and **3**. See "Displaying and Entering Node Information in the SDR" on page 26 for more information.

```
smit vsd_data
```

and selected the VSD Node Information option.

In the table *italics* show the information we entered. A `constant width` font shows the default data. We took most of the defaults except where noted.

Table 7. IBM Virtual Shared Disk Node Information

Node Number	Adapter	Initial Cache	Max Cache	Request Blocks	pbufs	Min Buddy Buffer	Max Buddy Buffer	# Buddy Buffers
1	*css0*	64	256	256	48	4096	24576	4
2	*css0*	64	256	256	48	4096	24576	4
3	*css0*	64	256	256	48	4096	24576	4

Next we used SMIT to enter the IBM Virtual Shared Disk global volume group information for **vg1n1**, **vg1n2** and **vg1n3**. See "Entering Global Volume Group Information in the SDR for One Node" on page 99 for more information.

```
smit vsd_data
```

and selected the VSD Global Volume Information option.

Table 8. IBM Virtual Shared Disk Volume Group Information

Global Group Name	Local VG Name	Primary Server	Secondary Server
vg1n1	*vg1n1*	*1*	–
vg1n2	*vg1n2*	*2*	–
vg1n3	*vg1n3*	*3*	–

We did not enter information about secondary servers because that is only for IBM Recoverable Virtual Shared Disk users.

Next we entered the IBM Virtual Shared Disk definition information for **vsd1vg1n1**, **vsd1vg1n2** and **vsd1vg1n3**. See "Entering IBM Virtual Shared Disk Information in the SDR for One Node" on page 100 for more information.

Again we used the SMIT command:

```
smit vsd_data
```

and selected the Define a Virtual Shared Disk option.

Table 9. IBM Virtual Shared Disk Definition Information

LV Name	Global Group	Name	Option
lv1vg1n1	*vg1n1*	*vsd1vg1n1*	nocache
lv1vg1n2	*vg1n2*	*vsd1vg1n2*	nocache
lv1vg1n3	*vg1n3*	*vsd1vg1n3*	nocache

Now we have defined our IBM Virtual Shared Disk in the SDR.

- We checked all the data base information using SMIT:

  ```
  smit
  list_vsd
  ```

- We did the following steps next to get our system running. They are described in "Managing IBM Virtual Shared Disks" on page 105.

  1. Configuring on "Configuring an IBM Virtual Shared Disk" on page 107.
  2. Starting on "Starting an IBM Virtual Shared Disk" on page 107.

Displaying IBM Virtual Shared Disk Information Stored in the SDR

Once you have defined IBM Virtual Shared Disk information in the SDR, you can display this information.

To list IBM Virtual Shared Disk definition information using SMIT (from the SP Configuration Database Management menu):

SELECT List Database Information

▲ The List Database Information menu appears.

SELECT List VSD Database Information

▲ The List VSD Database Information menu appears.

Or you can use the fastpath invocation for this menu:

```
smit list_vsd
```

At this point, you can select options for listing node, global volume group, and defined IBM Virtual Shared Disk information.

You can also list IBM Virtual Shared Disk information using the **vsdsklst** command with the appropriate flags to indicate what kind of information you want to display. The following example shows the format of the information returned by **vsdsklst** for one node from the `vsdsklst -dv -a` command.

```
k7n12.ppd.pok.ibm.com
Node Number:12; Node Name:k7n12.ppd.pok.ibm.com
    Volume group:rootvg; Partition Size:4; Total:537; Free:315
        Physical Disk:hdisk0; Total:537; Free:315
    Volume group:vsdvg; Partition Size:4; Total:537; Free:533
        Physical Disk:hdisk1; Total:537; Free:533
        VSD Name:1HsD8n12{lv1HsD8n12}; Size:2
        VSD Name:1HsD20n12{lv1HsD20n12}; Size:2
```

Listing Node Information

To view IBM Virtual Shared Disk information using SMIT:

SELECT List VSD Node Information

▲ A listing of VSD node information appears.

You can also display IBM Virtual Shared Disk node information using the **vsdatalst** command:

```
vsdatalst -n
```

Listing Global Volume Group Information

To view IBM Virtual Shared Disk global volume group information using SMIT:

SELECT List VSD Global Volume Group Information

▲ A listing of VSD global volume group information appears.

You can also display IBM Virtual Shared Disk node information using the **vsdatalst** command:

```
vsdatalst -g
```

Listing Defined IBM Virtual Shared Disks

To view defined IBM Virtual Shared Disk information using SMIT:

SELECT List Defined Virtual Shared Disks

▲ A listing of defined IBM Virtual Shared Disks appears.

You can also display defined IBM Virtual Shared Disk information using the **vsdatalst** command:

```
vsdatalst -v
```

Deleting IBM Virtual Shared Disk Information from the SDR

You can delete the IBM Virtual Shared Disk information you have stored in the SDR. Once you have defined IBM Virtual Shared Disk information in the SDR, you may want to delete or change (delete and add) information. To delete IBM Virtual Shared Disk definition information using SMIT (from the SP Configuration Database Management menu):

SELECT Delete Database Information

▲ The Delete Database Information menu appears.

SELECT Delete VSD Database Information

▲ The Delete VSD Database Information menu appears.

The fastpath invocation for this menu is:

```
smit delete_vsd
```

At this point, you can select options for deleting node and global volume group information and for undefining a IBM Virtual Shared Disk.

You can also use the **removevsd** or **undefvsd** commands to remove an IBM Virtual Shared Disk, a list of IBM Virtual Shared Disks, or all the IBM Virtual Shared Disks in a system or system partition. See the command reference pages for **removevsd** and **undefvsd** for more information.

Managing IBM Virtual Shared Disks

Once IBM Virtual Shared Disks have been defined in the SDR, they can be configured to the system and managed to various states by using SMIT menus or commands.

Running any of the following IBM Virtual Shared Disk commands affects the IBM Virtual Shared Disk states only on the node on which the command is run:

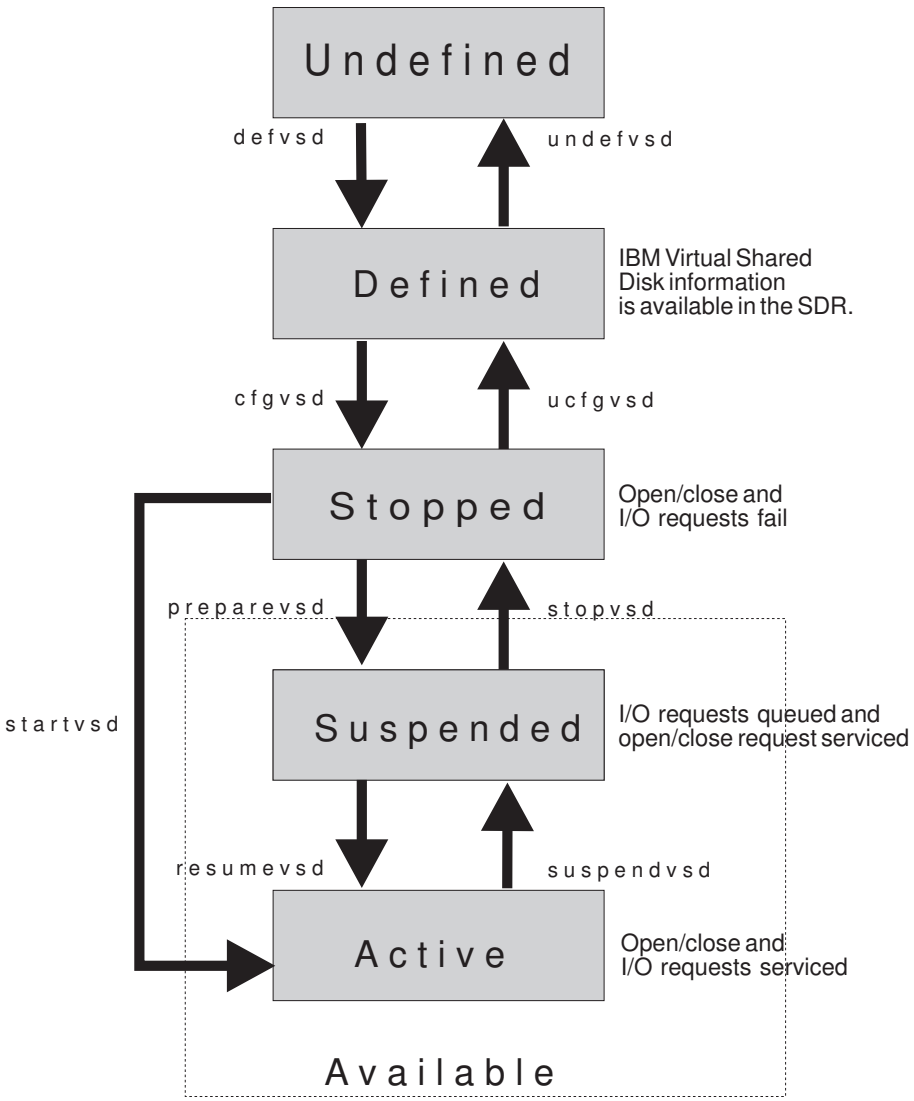
- **varyonvg**
- **cfgvsd**
- **startvsd**
- **preparevsd**
- **resumevsd**
- **suspendvsd**
- **stopvsd**
- **varyoffvg**
- **ucfgvsd**

The SMIT panels use **dsh** to run the command on any number of nodes you select. The SMIT menus may be invoked from the control workstation. See Figure 14 on page 106 for an overview of IBM Virtual Shared Disk states and associated commands.

In order for an application to be able to read or write **vsd.***nnn*, **vsd.***nnn* must be in the active state on both the client and server. Both **cfgvsd** and **startvsd** must be run on **vsd.***nnn*. Once **vsd.***nnn* is in the active state, the application may open /**dev**/**rvsd.***nnn*, providing the permissions on the file permit access. For example, to allow an application on node **1** to access /**dev**/**rvsd1vg1n1**, **cfgvsd** and **startvsd** must have been run on node **1** for **vsd1vg1n1**. To allow access to **vsd1vg1n1** from node **2**, **cfgvsd** and **startvsd** must also be run on node **2**. You may use **cfgvsdhsd** on all nodes instead of **cfgvsd** on each node.

Figure 14 on page 106 summarizes IBM Virtual Shared Disk states, how I/O requests are treated in each state, and the commands (sometimes called methods) used to change states. New users should study Figure 14 on page 106 to understand IBM Virtual Shared Disk operation.

If you use the IBM Recoverable Virtual Shared Disk, do not issue any commands that change the state of the IBM Virtual Shared Disks (**cfgvsd**, **statvsd**, **stopvsd**, or **ucfgvsd**.) The IBM Recoverable Virtual Shared Disk issues these commands for you.



Note: Methods on arrows cause transitions

Figure 14. IBM Virtual Shared Disk States and Associated Commands

To manage IBM Virtual Shared Disks using SMIT:

TYPE **smit**

▲ The System Management menu appears

SELECT SP System Management

▲ The SP System Management menu appears.

SELECT SP Cluster Management

▲ The SP Cluster Management menu appears.

SELECT VSD Management

▲ The VSD Management menu appears.

The fastpath invocation for this menu is:

```
smit vsd_mgmt
```

Note: If you are not using the IBM Recoverable Virtual Shared Disk product and you want to bring up IBM Virtual Shared Disks automatically at boot time, all **cfgvsd** and **startvsd** (or **preparevsd** and **resumevsd**) commands can be run at boot time on all IBM Virtual Shared Disk nodes. Put these commands in a script that is invoked by either **/etc/init.rc** or **/etc/inittab**. This approach allows all IBM Virtual Shared Disk configuration to occur in parallel across the system partition as each machine boots up. You must ensure that each global volume group is varied on (using the **varyonvg** command) on the server node before running **startvsd** or **resumevsd** on the server node. If you subsequently install the IBM Recoverable Virtual Shared Disk product, make sure to undo these steps.

Configuring an IBM Virtual Shared Disk

Configuring puts a defined IBM Virtual Shared Disk in the stopped state, but does not make it available. The IBM Virtual Shared Disk configuration method (**cfgvsd**) issues **/usr/lpp/csd/bin/readSDR** to extract information from the SDR and puts it into the flat files **VSD_Global_Volume_Group**, **VSD_Table**, **VSD_ipaddr** and **Node** in the **/usr/lpp/csd/vsdfiles** directory. Other IBM Virtual Shared Disk commands access these files.

Note: Do not alter these files. Each **cfgvsd** command invocation recreates the files from the SDR information and any changes you make to the files will be lost. To make changes, update the SDR information.

Use **cfgvsd** with **dsh** to configure IBM Virtual Shared Disks on more than one node at a time.

To configure an IBM Virtual Shared Disk using SMIT (from the VSD Management menu):

SELECT Configure a VSD

▲ The VSD and node selection lists appear.

You can also configure an IBM Virtual Shared Disk using the **cfgvsd** command.

To configure all IBM Virtual Shared Disks defined in the SDR to the node you are on, enter:

```
cfgvsd -a
```

To configure specific IBM Virtual Shared Disks defined in the SDR, enter:

cfgvsd *vsd_name* **...**

Starting an IBM Virtual Shared Disk

Starting an IBM Virtual Shared Disk puts a stopped one in the active (and available) state. (This is equivalent to preparing and resuming an IBM Virtual Shared Disk.) Note that for an IBM Virtual Shared Disk to be usable, it must be in the active state on both the server and client nodes.

To start an IBM Virtual Shared Disk using SMIT (from the VSD Management menu):

SELECT Start a VSD

▲ The VSD and node selection lists appear.

You can also start an IBM Virtual Shared Disk using the **startvsd** command.

To start all IBM Virtual Shared Disks defined in the SDR to the node you are on, enter:

```
startvsd -a
```

To start specific IBM Virtual Shared Disks defined in the SDR, enter:

startvsd *vsd_name* **...**

Preparing an IBM Virtual Shared Disk

Preparing an IBM Virtual Shared Disk puts a stopped one in the suspended state. In the suspend state, open and close requests are honored. Read and write requests are held until the IBM Virtual Shared Disk is brought to the active state.

To prepare an IBM Virtual Shared Disk using SMIT (from the VSD Management menu):

SELECT Prepare a VSD

▲ The VSD and node selection lists appear.

You can also prepare an IBM Virtual Shared Disk using the **preparevsd** command.

To prepare all IBM Virtual Shared Disks in the SDR, enter:

```
preparevsd -a
```

To prepare specific IBM Virtual Shared Disks defined in the SDR, enter:

preparevsd *vsd_name* **...**

Resuming an IBM Virtual Shared Disk

Resuming an IBM Virtual Shared Disk puts a suspended one in the active state. The IBM Virtual Shared Disk remains available and read and write requests that were held are resumed.

To resume an IBM Virtual Shared Disk using SMIT (from the VSD Management menu):

SELECT Resume a VSD

▲ The VSD and node selection lists appear.

You can also resume an IBM Virtual Shared Disk using the **resumevsd** command.

To resume all IBM Virtual Shared Disks defined in the SDR, enter:

```
resumevsd -a
```

To resume a specific IBM Virtual Shared Disk defined in the SDR, enter:

resumevsd *vsd_name* **...**

Suspending an IBM Virtual Shared Disk

Suspending an IBM Virtual Shared Disk puts an active one in the suspended state. The IBM Virtual Shared Disk remains available and read and write requests that were active are suspended and held. All read and write requests subsequent to those that were active are also held.

To suspend an IBM Virtual Shared Disk using SMIT (from the VSD Management menu):

SELECT Suspend a VSD

▲ The VSD and node selection lists appear.

You can also suspend an IBM Virtual Shared Disk using the **suspendvsd** command.

To suspend all IBM Virtual Shared Disks defined in the SDR, enter:

```
suspendvsd -a
```

To suspend specific IBM Virtual Shared Disks defined in the SDR, enter:

```
suspendvsd vsd_name ...
```

Stopping an IBM Virtual Shared Disk

Stopping an IBM Virtual Shared Disk puts a suspended one in the stopped state. The IBM Virtual Shared Disk becomes unavailable. All applications that have outstanding requests to a stopped IBM Virtual Shared Disk terminate in error.

To stop an IBM Virtual Shared Disk using SMIT (from the VSD Management menu):

SELECT Stop a VSD

▲ The VSD and node selection lists appear.

You can also stop an IBM Virtual Shared Disk using the **stopvsd** command.

To stop all IBM Virtual Shared Disks defined in the SDR, enter:

```
stopvsd -a
```

To stop specific IBM Virtual Shared Disks defined in the SDR, enter:

```
stopvsd vsd_name ...
```

Unconfiguring an IBM Virtual Shared Disk

Unconfiguring a stopped IBM Virtual Shared Disk makes it inaccessible. It does not, however, undefine or change the definition information for the IBM Virtual Shared Disk in the SDR.

To unconfigure an IBM Virtual Shared Disk using SMIT (from the VSD Management menu):

SELECT Unconfigure a VSD

▲ The VSD and node selection lists appear.

You can also unconfigure an IBM Virtual Shared Disk using the **ucfgvsd** command.

To unconfigure all IBM Virtual Shared Disks defined in the SDR, enter:

```
ucfgvsd -a
```

To unconfigure specific IBM Virtual Shared Disks in the SDR, enter:

```
ucfgvsd vsd_name ...
```

Displaying Managed IBM Virtual Shared Disk Characteristics

Configuration information for an IBM Virtual Shared Disk includes:

- Name
- Minor number
- State
- Current server node number
- *cache*|*nocache* option
- At the server, the major and minor number of the underlying logical volume. At other nodes, this information is not available.

To display managed IBM Virtual Shared Disk characteristics for all IBM Virtual Shared Disks using SMIT (from the VSD Management menu):

SELECT Show All Managed VSD Characteristics

▲ The VSD and node selection lists appear.

You can also display managed IBM Virtual Shared Disk characteristics using the **lsvsd** command. To display configuration information for specific IBM Virtual Shared Disks in your system, enter:

```
lsvsd -l vsd_name ...
```

Displaying Managed IBM Virtual Shared Disk Statistics

Usage information includes:

- Number of local logical read and write operations
- Number of remote logical read and write operations
- Number of client logical read and write operations
- Number of physical reads and writes
- Number of cache hits for read
- Number of 512KB blocks read and written

To display managed IBM Virtual Shared Disk usage statistics for all IBM Virtual Shared Disks using SMIT (from the VSD Management menu):

SELECT Show All Managed VSD Statistics

▲ The VSD and node selection lists appear.

You can also display managed IBM Virtual Shared Disk usage statistics using the **lsvsd** command. To display usage information for IBM Virtual Shared Disks in your system, enter:

```
lsvsd -s vsd_name ...
```

Note: The number of blocks read and written is cumulative. To reset this count before you measure it, issue **ctlvsd -V** command .

Displaying IBM Virtual Shared Disk Device Driver Statistics

You may wish to display usage information about the IBM Virtual Shared Disk device driver. Usage information includes:

- IBM Virtual Shared Disk parallelism value
- Nonzero sequence numbers and associated nodes
- IBM Virtual Shared Disk device driver counters
- Node outcast status

- IP message size
- Timeouts
- Retries
- Cache shortages
- Buddy buffer shortages
- Indirect I/O
- Communications buffer pool shortage

To display usage information about the IBM Virtual Shared Disk device driver using SMIT (from the VSD Management menu):

SELECT Show VSD Device Driver Statistics

▲ A node selection list appears.

You can also display IBM Virtual Shared Disk device driver usage statistics using the **statvsd** command:

```
statvsd
```

See Appendix A, “IBM Virtual Shared Disk and IBM Recoverable Virtual Shared Disk Commands” on page 87 for a description of these statistics.

Setting and Displaying IBM Virtual Shared Disk Device Driver Operational Parameters

You may wish to display and, in some cases, set IBM Virtual Shared Disk device driver operational parameters. Operational parameters include:

- Current cache buffer count
- Maximum cache buffer count
- Request block count
- pbuf count
- Level of IBM Virtual Shared Disk parallelism
- Buddy buffer configuration

To display and set IBM Virtual Shared Disk device driver operational parameters using SMIT (from the VSD Management menu):

SELECT Set/Show VSD Device Driver Operational Parameters

▲Show VSD Device Driver Operational Parameters

▲Reset Sequence Numbers

▲Increase Cache Size

▲Set Level of VSD Parallelism

▲Reset VSD Device Driver Counters

▲Reset a VSD's Statistics

▲Reset All VSDs' Statistics

You can also display and set IBM Virtual Shared Disk device driver operational parameters using the **ctlvsd** command.

Testing to Verify an IBM Virtual Shared Disk (the Verification Test Suite)

Warning: The **vsdvts** command writes data to the IBM Virtual Shared Disk. **Do not use the command after you have put real data on a IBM Virtual Shared Disk.** Refer to the **vsdvts** command reference page.

To test that you have successfully installed the IBM Virtual Shared Disk software and that you can successfully define and make an IBM Virtual Shared Disk active, and read and write to it, follow these steps:

1. Make sure the switch or other network adapter is up and running.
2. Create a logical volume. One physical partition will be enough space for it.
3. Define the logical volume as an IBM Virtual Shared Disk in the SDR. Then, in the following commands, substitute the name of the IBM Virtual Shared Disk you defined for *vsd.name*. Commands are in /**usr**/**lpp**/**csd**/**bin**.
4. Configure the IBM Virtual Shared Disk to your system. Enter:

   cfgvsd *vsd_name*
5. Make the IBM Virtual Shared Disk available. Enter:

   preparevsd *vsd_name*
6. Make the IBM Virtual Shared Disk active. Enter:

   resumevsd *vsd_name*
7. Display configuration information about the IBM Virtual Shared Disk. Enter:

   lsvsd -l *vsd_name*
8. Confirm that the IBM Virtual Shared Disk Verification Test is successful. Enter:

   vsdvts *vsd_name*

   The command should complete within 15 seconds and you should get the message:

   ```
   VSD verification test successful!
   ```

You can use the **vsdvts** command to verify all of your IBM Virtual Shared Disks. Use it to verify both local and remote IBM Virtual Shared Disks. Run the **cfgvsd** and **startvsd** commands on all client and server nodes. Use the SMIT panels to verify that the panels work.

Command and SMIT Interfaces for Data Striping Devices

Defining HSD Device Driver Information in the SDR

The HSD must be defined before you can configure and use it. The data striping device information is stored in the System Data Repository (SDR). You can use either SMIT or a command to enter the information in the SDR.

Creating HSDs with SMIT

The SMIT panels refer to the data striping devices as “hsd”

To define the Data Striping Device information in the SDR using SMIT:

SELECT SP System Management

▲ The SP System Management menu appears.

SELECT SP Configuration Database Management

▲ The Configuration Database Management appears.

SELECT Enter Database information

▲ The Enter Database information menu appears.

SELECT IBM Virtual Shared Disk Database information

▲ The IBM Virtual Shared Disk Database information menu appears

SELECT Define a Hashed Shared Disk

▲ The Define A Hashed Shared Disk dialog appears.

The fastpath invocation for this menu is:

smit vsd_data

An Example of Creating an HSD

Example 1

We will define our HSD, called **hsd1**, with four virtual shared disks at the same node and a stripe size of 4096 bytes. The IBM Virtual Shared Disks are named vsd1vg1n1, vsd2vg1n1, vsd3vg1n1, and vsd4vg1n1.

The HSDs were defined:

```
defhsd hsd1 4096 vsd1vg1n1 vsd2vg1n1 vsd3vg1n1 vsd4vg1n1
```

We have 16384 bytes (four 4KB blocks) of data to write. The first 4096 bytes are stored in vsd1vg1n1. The second 4096 bytes are stored in vsd2vg1n1, the third 4096 bytes are stored in vsd3vg1n1, and the last 4096 bytes are stored in vsd4vg1n1.

Example 2

Assume now we have two striped devices defined, **hsd1** and **hsd2**. Our stripe size is 4096.

We had identified the IBM Virtual Shared Disks to be included in the Data Striping Devices. Assume there are four nodes in the system and two physical hard disks per node. We wanted to stripe across all the four nodes and all the disks. We defined a logical volume on the disks and nodes. Then we defined the IBM Virtual Shared Disks:

hsd name	node	IBM Virtual Shared Disks in the HSD
hsd1	node1	vsd1vg1n1
	node2	vsd1vg1n2
	node3	vsd1vg1n3
	node4	vsd1vg1n4
hsd2	node1	vsd2vg1n1
	node2	vsd2vg1n2
	node3	vsd2vg1n3
	node4	vsd2vg1n4

We defined the HSDs with the following commands:

```
defhsd hsd1 4096 vsd1vg1n1 vsd1vg1n2 vsd1vg1n3 vsd1vg1n4
defhsd hsd2 4096 vsd2vg1n1 vsd2vg1n2 vsd2vg1n3 vsd2vg1n4
```

Displaying Data Striping Device Information in the SDR

To list HSD definition information in the SDR do one of the following:

Using SMIT

From the SP Configuration Database Management menu:

SELECT List Database Information

▲ The List Database Information menu appears.

SELECT List HSD Database Information

▲ The List IBM Virtual Shared Disk Database Information menu appears.

SELECT List Defined Hashed Shared Disks

▲ A listing of defined Hashed Shared Disks appears.

You can use the fastpath invocation for this menu :

```
smit list_vsd
```

At this point, you can select options for listing defined HSD information.

To view defined HSD information using SMIT:

SELECT List Defined Hashed Shared Disks

▲ A listing of defined Hashed Shared Disks appears.

From the Command Line

You can also display defined HSDs information using the **hsdatalst** command:

```
hsdatalst
```

Undefining HSD Information in the SDR

To delete Data Striping Device definition information

Using SMIT

From the SP Configuration Database Management menu:

SELECT Delete Database Information

▲ The Delete Database Information menu appears.

SELECT Delete IBM Virtual Shared Disk Database Information

▲ The Delete IBM Virtual Shared Disk Database Information menu appears.

The fastpath invocation for this menu is:

```
smit delete_vsd
```

At this point, you can select options for undefining an HSD.

To undefine an HSD using SMIT:

SELECT Undefine a Hashed Shared Disk

▲ The Undefine a Hashed Shared Disk selection list appears.

You can either type in the *hsd_name* or select an *hsd_name* from the list function. The *hsd_name* must be in the unconfigured state.

From the Command Line

You can also undefine a Data Striping Device using the **removehsd** or **undefhsd** commands:

undefhsd -v *hsd_names* [**–f**]

removehsd -v *hsd_names*
[**–f**]

If the *hsd_names* are configured, you can use the –**f** parameter of **removehsd** to force the system to unconfigure and remove them.

Managing Data Striping Devices

To manage Data Striping Devices using SMIT:

TYPE **smit**

▲ The System Management menu appears

SELECT SP System Management

▲ The SP System Management menu appears.

SELECT SP Cluster Management

▲ The SP Cluster Management menu appears.

SELECT HSD Management

▲ The HSD Management menu appears.

The fastpath invocation for this menu is:

```
smit hsd_mgmt
```

Configuring an HSD

Configuration makes a defined HSD available.

Using SMIT: From the HSD Management menu,

SELECT Configure an HSD

▲ The HSD and node selection lists appear.

From the Command Line

To configure specific HSDs defined in the SDR on one node, enter:

cfghsd *hsd_name* **...**

Unconfiguring an HSD

Unconfiguring an HSD makes it unavailable. It does not, however, undefine or change the definition information for the HSD in the SDR. You must do this on each node.

Using SMIT: From the HSD Management menu,

SELECT Unconfigure an HSD

▲ The HSD and node selection lists appear.

From the Command Line

To unconfigure specific HSDs in the SDR on one node, enter:

ucfghsd *hsd_name* **...**

Displaying HSD Information at a Node

You can display managed IBM Virtual Shared Disk Data Striping Devicecharacteristics for all configured Data Striping Devices.

Using SMIT: From the HSD Management menu,

SELECT Show All Managed HSD Characteristics

▲ The HSD and node selection lists appear.

From the Command Line: You can also display managed Data Striping Device characteristics using the **lshsd** command. To display information for a specific Data Striping Device in your system, enter:

lshsd -l [*hsd_name* **...**]

Configuration information displayed for an HSD includes:

Name
Minor number
Stripe size
protect_lvcb / **not_protect_lvcb** option
Underlying IBM Virtual Shared Disks in the HSD

You can display managed IBM Virtual Shared Disk Data Striping Device usage statistics for all configured Data Striping Devices..

Using SMIT: From the HSD Management menu,

SELECT Show All Managed HSD Characteristics

▲ The HSD and node selection lists appear.

From the Command Line: You can also display managed Data Striping Device usage statistics using the **lshsd** command. To display usage information for a specific Data Striping Device in your system, enter:

lshsd -s [*hsd_name* **...**]

Usage information displayed for an HSD includes:

- Number of reads on each underlying IBM Virtual Shared Disk
- Number of writes on each underlying IBM Virtual Shared Disk

Setting and Displaying IBM Virtual Shared Disk HSD Operational Parameters

You can display and, in some cases, set IBM Virtual Shared Disk Data Striping Device parallelism level, as well as reset the device counters and statistics. Operational parameters include:

- Current level of Data Striping Device parallelism
- Number of read requests not at page boundary count
- Number of write requests not at page boundary count

Using SMIT: From the HSD Management menu,

SELECT Set/Show HSD Device Driver Operational Parameters

▲ Show HSD Device Driver Operational Parameters

▲ Set Level of HSD Parallelism

▲ Reset HSD Device Driver Counters

▲ Reset an HSD's Statistics

▲ Reset All HSDs' Statistics

You can also display and set IBM Virtual Shared Disk Data Striping Device operational parameters using the **ctlhsd** command. Refer to the *IBM Parallel System Support Programs for AIX: Command and Technical Reference* for more information.

Verifying Data Striping Device Installation

After you install the HSD software and define and configure several HSDs, you can use the verification test. Do this **before** you store data on these devices.

CAUTION:
The hsdvts command writes data to the configured *hsd_name***. Do not use the command after you have put real data on that** *hsd_name* **or the underlying IBM Virtual Shared Disks.**

Refer to the **hsdvts** command reference page. Follow these steps:

1. Enter:

   hsdvts *hsd.name*

   You should get the message:

   ```
   HSD verification test successful!
   ```

You can use the **hsdvts** command to verify all of your HSDs. Use it to verify both local and remote HSDs. Issue the **cfghsd** command on all client and server nodes.

Management and Usage Notes

Configuring an HSD

You can define, configure, unconfigure, and undefine HSDs, using the **defhsd**, **cfghsd**, **ucfghsd**, and **undefhsd** commands.

Configuring an HSD creates the special device file /**dev**/**r***hsd_name*, which loads the HSD device driver into the kernel and makes the device available for use by user applications.

Use **cfghsdvsd** and **ucfghsdvsd** to configure and unconfigure HSDs and their underlying IBM Virtual Shared Disks on multiple nodes at once.

Unconfiguring an HSD makes the device unavailable to applications.

Defining an HSD creates an HSD object in the SDR.

Undefining an HSD deletes the HSD object from the SDR and removes the special device file /**dev**/*hsd_name*.

As a device driver, the Data Striping Device provides interfaces to application programs through the **open**, **close**, **read**, **write**, and related system calls. A **read** or **write** first uses the offset in the data file to figure out which IBM Virtual Shared Disk the request is for. The HSD then passes the request to the IBM Virtual Shared Disk to do the I/O. Usually the **ioctl** call is device-dependent. The **ioctl** call for the HSD provides minimum information about the device.

The Data Striping Device is a character (raw) device which must be opened as /**dev**/**r***hsd_name*; all the data addresses must be at 512 block boundaries.

Command and SMIT Interfaces for the IBM Recoverable Virtual Shared Disk Licensed Program

This section tells how to maintain twin-tailed Logical Volume Manager (LVM) components shared by two nodes in an SP system running the IBM Recoverable Virtual Shared Disk software. It includes procedures for volume groups, logical volumes, and physical volumes.

Defining Twin-tailed LVM Components in the IBM Recoverable Virtual Shared Disk Environment

Some of the definitions you need to understand are:

twin-tailed volume group A volume group consisting of supported disks that are all twin-tailed to the same nodes.

twin-tailed physical volume
A disk in a twin-tailed volume group.

twin-tailed logical volume
A logical volume in a twin-tailed volume group.

Coordinating the IBM Recoverable Virtual Shared Disk Environment

The twin-tailed components must have the same definition on both twin-tailed nodes. Any change to an LVM component **must** be reflected in the Object Data Manager (ODM) definitions on both nodes.

Follow the procedures in this chapter carefully in the order given to keep the LVM ODM definitions on both nodes synchronized.

Note: Do not reboot or issue **ha_vsd reset** until you have completed all the steps in a task.

Overview of Shared Volume Groups Administrative Tasks

The overall procedure for modifying a twin-tailed LVM component is the same for all tasks. In general, you export the volume group from a secondary node, make the change on the primary node, and then reimport the volume group on the secondary node. Specific operations, however, have unique steps. They are described in the following sections.

The table below summarizes the steps you must complete on the primary node and the secondary node to change a twin-tailed LVM component in an IBM Recoverable Virtual Shared Disk system. Perform all the steps in the correct order so the data does not become corrupted.

Table 10. General Procedure for Changing a Twin-tailed LVM Component

Step Number	Primary Node	Secondary Node
Step 1	Complete prerequisite tasks	Complete prerequisite tasks
Step 2		Export a volume group
Step 3	Vary on a volume group	
Step 4	Make changes to the twin-tailed LVM component	
Step 5	Vary off the volume group	
Step 6		Import a volume group
Step 7		Change a volume group to remain dormant at start up
Step 8		Vary off the volume group
Step 9	Complete follow-up tasks	Complete follow-up tasks

Prerequisite Tasks

The prerequisite tasks, while not directly involved in modifying LVM components, must be completed before you begin to make the change.

Note: You must suspend, stop, and unconfigure all IBM Virtual Shared Disks on all nodes in the volume group to be changed before changing a twin-tailed LVM component.

The prerequisite tasks can vary for the different operations. The descriptions for each operation on the following pages have a list of specific prerequisite tasks.

Maintaining Twin-Tailed Volume Groups

Maintaining twin-tailed volume groups requires the following administrative tasks:

- "Creating a Twin-Tailed Volume Group"
- "Adding Physical Volumes To a Twin-Tailed Volume Group" on page 124
- "Adding or Removing Logical Volumes to or from a Twin-Tailed Volume Group" on page 128
- "Reducing a Twin-Tailed Volume Group" on page 130
- "Removing a Twin-Tailed Volume Group" on page 136

Creating a Twin-Tailed Volume Group

The table below summarizes the steps you must complete on the primary node and the secondary node to create a twin-tailed volume group. Perform all the steps in the correct order so the data does not become corrupted.

Note: Twin-tailed disks that have SCSI IDs must have a different ID on the secondary than the one on the primary.

Table 11. Steps to Create a Twin-tailed Volume Group

Step Number	Primary Node	Secondary Node
Step 1	Complete prerequisite tasks	Complete prerequisite tasks
Step 2	Create a twin-tailed volume group	
Step 3	Create logical volumes	
Step 4	Vary off the volume group	
Step 5		Import a volume group
Step 6		Change a volume group to remain dormant at startup
Step 7		Vary off the volume group
Step 8	Complete follow-up tasks	Complete follow-up tasks

Complete the following steps to create a twin-tailed volume group.

Step 1: Complete Prerequisite Tasks

1. The physical volumes (**hdisks**) should be installed, configured, named, and available.
2. Make sure the primary and the secondary nodes are active.

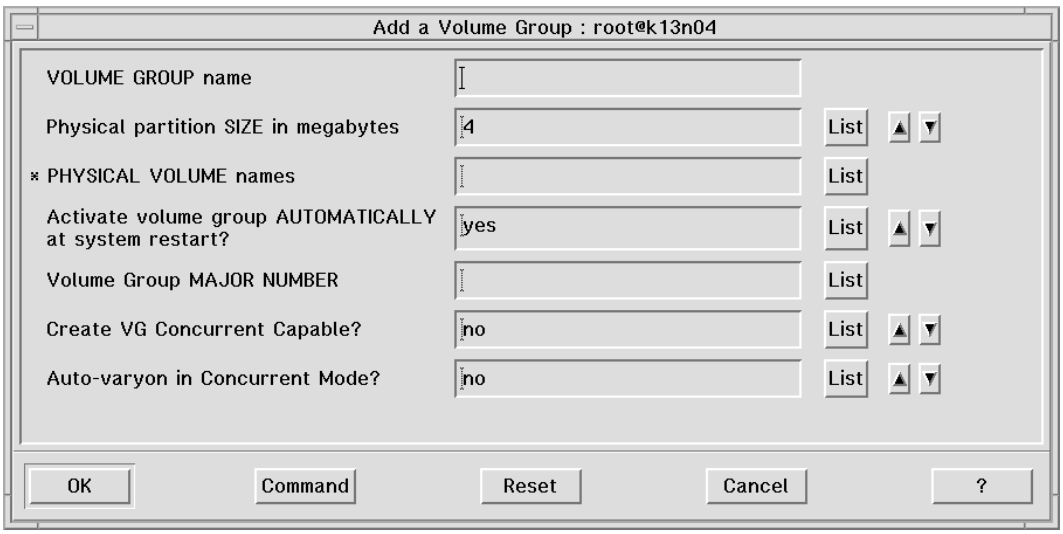
Step 2: Create a Twin-tailed Volume Group on the Primary Node

Use the **smit mkvg** fastpath to create a volume group.

1. As the root user at the primary node, enter:

   ```
   smit mkvg
   ```

   SMIT returns a screen similar to the following:



2. Enter field values as follows:

   VOLUME GROUP name Enter the name of the twin-tailed volume group.

   Physical partition size in megabytes
   Accept the default value, unless your site has another specific partitioning requirement.

   PHYSICAL VOLUME names
   Enter the names of all the disks to be used in the volume group. You can click on List to display all available physical volumes. . Click on OK when you have selected all the entries you want. The physical volumes you selected are automatically entered in the PHYSICAL VOLUME names field.

   Activate volume group AUTOMATICALLY at system restart?
   Set the field to **no** so that the volume group can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts.

   ACTIVATE volume group after it is created?
   Set this field to **yes**.

   Volume Group MAJOR NUMBER
   Use the default, the next available number in the valid range.

   Create VG Concurrent Capable?
   Accept the default of **no**. Do not specify **yes**.

   Auto–varyon in Concurrent Mode?
   Accept the default of **no**. Do not specify **yes**.

3. Click on OK. The system asks if you are sure. Check and correct your entries, if needed.
4. Create logical volumes. Refer to the LVM reference pages.
5. After the command completes, press F12 to exit SMIT and return to the command line.

Step 3: Create Logical Volumes

Use the **smit mklv** fastpath to create logical volumes. Create the logical volumes you want for this IBM Virtual Shared Disk.

Step 4: Vary Off the Volume Group on the Primary Node

Use the **varyoffvg** command to quiesce the affected volume group.

To vary off the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts, enter:

varyoffvg *volume_group_name*

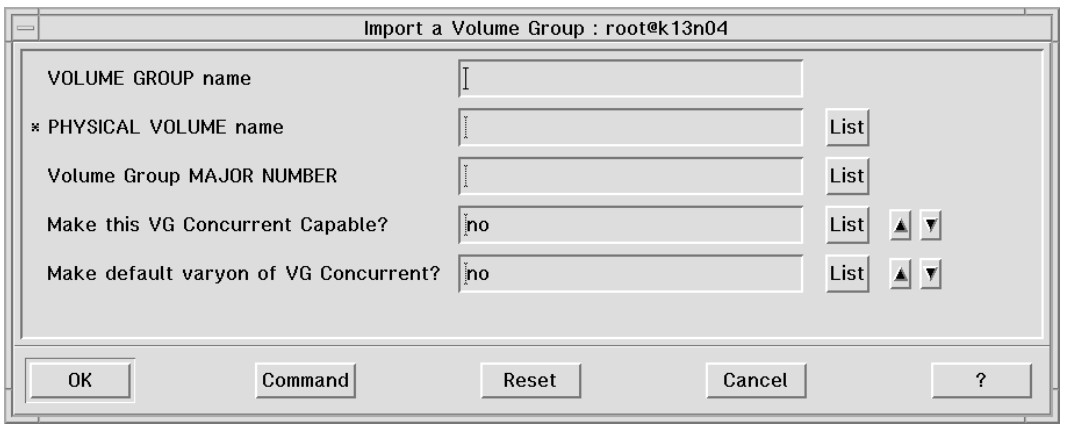
Step 5: Import Volume Group Information onto the Secondary Node

You now return to the secondary node. First, import the volume group onto the secondary node. Importing the volume group onto the secondary node synchronizes the ODM definition of the volume group on both nodes.

1. To use the **smit importvg** fastpath to import the volume group, enter:

   smit importvg

   SMIT returns a screen similar to the following:



2. Enter field values as follows:

   VOLUME GROUP name Enter the name of the volume group that you are importing. Make sure the volume group name is the same name that you used on the primary node.

   PHYSICAL VOLUME name
   Enter the name of one of the physical volumes that resides in the volume group. Note that a disk can have a different physical name on different nodes. Make sure that you use the disk name as it is defined on the *secondary* node. Every **hdisk** has an ID. If you know that the name of a disk on the primary node is *hdisk01*, use the name that matches that ID on the secondary node. You can use the **lspv** command to determine the name.

   ACTIVATE volume group after it is imported?
   Set the field to **yes**.

Volume Group MAJOR NUMBER
Use the default, the next available number in the valid range.

Make this VG Concurrent Capable?
Accept the default of **no**. Do not specify **yes**.

Make default varyon of VG Concurrent?
Accept the default of **no**. Do not specify **yes**.

3. Click on OK.
4. Press F12 to exit SMIT and return to the command line.

Step 6: Change a Volume Group to Remain Dormant at Start Up

By default, a volume group that was just imported is configured to automatically become active at system restart. However, an IBM Recoverable Virtual Shared Disk volume group should be varied on as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts. Therefore, after importing a volume group, use the SMIT Change a Volume Group screen to reconfigure the volume group to remain dormant at start up.

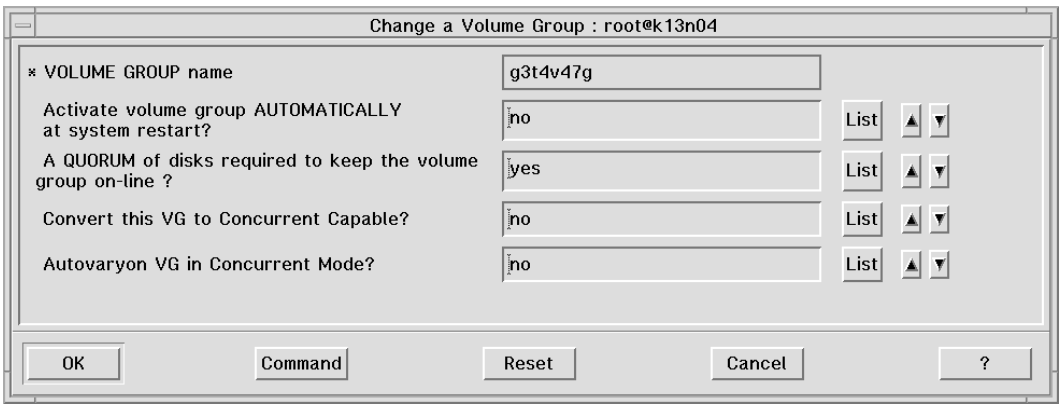
1. To use the **smit chvg** fastpath to change the characteristics of a volume group, enter:

   ```
   smit chvg
   ```

   SMIT prompts you to select the volume group.

2. Enter the name of the volume group you just imported, or click on List and select the name from the display. Click on OK.

   SMIT returns a screen similar to the following: The first field contains the volume group name you specified.



3. Enter remaining field values as follows:

   ACTIVATE volume group automatically at system restart?
   Set this field to **no**.

   A QUORUM of disks required to keep the volume group on-line?
   Accept the default, **yes**.

   Convert this VG to Concurrent Capable?
   Accept the default **no**. Do not specify **yes**.

Autovaryon VG in Concurrent Mode?
Accept the default **no**. Do not specify **yes**.

4. Click on OK.
5. Press F12 to exit SMIT and return to the command line.

Step 7: Vary Off the Volume Group on the Secondary Node

Use the **varyoffvg** command to quiesce the affected volume group after making the change.

To vary off the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts, enter:

varyoffvg *volume_group_name*

Step 8: Complete Follow-up Tasks

Once you have created the volume group, do the following tasks:

- Make sure the primary and secondary nodes are active.
- Use the **vsdvg** command to define the twin-tailed volume groups.
- Use **defvsd** to define IBM Virtual Shared Disks on all the twin-tailed nodes.
- Use the **cfgvsd** command for all the IBM Virtual Shared Disks in this volume group at all the nodes.
- Use the **startvsd** command at all the IBM Virtual Shared Disk nodes to start these new IBM Virtual Shared Disks.

Adding Physical Volumes To a Twin-Tailed Volume Group

The following table summarizes the steps you do on both the primary and secondary nodes to extend (add one or more physical volumes to) a twin-tailed volume group. Perform all the steps in the correct order so the data does not become corrupted.

Table 12. Procedure for Extending a Twin-tailed LVM Component

Step Number	Primary Node	Secondary Node
Step 1	Complete prerequisite tasks	Complete prerequisite tasks
Step 2		Export a volume group
Step 3	Vary on a volume group	
Step 4	Extend a twin-tailed volume group	
Step 5	Vary off the volume group	
Step 6		Import a volume group
Step 7		Change a volume group's characteristics
Step 8		Vary off the volume group
Step 9	Complete follow-up tasks	Complete follow-up tasks

Step 1: Complete Prerequisite Tasks

1. The physical volumes (**hdisks**) should be installed, configured, named, and available.
2. Suspend, stop, and unconfigure all IBM Virtual Shared Disks in the volume group involved in the change on all nodes.

Step 2: Export Volume Group Information from the Secondary Node

Before making any changes to the LVM elements on the primary node, you must export the appropriate volume group from the secondary node. Exporting the volume group deletes the information about this volume group from the ODM.

To use the **exportvg** command to export the volume group containing the component you are going to change on the secondary node, enter:

exportvg *volume_group_name*

Step 3: Vary on a Volume Group on the Primary Node

To use the **varyonvg** command to activate the affected volume group after making the change, enter:

varyonvg *volume_group_name*

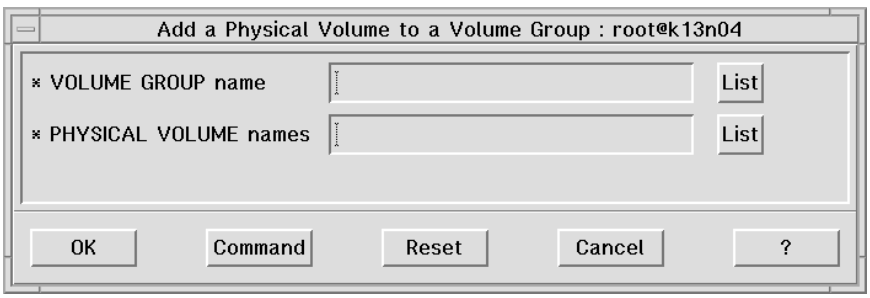
Step 4: Extend the Twin-Tailed Volume Group on the Primary Node

Use the **smit extendvg** fastpath to extend a volume group.

1. As the root user at the primary node, enter:

   smit extendvg

   SMIT returns a screen similar to the following:



2. Enter field values as follows:

   VOLUME GROUP name Enter the name of the volume group that you are extending.

   PHYSICAL VOLUME names
   Enter the names of the physical volumes you are adding to the volume group.

3. Click on OK to extend the volume group.
4. After the command completes, press F12 to exit SMIT and return to the command line.
5. Create new logical volumes or extend existing ones.

Step 5: Vary Off the Volume Group on the Primary Node

Use the **varyoffvg** command to quiesce the affected volume group after making the change.

To vary off the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts, enter:

varyoffvg *volume_group_name*

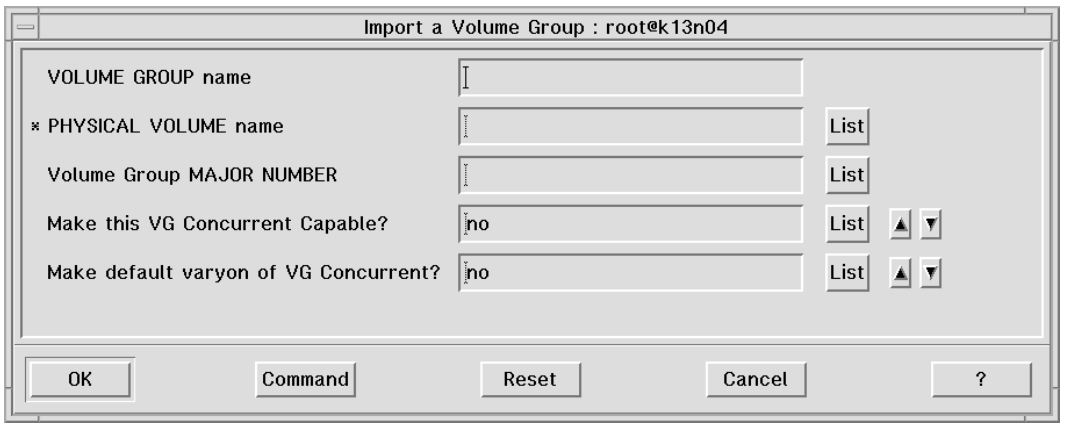
Step 6: Import Volume Group Information onto the Secondary Node

You now return to the secondary node. First, import the volume group onto the secondary node. Importing the volume group onto the secondary node synchronizes the ODM definition of the volume group with the primary node.

1. To use the **smit importvg** fastpath to import the volume group, enter:

   smit importvg

   SMIT returns a screen similar to the following:



2. Enter field values as follows:

   VOLUME GROUP name Enter the name of the volume group that you are importing. Make sure the volume group name is the same name that you used on the primary node.

   PHYSICAL VOLUME name
   Enter the name of one of the physical volumes that resides in the volume group. Note that a disk can have a different physical name on different nodes. Make sure that you use the disk name as it is defined on the *secondary* node. Every **hdisk** has an ID. If you know that the name of a disk on the primary node is *hdisk01*, use the name that matches that ID on the secondary node. You can use the **lspv** command to determine the name.

   Volume Group MAJOR NUMBER
   Use the default, the next available number in the valid range.

Make this VG Concurrent Capable?
Accept the default of **no**. Do not specify **yes**.

Make default varyon of VG Concurrent?
Accept the default of **no**. Do not specify **yes**.

3. Click on OK.
4. Press F12 to exit SMIT and return to the command line.

Step 7: Change Volume Group Characteristics on the Secondary Node

By default, a volume group that was just imported is configured to automatically become active at system restart. However, an IBM Recoverable Virtual Shared Disk volume group should be varied on as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts. Therefore, after importing a volume group, use the SMIT Change a Volume Group screen to reconfigure the volume group to remain dormant at start up.

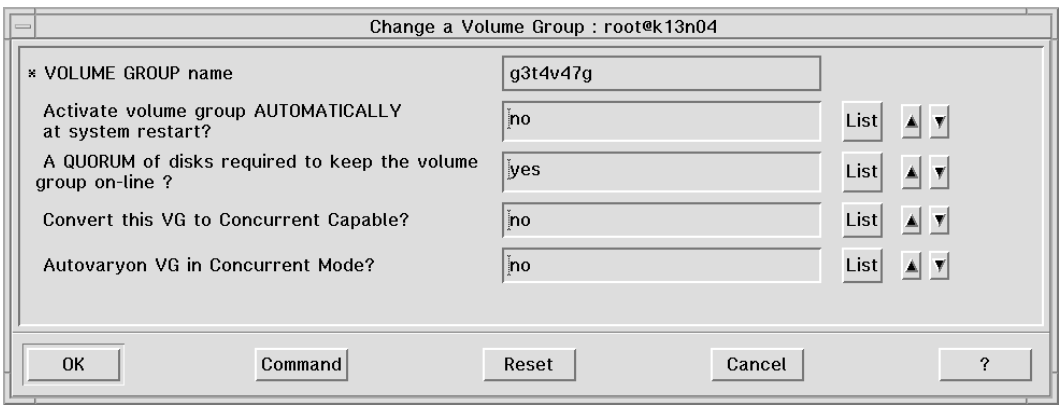
1. To use the **smit chvg** fastpath to change the characteristics of a volume group, enter:

   ```
   smit chvg
   ```

   SMIT prompts you to select the volume group.

2. Enter the name of the volume group you just imported, or click on List and select the name from the display. Click on OK.

   SMIT returns a screen similar to the following. The first field contains the volume group name you specified.



3. Enter remaining field values as follows:

   ACTIVATE volume group automatically at system restart?
   Set this field to **no**.

   A QUORUM of disks required to keep the volume group on-line?
   Accept the default, **yes**.

   Convert this VG to Concurrent Capable?
   Accept the default **no**. Do not specify **yes**.

   Autovaryon VG in Concurrent Mode?
   Accept the default **no**. Do not specify **yes**.

4. Click on OK.

5. Press F12 to exit SMIT and return to the command line.

Step 8: Vary Off the Volume Group on the Secondary Node

Use the **varyoffvg** command to quiesce the affected volume group after making the change.

To vary off the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts, enter:

varyoffvg *volume_group_name*

Step 9: Complete Follow-up Tasks

Once the you have extended the volume group, do the following tasks:

- Make sure the primary and secondary nodes are active.
- Use **defvsd** to define any new IBM Virtual Shared Disks.
- Use the **cfgvsd** command for all the IBM Virtual Shared Disks in this volume group at all the nodes.
- Use the **startvsd** command at all the IBM Virtual Shared Disk nodes to start these new IBM Virtual Shared Disks.

Adding or Removing Logical Volumes to or from a Twin-Tailed Volume Group

The following table summarizes the steps you do on both the primary and secondary nodes to add or remove logical volumes to or from a twin-tailed volume group. Perform all the steps in the correct order so the data does not become corrupted.

Table 13. Procedure for Modifying a Twin-tailed LVM Component

Step Number	Primary Node	Secondary Node
Step 1	Complete prerequisite tasks	Complete prerequisite tasks
Step 2		Export a volume group
Step 3	Create or remove logical volumes	
Step 4	Vary off the volume group	
Step 5		Import a volume group
Step 6		Change a volume group's characteristics
Step 7		Vary off the volume group
Step 8	Vary on the volume group	
Step 9	Complete follow-up tasks	Complete follow-up tasks

Step 1: Complete Prerequisite Tasks

1. Shutdown your applications using the volume group to ensure data consistency and ODM integrity.
2. Suspend, stop, and unconfigure all IBM Virtual Shared Disks on all the IBM Virtual Shared Disk nodes.

3. Undefine IBM Virtual Shared Disks on any logical volumes being removed. Ensure all data is moved from the logical volumes before removing.

Step 2: Export Volume Group Information from the Secondary Node

Before making any changes to the LVM elements on the primary node, you must export the appropriate volume group from the secondary node. Exporting the volume group deletes the information about this volume group from the ODM.

To use the **exportvg** command to export the volume group containing the component you are going to change on the secondary node, enter:

exportvg *volume_group_name*

Step 3: Create or Remove Required Logical Volumes on the Primary Node

To create the logical volumes, use the fastpath invocation to the SMIT panel:

smit mklv

Alternatively, you can use the command line interface to create the logical volumes:

mklv

To remove the logical volumes, use the fastpath invocation to the SMIT panel:

smit rmlv

Alternatively, you can use the command line interface to remove the logical volumes:

rmlv

Step 4: Vary Off the Volume Group on the Primary Node

Use the **varyoffvg** command to quiesce the affected volume group after making the change.

To vary off the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts, enter:

varyoffvg *volume_group_name*

Step 5: Import Volume Group Information onto the Secondary Node

You now return to the secondary node. First, import the volume group onto the secondary node. Importing the volume group onto the secondary node synchronizes the ODM definition of the volume group with the primary node.

1. To use the **smit importvg** fastpath to import the volume group, enter:

   smit importvg

Step 6: Change Volume Group Characteristics on the Secondary Node

By default, a volume group that was just imported is configured to automatically become active at system restart. However, an IBM Recoverable Virtual Shared Disk volume group should be varied on as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts. Therefore, after importing a volume group, use the

SMIT Change a Volume Group screen to reconfigure the volume group to remain dormant at start up.

To use the **smit chvg** fastpath to change the characteristics of a volume group, enter:

```
smit chvg
```

For more information, refer to “Step 7: Change Volume Group Characteristics on the Secondary Node” on page 127.

Alternatively, to use the command line interface to change the volume group to remain dormant at startup, enter:

chvg -a n' -Q y' *volume_group_name*

Step 7: Vary Off the Volume Group on the Secondary Node

Use the **varyoffvg** command to quiesce the affected volume group after making the change.

To vary off the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts, enter:

varyoffvg *volume_group_name*

Step 8: Vary On the Volume Group on the Primary Node

Use the **varyonvg** command to quiesce the affected volume group after making the change.

To vary on the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts, enter:

varyonvg *volume_group_name*

Step 9: Complete Follow-up Tasks

Once the you have extended the volume group, do the following tasks:

- Make sure the primary and secondary nodes are active.
- Use **defvsd** to define any new IBM Virtual Shared Disks.
- Use the **cfgvsd** command for all the IBM Virtual Shared Disks in this volume group at all the nodes.
- Use the **startvsd** command at all the IBM Virtual Shared Disk nodes to start these new IBM Virtual Shared Disks.

Reducing a Twin-Tailed Volume Group

The following table summarizes the steps you do on both the primary and secondary nodes to reduce (remove one or more physical volumes from) a twin-tailed volume group:

Table 14. Procedure for Reducing a Twin-Tailed Volume Group		
Step Number	**Primary Node**	**Secondary Node**
Step 1	Complete prerequisite tasks	Complete prerequisite tasks
Step 2		Export a volume group
Step 3	Vary on a volume group	
Step 4	Remove data from the physical volume	
Step 5	Reduce a twin-tailed volume group	
Step 6	Vary off the volume group	
Step 7		Import volume group
Step 8		Change volume group characteristics
Step 9		Vary off the volume group
Step 10	Complete follow-up tasks	Complete follow-up tasks

Complete the following tasks to reduce a twin-tailed volume group.

Step 1: Complete Prerequisite Tasks

1. Both the primary and secondary nodes must be up and in the active group.
2. Suspend, stop, and unconfigure IBM Virtual Shared Disks in the affected volume group on all nodes.

Step 2: Export Volume Group Information from the Secondary Node

Before making any changes to the LVM elements on the primary node, you must export the appropriate volume group from the secondary node. Exporting the volume group deletes the information about this volume group from the ODM.

To use the **exportvg** command to export the volume group containing the component you are going to change on the secondary node, enter:

exportvg *volume_group_name*

Step 3: Vary on a Volume Group on the Primary Node

To use the **varyonvg** command to verify that the affected volume group is varied on after making the change, enter:

varyonvg *volume_group_name*

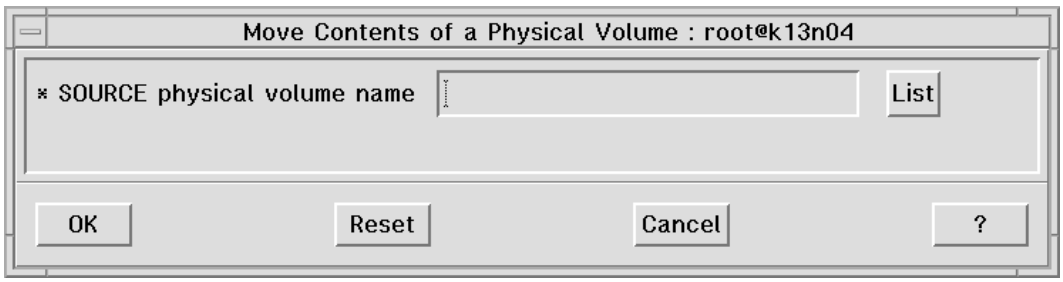
Step 4: Move Data from the Volume to Be Removed

Use the **smit migratepv** fastpath to move data on the physical volume being removed from the volume group to a different physical volume. If you do not, **data will be lost**.

1. As the root user on the primary node, enter:

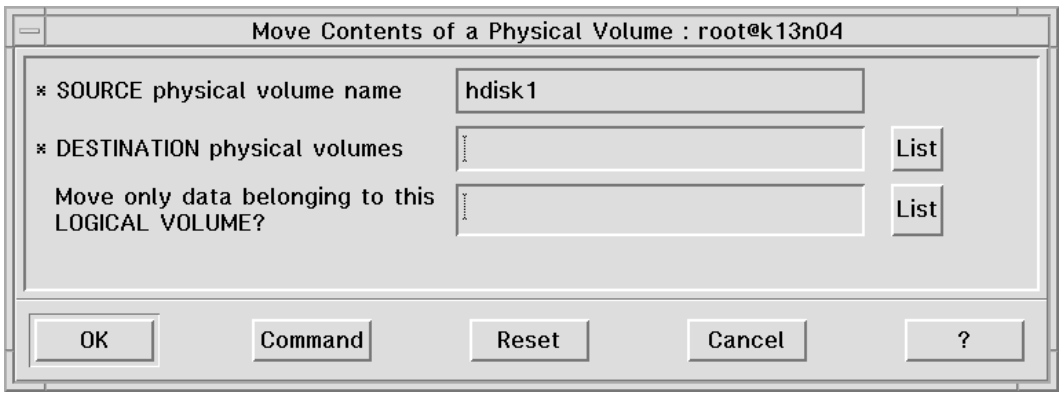
   smit migratepv

   SMIT returns a screen similar to the following:



2. Enter the names of the physical volumes being moved in the SOURCE physical volume name field or click on List and select the names from the display. Click on OK.

   SMIT returns a screen similar to the following. The physical volumes being migrated are entered in the SOURCE physical volume names field.



3. Enter field values as follows:

   DESTINATION physical volumes
   Enter the names of the physical volumes to which you want to move the data.

   Move only data belonging to this LOGICAL VOLUME?
   Set this field to **no**.

4. Click on OK to migrate the physical volumes.
5. After the command completes, press F12 to exit SMIT and return to the command line.

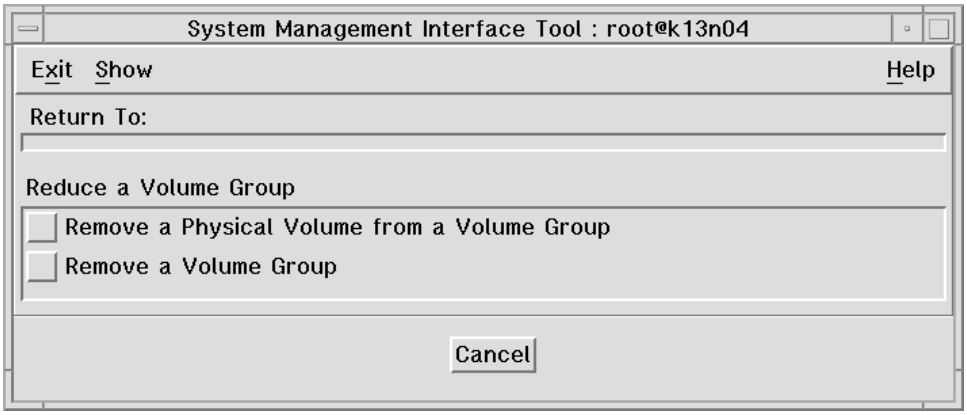
Step 5: Reduce the Shared Volume Group on the Primary Node

Use the **smit reducevg** fastpath to reduce a volume group.

1. As the root user at the primary node, enter:

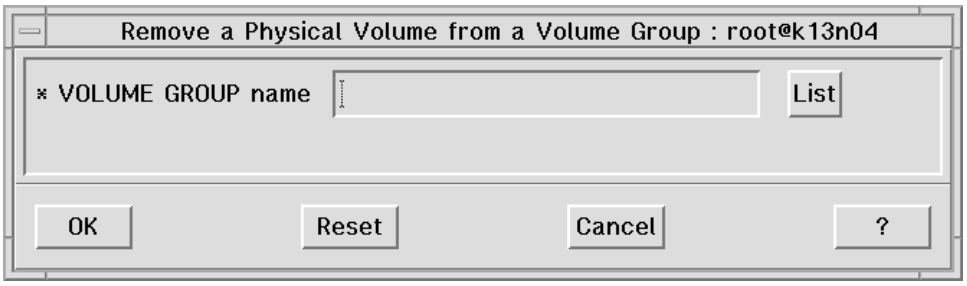
   ```
   smit reducevg
   ```

   SMIT returns a screen similar to the following:



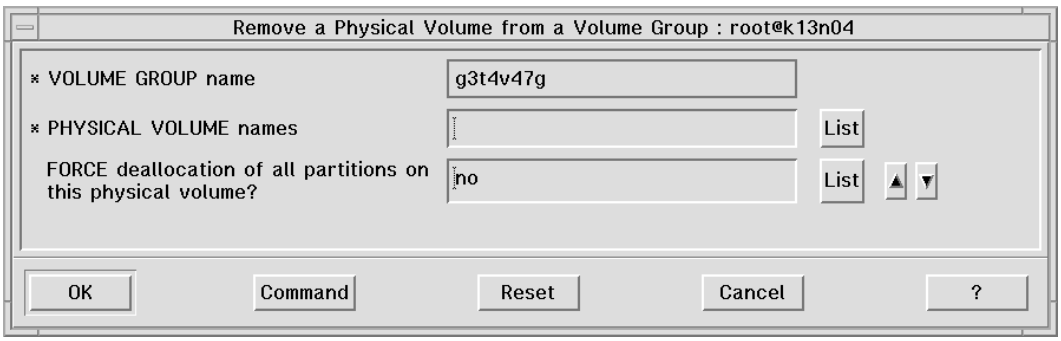
2. Select the **Remove a Physical Volume from a Volume Group** option and click on OK.

   SMIT returns a screen similar to the following:



3. Enter the name of the volume group from which you are removing the physical volumes in the VOLUME GROUP name field. Click on OK.

   SMIT returns a screen similar to the following. The name of the volume group you specified in the preceding screen is entered.



4. Enter field values as follows:

   PHYSICAL VOLUME names
   Enter the name of the physical volumes to remove from the volume group.

   FORCE deallocation of all partitions on this physical volume?
   Set this field to **no**.

5. Click on OK to remove the designated physical volumes from the volume group.
6. After the command completes, press F12 to exit SMIT and return to the command line.

Step 6: Vary Off the Volume Group on the Primary Node

Use the **varyoffvg** command to quiesce the affected volume group after making the change.

To vary off the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts, enter:

varyoffvg *volume_group_name*

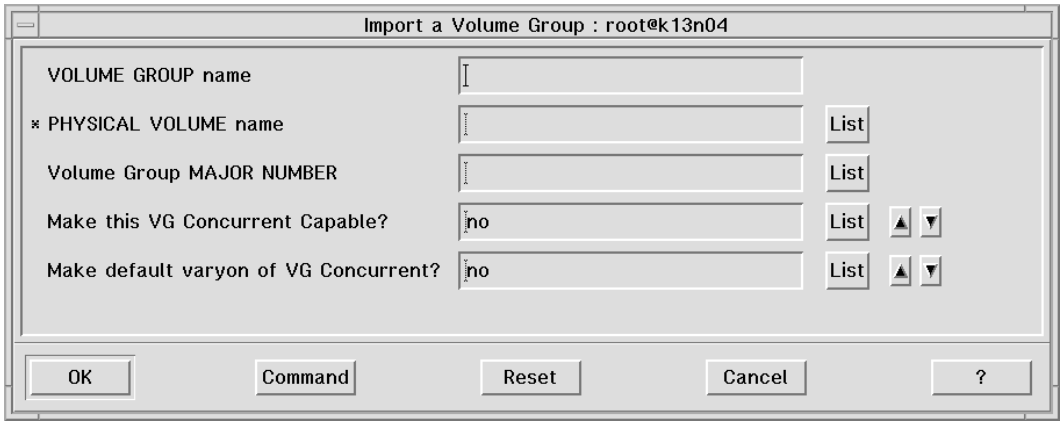
Step 7: Import Volume Group Information onto the Secondary Node

You now return to the secondary node. First, import the volume group on the secondary node. Importing the volume group onto the secondary node synchronizes the ODM definition of the volume group on the node.

1. To use the **smit importvg** fastpath to import the volume group, enter:

   smit importvg

   SMIT returns a screen similar to the following:



2. Enter field values as follows:

   VOLUME GROUP name Enter the name of the volume group that you are importing. Make sure the volume group name is the same name that you used on the primary node.

   PHYSICAL VOLUME name
   Enter the name of one of the physical volumes that resides in the volume group. Note that a disk can have a different physical name on different nodes. Make sure that you use the disk name as it is defined on the *secondary* node. Every **hdisk** has an ID. If you know that the name of a disk on the primary node is *hdisk01*, use the name that matches that ID on the secondary node. You can use the **lspv** command to determine the name.

Volume Group MAJOR NUMBER
Use the default, the next available number in the valid range.

Make this VG Concurrent Capable?
Accept the default of **no**. Do not specify **yes**.

Make default varyon of VG Concurrent?
Accept the default of **no**. Do not specify **yes**.

3. Click on OK.
4. Press F12 to exit SMIT and return to the command line.

Step 8: Change Volume Group Characteristics on the Secondary Node

By default, a volume group that was just imported is configured to automatically become active at system restart. However, an IBM Recoverable Virtual Shared Disk volume group should be varied on as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts. Therefore, after importing a volume group, use the SMIT Change a Volume Group screen to reconfigure the volume group to remain dormant at startup.

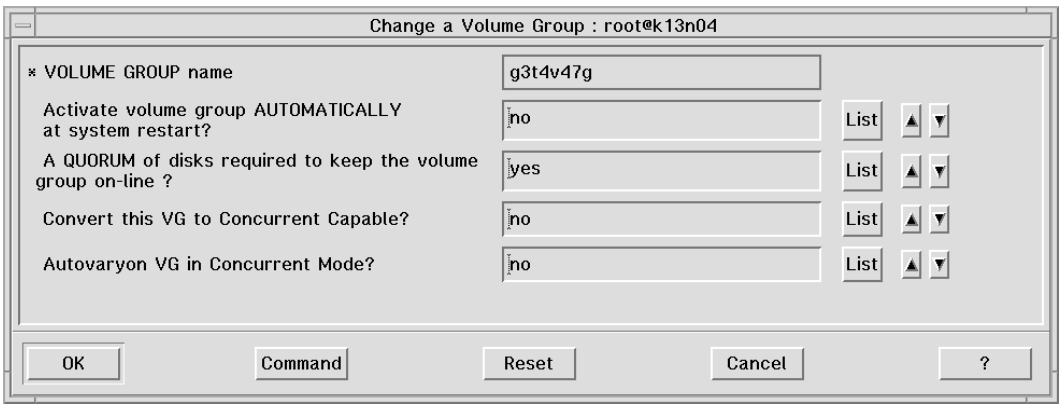
1. To use the **smit chvg** fastpath to change the characteristics of a volume group, enter:

   ```
   smit chvg
   ```

   SMIT prompts you to select the volume group.

2. To select the volume group you just imported, click on OK.

   A screen similar to the following appears. The first field contains the volume group name you specified.



3. Enter remaining field values as follows:

   ACTIVATE volume group automatically at system restart?
   Set this field to **no**.

   A QUORUM of disks required to keep the volume group on-line?
   Accept the default, **yes**.

   Convert this VG to Concurrent Capable?
   Accept the default **no**. Do not specify **yes**.

Autovaryon VG in Concurrent Mode?
Accept the default **no**. Do not specify **yes**.

4. Click on OK.
5. Press F12 to exit SMIT and return to the command line.

Step 9: Vary Off the Volume Group on the Secondary Node

Use the **varyoffvg** command to quiesce the affected volume group after making the change.

You vary off the volume group so that it can be activated as appropriate by the IBM Recoverable Virtual Shared Disk recovery scripts. Enter:

varyoffvg *volume_group_name*

Step 10: Complete Follow-up Tasks

Once the you have reduced the volume group, complete the following tasks:

- Make sure the primary and secondary nodes are active.
- Use **undefvsd** to undefine no longer required IBM Virtual Shared Disks.
- Use the **cfgvsd** command for all the IBM Virtual Shared Disks in this volume group at all the nodes.
- Use the **startvsd** command at all the IBM Virtual Shared Disk nodes to start these new IBM Virtual Shared Disks.

Removing a Twin-Tailed Volume Group

See the following table for an overview of the steps needed to remove a twin-tailed volume group.

Table 15. Steps to Remove a Shared Volume Group

Step Number	Primary Node	Secondary Node
Step 1	Complete prerequisite tasks	Complete prerequisite tasks
Step 2		Export volume group information
Step 3	Vary on a volume group	
Step 4		Delete a twin-tailed volume group

Complete the following steps to remove a twin-tailed volume group.

Step 1: Complete Prerequisite Tasks

1. Suspend, stop, and unconfigure the active IBM Virtual Shared Disks on this volume group on all nodes.
2. Undefine the volume group with **vsdelvg** at the control workstation.

Step 2: Export Volume Group Information from the Secondary Node

Before making any changes to the LVM elements on the primary node, you must export the appropriate volume group from the secondary node. Exporting the volume group deletes the information about this volume group from the ODM. To

use the **exportvg** command to export the volume group containing the component you are going to change on the secondary nodes, enter:

exportvg *volume_group_name*

Step 3: Vary On a Volume Group on the Primary Node

To use the **varyonvg** command to activate the affected volume group after making the change, enter:

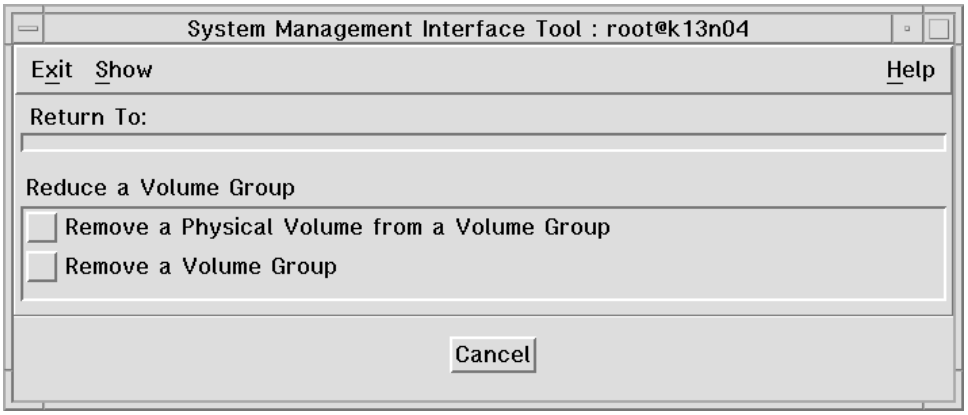
varyonvg *volume_group_name*

Step 4: Delete a Twin-tailed Volume Group on the Primary Node

1. As the root user at the primary node, enter:

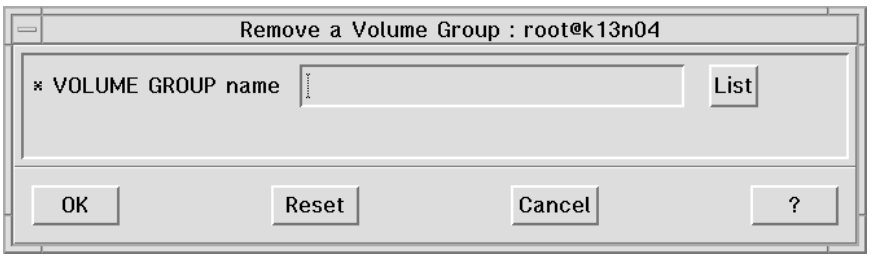
   smit reducevg

   SMIT returns a screen similar to the following:



2. Select the **Remove a Volume Group** option and click on OK.

   SMIT returns a screen similar to the following:



3. Enter the name of the volume group to be removed. Click on OK.
4. When the command completes, press 10 to leave SMIT and return to the command line.

Appendix C. Setting Up 7134 Disks as Twin-tailed Volume Groups

If you are using 7134 disks to set up twin-tailed volume groups, making and importing the volume groups is the same as on 9333 and 9334 disks, but there is some unique preparation work that needs to be done.

To install the 7134 disks as twin-tailed volume groups:

1. Connect the 7134 drawer to two nodes.
2. Reboot (or run **cfgmgr**) on both nodes.
3. On each node, run

   ```
   lsdev -C | grep "2.0 GB 16 Bit Differential SCSI Disk Drive"
   ```

   A list of the disks, similar to the following, is output:

   ```
   hdisk8  Available 00-14-01-00 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk9  Available 00-14-01-10 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk14 Available 00-14-01-20 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk15 Available 00-14-01-30 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk16 Available 00-14-01-40 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk17 Available 00-14-01-50 2.0 GB 16 Bit Differential SCSI Disk Drive
   ```

4. Rerun **lsdev** and **grep** for the slot number displayed for the disks in the previous step to get the name of the adapters for these disks. For example, the output from the previous step displayed that the disks are in slot 14:

   ```
   lsdev -C | grep "00\-14"
   ```

   The output from this command looks similar to the following:

   ```
   ascsi0       Available 00-14       Wide SCSI I/O Controller Adapter
   vscsi0       Available 00-14-00    SCSI I/O Controller Protocol Device
   vscsi1       Available 00-14-01    SCSI I/O Controller Protocol Device
   hdisk8       Available 00-14-01-00 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk9       Available 00-14-01-10 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk14      Available 00-14-01-20 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk15      Available 00-14-01-30 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk16      Available 00-14-01-40 2.0 GB 16 Bit Differential SCSI Disk Drive
   hdisk17      Available 00-14-01-50 2.0 GB 16 Bit Differential SCSI Disk Drive
   ```

5. As displayed in the above output, **vscsi1** is the controller for these disks because it is at 00-14-01; the disks are also coming from 00-14-01. Move the disks from the Available state to the Defined state by removing all the devices except **ascsi0** (but don't remove their definition from the ODM). For example:

   ```
   rmdev -l vscsi0
   rmdev -l vscsi1
   rmdev -l hdisk8
   rmdev -l hdisk9
   ⋮
   ```

   Note: Do steps 3, 4, and 5 on both nodes.

6. On one node, use SMIT to change the External SCSI ID from **7** to something else (**6**, for example).

   a. Enter **smit**

© Copyright IBM Corp. 1997, 1998

b. Select **Devices**

c. Select **SCSI Adapter**

d. Select **Change/Show Characteristics of a SCSI Adapter**

e. Select **ascsi0**

f. Change **Adapter card SCSI ID** to another ID

7. Run **cfgmgr** on both nodes.
8. Run **lspv** on both nodes. There should be a physical volume ID for all of the 7134 disks. If there is not a physical volume ID for the 7134 disks, try removing the devices and rerunning **cfgmgr**. Note that it should not say **none** in this column.
9. From here, proceed as you normally would to make volume groups, logical volumes, and so forth.

Appendix D. Glossary

C

Client. The representation of an IBM Virtual Shared Disk on a node that is not the server node. Requests for data from the IBM Virtual Shared Disk can be addressed to all clients. The data will actually be provided remotely by the server node for that IBM Virtual Shared Disk. Clients are created when an IBM Virtual Shared Disk definition is configured on a node. For example, a system partition P21S1 contains nodes 1, 2, 3, and 4 of an SP2 system. All four nodes have IBM Virtual Shared Disk setup data defined in the System Data Repository (SDR). If an IBM Virtual Shared Disk named OVSDn2 exists and the command cfgvsd OVSDn2 -a is issued, clients are configured and represented on all four nodes. All four nodes can read from and write to OVSDn2.

Cluster. A set of nodes that are all clients of the same IBM Virtual Shared Disk or sets of IBM Virtual Shared Disks.

D

Direct client. The IBM Virtual Shared Disk client on the server node.

P

Primary node. When physical disks are connected to two nodes (twin-tailed), one node is designated as the primary node for each disk and the other is designated the secondary, or backup, node. The primary node is the server node for IBM Virtual Shared Disks defined on the physical disks under normal conditions. The secondary node can become the server node for the disks if the primary node is unavailable (off-line or down).

S

Secondary node. When physical disks are connected to two nodes (twin-tailed), one node is designated as the primary node for each disk and the other is designated the secondary, or backup, node. The secondary node acts as the server node for the IBM Virtual Shared disks defined on the physical disks if the primary node is unavailable (off-line or down).

Server node. The node that currently maintains and controls local access to the physical disks on which an IBM Virtual Shared Disk is created. With twin-tailed disks, where the physical disks are locally attached to two nodes, the primary node is the server node under normal conditions. If the primary node becomes unavailable, the secondary node acts as the server node. The server node satisfies requests for data from the IBM Virtual Shared Disks created on the physical disks attached to it.

© Copyright IBM Corp. 1997, 1998

Index

A

adapter recovery 71
adding logical volumes to a twin-tailed volume group 128
administrative tasks
  adding logical volumes to a twin-tailed volume group 128
  creating a twin-tailed volume group 120
  deleting logical volumes from a twin-tailed volume group 128
  extending a twin-tailed volume group 124
  overview 119
  reducing a twin-tailed volume group 130
  removing a twin-tailed volume group 136
administrative tasks. 118
application program recovery 66
  data integrity considerations 67

B

block device
  use with IBM Virtual Shared Disks 63
buddy buffers
  tuning 56
buffer allocation
  tuning 57

C

cache buffer
  tuning 57
cfghsd command 87
cfghsdvsd command 87
cfgvsd command 87
changing IBM Virtual Shared Disks on twin-tailed volume groups 47
changing IBM Virtual Shared Diskstates 45
character (raw) device
  use with IBM Virtual Shared Disks 63
checksum processing 63
cksumvsd command 87
client 1, 69
client information
  displaying 41, 42
commands 73
  cfghsd 87
  cfghsdvsd 87
  cfgvsd 87
  cksumvsd 87
  createhsd 34, 88
  createvsd 88
  ctlhsd 88
commands *(continued)*
  ctlvsd 89
  defhsd 89
  defvsd 89
  fencevsd 67, 89
  ha_vsd 90
  ha.vsd 90
  hc.vsd 90
  hsdatalst 90
  hsdvts 91
  lsfencevsd 67
  lshsd 91
  lsvsd 91
  preparevsd 91
  removehsd 92
  removevsd 92
  resumevsd 92
  spvsd 92
  startvsd 93
  statvsd 93
  stopvsd 93
  suspendvsd 93
  ucfghsd 94
  ucfghsdvsd 94
  ucfgvsd 94
  undefhsd 94
  undefvsd 95
  unfencevsd 67, 95
  updatehsd 95
  updatevsdnode 95
  updatevsdtab 96
  verparvsd 96
  vsdatalst 96
  vsdchgserver 67, 96
  vsddiag 97
  vsdelnode 97
  vsdelvg 97
  vsdnode 97
  vsdsklst 98
  vsdvg 98
  vsdvgts 98
  vsdvts 98
communication adapter recovery 71
communications failures 69
component definitions 118
configuring an IBM Virtual Shared Disk 45, 107
consideration
  data striping 73
consideration for IBM Recoverable Virtual Shared Disk environment 119
createhsd command 34, 88
  examples 37

© Copyright IBM Corp. 1997, 1998

createhsd command *(continued)*
examples with IBM Recoverable Virtual Shared Disk 37
createvsd command 88
examples 32
examples with IBM Recoverable Virtual Shared Disk 33
creating a twin-tailed volume group 120
creating Data Striping Devices
examples 37
creating HSDs
examples with IBM Recoverable Virtual Shared Disk 37
creating HSDs on multiple nodes 34
creating IBM Virtual Shared Disks
examples 32
examples with IBM Recoverable Virtual Shared Disk 33
creating IBM Virtual Shared Disks on multiple nodes 29
ctlhsd command 88
ctlvsd command 89

D

daemons
hcd 69
rvsdd 69
data integrity during recovery 67
data striping considerations 73
Data Striping Device 113, 116
cfghsd command 116
configuring 116
defining 112
displaying configuration information 116
hsdatalst command 114
management notes 118
managing
SMIT panels for 115
overview 2
performance 60
removehsd command 115
System Data Repository 112
deleting data 115
displaying data 114
undefining data 115
tuning performance 60
unconfiguring 116
undefhsd command 115
usage notes 118
Data Striping Device overview 1
debug files 81
defhsd command 89
definition
twin-tailed disk 4
twin-tailed volume group 4
definitions of twin-tailed components 118
defvsd command 89
deleting logical volumes from a twin-tailed volume group 128
deleting old rollback files 26
diagnosis information 81

E

environment, special consideration 119
error messages 81
extending a twin-tailed volume group 124

F

failure
communication adapter 71
HA topology services 78
heartbeatd recovery scenario 78
node recovery scenario 76
recovery service scenario 78
switch recovery scenario 77
fence subroutine 83
fencevsd command 67, 89

H

HA topology services failure 78
ha_vsd command 90
ha.vsd command 90
hc recovery subsystems 69
hc subsystem 66, 71
hc.activate script 66
hc.deactivate script 66
hc.h header file 66
hc.vsd command 90
hcdd daemon 69
header file, hc.h 66
heartbeatd recovery scenario 78
HSD
displaying statistics for 44
HSD considerations 73
hsdatalst command 90
HSDs
creating 34
displaying definition and client information 42
displaying node information 41
hsdvts command 91

I

IBM Recoverable Virtual Shared Disk overview 1
IBM Recoverable Virtual Shared Disk recovery subsystems 69
IBM Virtual Shared Disk
defvsd command 101
example of defining 101

IBM Virtual Shared Disk *(continued)*
global volume group information 99
IBM Virtual Shared Disk definition information 100
vsdvg command 100
IBM Virtual Shared Disk overview 1
IBM Virtual Shared DiskPerspective 19
information area 23
menu bar 19
panes 21
primary window 19
tool bar 20
IBM Virtual Shared Disks
cfgvsd command 107
changing states 45
configuring 45, 107
creating 29
ctlvsd command 111
data integrity 63
displaying client information 41
displaying configuration information 110
displaying node information 41
displaying operational parameters for IBM Virtual Shared Disk device driver 111
displaying statistics for 43
displaying usage information for an IBM Virtual Shared Disk 110
displaying usage information for the IBM Virtual Shared Disk device driver 110
lsvsd command 110
managing 105
SMIT panels for 106
migration 8
node information 27
overview 1
performance 53
planning 7
preparevsd command 108
preparing 45, 108
resumevsd command 108
resuming 46, 108
sequence numbers 64
setting operational parameters for IBM Virtual Shared Disk device driver 111
starting 45, 107
startvsd command 108
states 106
statvsd command 111
stopping 46, 109
stopvsd command 109
suspending 46, 108
suspendvsd command 109
System Data Repository
deleting data 104
transmission protocol 63
tuning performance 53
ucfgvsd command 109
IBM Virtual Shared Disks *(continued)*
unconfiguring 46, 109
verification test suite 112
verifying 40, 112
verifying IBM Virtual Shared Disks without IBM Recoverable Virtual Shared Disk 40
IBM Virtual Shared Diskss
System Data Repository
displaying data 103
vsdsklst command 103
incarnation number 71
information area 23
installation
IBM Recoverable Virtual Shared Disk hardware requirements 13
notes 15
space requirements 12
installation notes 15

L

lsfencevsd command 67
lshsd command 91
lsvsd command 91
LVM
considerations in tuning 54

M

management, twin-tailed disks 118
manual pages for public code xxiii
maximum I/O request size
tuning 58
mbufs
tuning 56
membership list
hc 71
menu bar 19
messages 81
migrating IBM Virtual Shared Disks to the IBM Recoverable Virtual Shared Disk 17
Migrating to a new level of PSSP 9

N

network failures 69
node failure recovery scenario 76
node information
displaying 41
node statistics 43

O

overview
Data Striping Device 1
IBM Recoverable Virtual Shared Disk 1
IBM Virtual Shared Disk 1

overview of administrative tasks 119

P

panes 21
parameters
  tuning 54
pbufs
  tuning 59
Perspective, IBM Virtual Shared Disk 19
  information area 23
  menu bar 19
  panes 21
  primary window 19
  tool bar 20
Perspectives graphical user interface
  IBM IBM Virtual Shared Disk Nodes panel 26
  initial panel 26
  SDR notebook page 26
preparevsd command 91
primary node 69
primary window for IBM Virtual Shared Disks 19
programming considerations 66
programming, application 66
PSSP
  migrating to a new level 9

Q

quorum 69

R

raw (character) device
  use with IBM Virtual Shared Disks 63
recognizing recovery 75
recoverable applications 66
recovery 75
  communication adapter 71
  HA topology services 78
  heartbeatd failure 78
  High Performance Switch failure 77
  node failure 76
  planning for 75
  recognizing 75
  recovery service daemon failure 78
recovery scripts 73
recovery service daemon failure 78
recovery subsystems 69
reducing a twin-tailed volume group 130
removehsd command 92
removevsd command 92
removing a twin-tailed volume group 136
request blocks
  tuning 59
requirements, installation 13
resumevsd command 92
rollback files 26
rvsd subsystem 69
rvsdd daemon 69

S

scenarios for recovery 75
secondary node 69
server 1
setting timestamps 47
space requirements for installation 12
spvsd command 92
startup 72
startvsd command 93
states
  changing 45
statistics
  HSD 44
  IBM Virtual Shared Disk 43
statvsd command 93
stopvsd command 93
subroutines
  fence 83
subsystems
  hc 66, 71
  rvsd 69
suspendvsd command 93
switch failure recovery scenario 77
switch parameters
  tuning 55
switch pool
  tuning 56
system startup 72

T

timestamps, setting 47
tool bar 20
tunable parameters 54
tuning
  buddy buffers 56
  buffer allocation 57
  cache buffer 57
  LVM considerations 54
   maximum I/O request size 58
  mbufs 56
  pbufs 59
  request blocks 59
  switch parameters 55
  switch pool 56
twin-tailed disk definition 4
twin-tailed disk management 118
twin-tailed disks
  changing IBM Virtual Shared Disks on 47

twin-tailed volume group definition 4

U

ucfghsd command 94
ucfghsdvsd command 94
ucfgvsd command 94
undefhsd command 94
undefvsd command 95
unfencevsd command 67, 95
updatehsd command 95
updatevsdnode command 95
updatevsdtab command 96
using the IBM Recoverable Virtual Shared Disk 69

V

verparvsd command 96
VSD_Global_Volume Group object 32
VSD_Table object 32
vsdatalst command 96
vsdchgserver command 67, 96
vsddiag command 97
vsdelnode command 97
vsdelvg command 97
vsdnode command 97
vsds
  migrating to the IBM Recoverable Virtual Shared Disk 17
vsdsklst command 98
vsdvg command 98
vsdvgts command 98
vsdvts command 98

Communicating Your Comments to IBM

IBM Parallel System Support Programs for AIX
Managing Shared Disks

Publication No. SA22-7279-01

If you especially like or dislike anything about this book, please use one of the methods listed below to send your comments to IBM. Whichever method you choose, make sure you send your name, address, and telephone number if you would like a reply.

Feel free to comment on specific errors or omissions, accuracy, organization, subject matter, or completeness of this book. However, the comments you send should pertain to only the information in this manual and the way in which the information is presented. To request additional publications, or to ask questions or make comments about the functions of IBM products or systems, you should talk to your IBM representative or to your IBM authorized remarketer.

When you send comments to IBM, you grant IBM a nonexclusive right to use or distribute your comments in any way it believes appropriate without incurring any obligation to you.

If you are mailing an RCF from a country other than the United States, you can give the RCF to the local IBM branch office or IBM representative for postage-paid mailing.

- If you prefer to send comments by mail, use the RCF at the back of this book.
- If you prefer to send comments by FAX, use this number:
  - FAX: (International Access Code)+1+914+432-9405
- If you prefer to send comments electronically, use this network ID:
  - IBMLink: (United States customers only): KGNVMC(MHVRCFS)
  - IBM Mail Exchange: USIB6TC9 at IBMMAIL
  - Internet e-mail: mhvrcfs@vnet.ibm.com
  - World Wide Web: http://www.s390.ibm.com/os390

Make sure to include the following in your note:

- Title and publication number of this book
- Page number or topic to which your comment applies

Optionally, if you include your telephone number, we will be able to respond to your comments by phone.

Reader's Comments — We'd Like to Hear from You

IBM Parallel System Support Programs for AIX
Managing Shared Disks

Publication No. SA22-7279-01

You may use this form to communicate your comments about this publication, its organization, or subject matter, with the understanding that IBM may use or distribute whatever information you supply in any way it believes appropriate without incurring any obligation to you. Your comments will be sent to the author's department for whatever review and action, if any, are deemed appropriate.

Note: Copies of IBM publications are not stocked at the location to which this form is addressed. Please direct any requests for copies of publications, or for assistance in using your IBM system, to your IBM representative or to the IBM branch office serving your locality.

Today's date: _____

What is your occupation?

Newsletter number of latest Technical Newsletter (if any) concerning this publication:

How did you use this publication?

[ ]	As an introduction	[ ]	As a text (student)
[ ]	As a reference manual	[ ]	As a text (instructor)
[ ]	For another purpose (explain)		

Is there anything you especially like or dislike about the organization, presentation, or writing in this manual? Helpful comments include general usefulness of the book; possible additions, deletions, and clarifications; specific errors and omissions.

Page Number: Comment:

Name

Address

Company or Organization

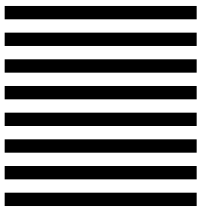
Phone No.

Reader's Comments — We'd Like to Hear from You
SA22-7279-01

Cut or Fold
Along Line

Fold and Tape **Please do not staple** Fold and Tape

NO POSTAGE
NECESSARY
IF MAILED IN THE
UNITED STATES



BUSINESS REPLY MAIL

FIRST-CLASS MAIL PERMIT NO. 40 ARMONK, NEW YORK

POSTAGE WILL BE PAID BY ADDRESSEE

IBM Corporation
Department 55JA, Mail Station P384
522 South Road
Poughkeepsie NY 12601-5400

Fold and Tape **Please do not staple** Fold and Tape

Cut or Fold
Along Line

Printed in the United States of America
on recycled paper containing 10%
recovered post-consumer fiber.

SA22-7279-01