

Bull

HACMP 4.4

Enhanced Scalability Installation and Administration Guide

Volume 1/2

AIX



Bull

HACMP 4.4

Enhanced Scalability Installation and Administration Guide

Volume 1/2

AIX

Software

August 2000

**BULL CEDOC
357 AVENUE PATTON
B.P.20845
49008 ANGERS CEDEX 01
FRANCE**

**ORDER REFERENCE
86 A2 62KX 02**

The following copyright notice protects this book under the Copyright laws of the United States of America and other countries which prohibit such actions as, but not limited to, copying, distributing, modifying, and making derivative works.

Copyright © Bull S.A. 1992, 2000

Printed in France

Suggestions and criticisms concerning the form, content, and presentation of this book are invited. A form is provided at the end of this book for this purpose.

To order additional copies of this book or other Bull Technical Publications, you are invited to use the Ordering Form also provided at the end of this book.

Trademarks and Acknowledgements

We acknowledge the right of proprietors of trademarks mentioned in this book.

AIX[®] is a registered trademark of International Business Machines Corporation, and is being used under licence.

UNIX is a registered trademark in the United States of America and other countries licensed exclusively through the Open Group.

Year 2000

The product documented in this manual is Year 2000 Ready.

Contents

About This Guide	xv
-------------------------	-----------

Part 1	Planning HACMP/ES Clusters	
Chapter 1	Building HACMP/ES Clusters	1-1
	Overview	1-1
	Defining Availability	1-1
	The Availability Costs and Benefits Continuum	1-2
	High Availability with HACMP/ES	1-2
	Enhancing Availability with the SP Hardware	1-3
	RS/6000 SP System	1-3
	Disk Subsystems	1-3
	Enhancing Availability with the AIX Software	1-5
	Journaled File System	1-6
	Disk Mirroring	1-6
	Process Control	1-6
	Error Notification	1-6
	Using HACMP/ES to Ensure Total System Availability	1-6
	High Availability Clustering	1-6
	Automatic Fallover Restores Services	1-7
	Fallover vs. Fallback	1-7
	Flexible Fallover Configurations	1-7
	Clients	1-8
	IP Address Takeover	1-8
	Applications	1-9
	User-Defined Events	1-9
	IBM RS/6000 Cluster Technology Availability Services ..	1-9
	Restrictions on HACMP/ES	1-11
	New Features in HACMP/ES 4.4	1-11
Chapter 2	Overview: Planning an HACMP/ES Cluster	2-1
	Design Goal: Eliminating Single Points of Failure	2-1
	Planning Guidelines	2-1
	Eliminating Cluster Objects as Single Points of Failure ..	2-2

The Planning Process	2-3
Using the Planning Worksheets	2-3
Step 1: Planning for Highly Available Applications	2-3
Step 2: Planning Cluster Network Connectivity	2-3
Step 3: Planning Shared Disk Devices	2-3
Step 4: Planning Shared LVM Components	2-3
Step 5: Planning Resource Groups	2-4
Step 6: Tailoring Cluster Event Processing	2-4
Step 7: Planning HACMP/ES Clients	2-4
Step 8: Installing an HACMP/ES Cluster	2-4

Chapter 3 Initial Cluster Planning 3-1

Prerequisites	3-1
Overview	3-1
Application Servers	3-3
Applications Integrated with HACMP/ES	3-3
Application Monitoring	3-4
Completing the Application and Application Server Worksheets	3-4
Completing the <i>Application Worksheet</i>	3-4
Completing the <i>Application Server Worksheet</i>	3-7
Completing the <i>Application Monitoring Worksheet</i>	3-8
Planning for AIX Fast Connect	3-12
Planning for AIX Connections	3-14
Planning for Communications Server for AIX (CS/AIX) .	3-16
Starting to Draw a Cluster Diagram	3-17
Where You Go From Here	3-19

Chapter 4 Planning Cluster Network Connectivity 4-1

Prerequisites	4-1
Overview	4-1
Designing the Network Topology	4-2
Eliminating the Network as a Single Point of Failure	4-2
Supported Network Types	4-4
Network Planning Considerations	4-5
Eliminating the TCP/IP Subsystem as a Single Point of Failure	4-6
Eliminating Network Adapters as a Single Point of Failure	4-8
Boot/Service/Standby Address Requirements for Resource Groups	4-10
Planning for the SP Switch Network	4-11
SP Switch Address Resolution Protocol (ARP)	4-13
Using HACMP/ES with NIS and DNS	4-14
How HACMP/ES Enables and Disables Nameserving ...	4-15

Planning for Cluster Performance	4-17
I/O Pacing	4-18
Syncd Frequency	4-18
Failure Detection Parameters	4-18
Completing the Network Worksheets	4-21
Completing the TCP/IP Networks Worksheet	4-21
Subnet Considerations for Cluster Monitoring with Tivoli	4-22
Completing the TCP/IP Network Adapter Worksheet	4-23
Completing the Serial Networks Worksheet	4-25
Completing the Serial Network Adapter Worksheet	4-25
Defining Hardware Addresses	4-26
Selecting an Alternate Hardware Address	4-26
Avoiding Network Conflicts	4-29
Adding the Network Topology to the Cluster Diagram	4-30
Where You Go From Here	4-30

Chapter 5

Planning Shared Disk Devices	5-1
Prerequisites	5-1
Overview	5-1
Choosing a Shared Disk Technology	5-2
SCSI Disk Planning Considerations	5-2
IBM Serial Storage Architecture Disk Subsystem	5-5
Disk Power Supply Considerations	5-6
SCSI Device Power Considerations	5-6
IBM SSA Disk Subsystem Configurations	5-6
Planning for Non-Shared Disk Storage	5-7
Planning a Shared SCSI-2 Disk Installation	5-8
Disk Adapters	5-8
Cables	5-8
Sample SCSI-2 Differential Configuration	5-9
Sample SCSI-2 Differential Fast/Wide Configuration	5-10
Sample IBM 7135-210 RAIDiant Disk Array Configuration	5-10
Sample IBM 2105 Versatile Storage Server Configuration	5-12
Planning a Shared IBM SSA Disk Subsystem Installation	5-13
IBM Manuals	5-13
Adapters	5-13
Using SSA Features For High Availability	5-15
Testing Loops	5-16
Planning for Concurrent SSA Volume Groups	5-17
SSA Disk Fencing in Concurrent Access Clusters	5-17
SSA Disk Fencing Implementation	5-17

Completing the Disk Worksheets	5-19
Completing the Shared SCSI-2 Disk Worksheet	5-19
Completing the Shared SCSI-2 Disk Array Worksheet ...	5-19
Completing the IBM Serial Storage Architecture Disk Subsystems Worksheet	5-20
Adding the Disk Configuration to the Cluster Diagram	5-20
Where You Go From Here	5-20

Chapter 6

Planning Shared LVM Components 6-1

Prerequisites	6-1
Overview	6-1
Planning for Non-Concurrent Access	6-1
Planning for Concurrent Access	6-1
LVM Components in the HACMP/ES Environment	6-2
Physical Volumes	6-2
Volume Groups	6-3
Logical Volumes	6-3
Filesystems	6-3
LVM Mirroring	6-4
Mirroring Physical Partitions	6-4
Mirroring Journal Logs	6-5
Quorum	6-5
Quorum at Varyon	6-6
Quorum after Varyon	6-6
Disabling and Enabling Quorum	6-6
Using Quorum in Non-Concurrent Access Configurations	6-7
Using Quorum in Concurrent Access Configurations	6-8
Using NFS with HACMP/ES	6-9
Reliable NFS Server Capability	6-10
Creating Shared Volume Groups	6-10
NFS Exporting Filesystems and Directories	6-10
NFS Mounting and Fallover	6-11
LVM Planning Considerations	6-14
Completing the Shared LVM Components Worksheets	6-14
Completing the Non-Shared Volume Group Worksheet ..	6-15
Completing the Shared Volume Group/Filesystem Worksheet	6-15
Completing the NFS-Exported Filesystem Worksheet	6-16
Concurrent Access Worksheets	6-17
Adding LVM Information to the Cluster Diagram	6-18
Where You Go From Here	6-18

Chapter 7	Planning Resource Groups	7-1
	Prerequisites	7-1
	Overview	7-1
	Planning Considerations	7-1
	Completing the <i>Resource Group Worksheet</i>	7-2
	Where You Go From Here	7-4
Chapter 8	Cluster Events: Tailoring and Creating	8-1
	Prerequisites	8-1
	Overview	8-1
	Tailoring Cluster Event Processing.....	8-1
	Event Notification	8-2
	Pre- and Post-Event Scripts	8-2
	Event Recovery and Retry	8-3
	User-defined Events	8-3
	Changing the Rules File	8-4
	Completing the Cluster Event Worksheet	8-6
	Where You Go From Here	8-7
Chapter 9	Planning for HACMP/ES Clients	9-1
	Prerequisites	9-1
	Overview	9-1
	Different Types of Clients: Computers and Terminal Servers .	9-1
	Client Application Systems	9-1
	NFS Servers	9-2
	Terminal Servers	9-2
	Clients Running Clinfo.....	9-2
	Reconnecting to the Cluster	9-2
	Tailoring the clinfo.rc Script	9-2
	Clients Not Running Clinfo	9-3
	Network Components	9-3
	Where You Go From Here	9-3

Part 2

Installing and Configuring HACMP/ES Clusters

Chapter 10

Overview: Installing and Configuring HACMP/ES 10-1

Prerequisites	10-1
Overview	10-1
Using the xhacmpm Utility to Configure HACMP/ES ..	10-2
Steps for Installing and Configuring an HACMP/ES Server .	10-2
Check Installed Hardware	10-2
Define Shared LVM Components	10-2
Tailor AIX for HACMP/ES	10-2
Install HACMP/ES Software	10-2
Update HACMP/ES Software	10-2
Set up the Cluster Information Program	10-2
Configure the HACMP/ES Software	10-2
Steps for Installing and Configuring an HACMP/ES Client .	10-3
Install the Base System on Clients	10-3
Edit the /usr/es/sbin/cluster/etc/clhosts File	10-3
Edit the /usr/es/sbin/cluster/etc/clinfo.rc Script	10-3
Update Non-Clinfo Clients	10-3
Reboot the Clients	10-3
Specified Operating Environment	10-3
Hardware Requirements	10-3
HACMP 4.3.1 Device Support	10-5

Chapter 11

Checking Installed Hardware 11-1

Overview	11-1
Checking Network Adapters	11-1
Ethernet, Token-Ring, and FDDI Adapters	11-1
Completing the TCP/IP Network Adapter Worksheets ..	11-1
Serial Networks	11-2
Checking an SP Switch	11-2
Configuring for Asynchronous Transfer Mode (ATM)	11-3
HACMP/ES Configuration Restrictions	11-3
ATM Configuration Restrictions	11-3
Configuring ATM ARP Servers for Use by HACMP/ES	
Nodes	11-4
Configuring ATM ARP Clients on HACMP/ES Cluster	
Nodes	11-6
Defining the ATM Network to HACMP/ES	11-7
ATM LAN Emulation	11-8
Defining the ATM LAN Emulation Network to	
HACMP/ES	11-9

Checking Shared External Disk Devices	11-9
Verifying Shared SCSI-2 Differential Disks	11-9
Verifying IBM SCSI-2 Differential Disk Arrays	11-12
Configuring Target Mode SCSI Connections	11-15
Checking the Status of SCSI Adapters and Disks	11-15
Enabling Target Mode SCSI Devices in AIX	11-16
Testing the Target Mode Connection	11-17
Follow-Up Task	11-18
Configuring Target Mode SSA Connections	11-18
Changing Node Numbers on Systems in SSA Loop	11-18
Configuring Target Mode SSA Devices	11-18
Testing the Target Mode Connection	11-19
Defining the Target Mode SSA Serial Network to HACMP/ES	11-19
Verifying Shared IBM SSA Disk Subsystems	11-20

Chapter 12

Defining Shared LVM Components 12-1

Overview	12-1
TaskGuide for Creating Shared Volume Groups	12-1
TaskGuide Requirements	12-1
Starting the TaskGuide	12-2
Defining Shared LVM Components for Non-Concurrent Access	12-2
Defining Shared LVM Components with AIX Mirroring	12-3
Defining Shared LVM Components without AIX Mirroring	12-3
Creating a Shared Volume Group on Source Node	12-4
Creating a Shared Filesystem on Source Node	12-4
Renaming jfslogs and Logical Volumes on Source Node	12-4
Adding Copies to Logical Volume on Source Node	12-5
Testing a Filesystem	12-5
Varying Off a Volume Group on the Source Node	12-6
Importing a Volume Group onto Destination Nodes	12-6
Changing a Volume Group's Startup Status	12-6
Varying Off Volume Group on Destination Nodes	12-6
Defining Shared LVM Components for Concurrent Access .	12-7
Creating a Concurrent Access Volume Group on a Source Node	12-7

Chapter 13	Tailoring AIX for HACMP/ES	13-1
	Tailoring AIX	13-1
	I/O Pacing	13-1
	Checking User and Group IDs	13-2
	Network Options	13-2
	Editing the /etc/hosts File and nameserver Configuration	13-2
	Editing the /.rhosts File	13-3
	Editing the /etc/rc.net File on NFS Clients	13-4
	SP Switch Address Resolution Protocol (ARP)	13-4
	AIX Automounter Daemon	13-5
	AIX Error Notification Facility	13-5
	HACMP Automatic Error Notification	13-6
	Emulation of Error Log Entries	13-8
Chapter 14	Installing the HACMP/ES Software	14-1
	Prerequisites	14-1
	Overview	14-2
	Contents of the Installation Media	14-2
	Conversion and Migration from HACMP for AIX to HACMP/ES 4.4	14-3
	Setting the LANG Variable	14-3
	Changes to Symbolic Links in HACMP/ES 4.4	14-3
	HACMP/ES Installation Choices	14-4
	Installation Server	14-4
	Installation Media	14-7
	Hard Disk Installation	14-8
	Installing the Concurrent Resource Manager	14-10
	Problems During the Installation	14-11
	Processor ID Licensing Issues	14-11
	Verifying Cluster Software	14-12
	Rebooting Cluster Nodes and Clients	14-12
	Converting from HACMP for AIX to HACMP/ES 4.4	14-12
	Step 1: Check on Committed Software	14-12
	Step 2: Save HACMP for AIX Cluster Configuration in a Snapshot	14-12
	Step 3: Remove the HACMP for AIX Software	14-13
	Step 4: Install HACMP/ES 4.4	14-13
	Step 5: Install the Saved Snapshot	14-13
	Step 6: Reinstall Saved Customized Event Scripts	14-13
	Step 7: Verify the Configuration	14-13
	Step 8: Reboot Cluster Nodes and Clients	14-13

	Node-by-Node Migration from HACMP for AIX Version 4.4 to HACMP/ES Version 4.4	14-14
	Prerequisites for Node-by-Node Migration	14-14
	How to Perform a Node-by-Node Migration	14-15
	Notes About the Migration Process	14-17
	Handling Node Failure During the Migration Process ...	14-18
	Backout Procedure	14-19
Chapter 15	Upgrading an HACMP/ES Cluster	15-1
	Preparing for an Upgrade	15-1
	Upgrading your HACMP/ES Cluster	15-2
	Installing HACMP/ES Software	15-3
	Upgrading Existing HACMP/ES Software to Version 4.4 ...	15-3
	cl_convert and clconvert_snapshot	15-5
	Upgrading the Concurrent Resource Manager	15-5
	Problems During the Installation	15-6
	Making Additional Changes to the Cluster	15-7
	Modifying Cluster Snapshots	15-7
	Upgrading Clinfo Applications to HACMP/ES 4.4	15-8
Chapter 16	Configuring Clinfo Scripts and Files	16-1
	Editing the /usr/es/sbin/cluster/etc/clhosts File	16-1
	Editing the /usr/es/sbin/cluster/etc/clinfo.rc Script	16-2
Chapter 17	Installing and Configuring Clients	17-1
	Prerequisites	17-1
	Overview	17-1
	Installing the Base System Client Images	17-2
	Editing the /usr/es/sbin/cluster/etc/clhosts File on Clients ...	17-2
	Editing the /usr/es/sbin/cluster/etc/clinfo.rc Script	17-3
	Updating Non-Clinfo Clients	17-4
	Pinging Non-Clinfo Clients	17-4
	Updating the ARP Cache on Non-Clinfo Clients after IP Address Takeover	17-4
	Rebooting the Clients	17-5

Chapter 18	Configuring an HACMP/ES Cluster	18-1
	Overview	18-1
	Defining the Cluster Topology	18-1
	Defining the Cluster ID and Name	18-1
	Defining Nodes	18-2
	Defining Adapters	18-2
	Configuring Global Networks	18-5
	Checking Topology Services and Group Services	18-5
	Configuring Cluster Performance Tuning	18-6
	Synchronizing the Cluster Definition across Nodes	18-8
	Configuring Resource Groups and Resources	18-9
	Adding Resource Groups	18-9
	Configuring Application Servers	18-10
	Configuring Applications Integrated with HACMP: AIX Fast Connect, AIX Connections, and CS/AIX	18-11
	Configuring AIX Fast Connect	18-11
	Configuring AIX Connections	18-13
	Configuring CS/AIX Communications Links	18-14
	Configuring Application Monitoring	18-17
	Overview	18-17
	Application Monitoring Prerequisites and Considerations	18-18
	Configuring a Process Application Monitor	18-19
	Configuring a User-defined Application Monitor	18-21
	Configuring Resources in a Resource Group	18-24
	Resource Configuration Considerations	18-24
	Entering Resource Information in SMIT	18-25
	Configuring Run-Time Parameters	18-29
	Synchronizing Cluster Resources	18-30
	Configuring Cluster Security	18-31
	Configuring Kerberos Security with HACMP/ES for AIX	18-31
	Configuring Kerberos Automatically	18-32
	Configuring Kerberos Manually	18-33
	PSSP 3.2 Enhanced Security Options	18-36
	DCE Authentication	18-37
	Verifying the Cluster Environment	18-37
	Verifying Cluster and Node Environment	18-37
	Checking Cluster Topology	18-39
	Customizing Cluster Log Files	18-40
	Configuring Cluster Events	18-41
	Configuring Custom Cluster Events	18-41
	Configuring Cluster Event Processing	18-42
	Sample Custom Scripts	18-44
	Making cron jobs Highly Available	18-44
	Making Print Queues Highly Available	18-45

Part 3	Volume 1 Appendixes	
Appendix A	Planning Worksheets	A-1
Appendix B	Using the Online Cluster Planning Worksheet Program	B-1
Appendix C	Applications and HACMP	C-1
Appendix D	Installing and Configuring Cluster Monitoring with Tivoli	D-1
HACMP/ES Master Index for Vols 1-2		MIX-1

Contents

About This Guide

This guide, in two volumes, provides information necessary to plan, install, configure, maintain, and troubleshoot the High Availability Cluster Multi-Processing for AIX Enhanced Scalability (HACMP/ES) software.

Who Should Use This Guide

This guide is intended for system administrators and customer engineers responsible for:

- Planning hardware and software resources for an HACMP/ES cluster
- Installing and configuring an HACMP/ES cluster
- Maintaining and troubleshooting an HACMP/ES cluster.

As a prerequisite to installing the HACMP/ES software, you should be familiar with

- RS/6000 system components (including disk devices, cabling, and network adapters)
- The AIX operating system, including the Logical Volume Manager subsystem
- The System Management Interface Tool (SMIT)
- Communications, including the TCP/IP subsystem.

Overview of Contents

Volume 1, Planning and Installation

Volume 1 contains the following chapters on planning, installing, and configuring your cluster environment:

- Chapter 1, Building HACMP/ES Clusters, describes how to use HACMP/ES software to build highly available clusters.
- Chapter 2, Overview: Planning an HACMP/ES Cluster, provides an overview of the recommended planning process.
- Chapter 3, Initial Cluster Planning, describes the initial steps you take to plan an HACMP/ES cluster to make applications highly available.
- Chapter 4, Planning Cluster Network Connectivity, describes planning the network support for an HACMP/ES cluster.
- Chapter 5, Planning Shared Disk Devices, discusses information you must consider before configuring shared external disks in an HACMP/ES cluster.
- Chapter 6, Planning Shared LVM Components, describes planning shared volume groups for an HACMP cluster.
- Chapter 7, Planning Resource Groups, describes how to plan resource groups within an HACMP/ES cluster.
- Chapter 8, Cluster Events: Tailoring and Creating, describes tailoring and creating cluster event processing for your cluster.

- Chapter 9, Planning for HACMP/ES Clients, discusses planning considerations for HACMP/ES clients.
- Chapter 10, Overview: Installing and Configuring HACMP/ES, lists the steps to install and configure HACMP/ES software, and lists required and supported hardware.
- Chapter 11, Checking Installed Hardware, describes how to verify that installed hardware is ready to support an HACMP/ES cluster.
- Chapter 12, Defining Shared LVM Components, describes how to define the LVM components shared by the nodes in an HACMP/ES cluster.
- Chapter 13, Tailoring AIX for HACMP/ES, discusses several general tasks necessary to ensure that your HACMP/ES environment works only as planned.
- Chapter 14, Installing the HACMP/ES Software, describes how to install the HACMP/ES for AIX, version 4.4 LPP on cluster nodes and clients in a cluster environment. It contains instructions for a new installation and for converting HACMP to HACMP/ES.
- Chapter 15, Upgrading an HACMP/ES Cluster, describes how to upgrade an HACMP/ES cluster to the most recent software.
- Chapter 16, Configuring Clinfo Scripts and Files, describes how to edit Clinfo-related files and scripts.
- Chapter 17, Installing and Configuring Clients, describes how to install the HACMP/ES for AIX, version 4.4 LPP on clients and how to configure clients for an HACMP/ES cluster.
- Chapter 18, Configuring an HACMP/ES Cluster, describes how to configure an HACMP/ES cluster.
- Appendix A, Planning Worksheets, contains worksheets and worksheet samples for all the planning tasks associated with developing an HACMP/ES cluster.
- Appendix B, Using the Online Cluster Planning Worksheet Program, contains instructions for using the online planning worksheets.
- Appendix C, Applications and HACMP, discusses points to keep in mind when planning for highly available applications in HACMP.
- Appendix D, Installing and Configuring Cluster Monitoring with Tivoli provides details on the prerequisites and procedures for setting up cluster monitoring with Tivoli Framework.

Volume 2, Administration and Troubleshooting

Volume 2 contains the following chapters and appendixes:

- Chapter 19, Maintaining an HACMP/ES Cluster, provides a list of the tasks you perform to maintain an HACMP/ES system, related administrative tasks, and a list of AIX files modified by HACMP/ES.
- Chapter 20, Starting and Stopping Cluster Services, explains how to start and stop cluster services on cluster nodes and clients. It also describes how to use C-SPOC to start and stop cluster services.
- Chapter 21, Monitoring an HACMP/ES Cluster, describes tools you can use to monitor an HACMP/ES cluster.
- Chapter 22, Maintaining Shared LVM Components, explains how to maintain LVM components shared by nodes in an HACMP/ES cluster, including specific procedures for managing volume groups, file systems, logical volumes, and physical volumes. It also includes information about using C-SPOC to maintain LVM components.

- Chapter 23, Maintaining Shared LVM Components in a Concurrent Access Environment, explains how to maintain LVM components in a concurrent access environment, including specific procedures for managing volume groups, logical volumes, and physical volumes. It also includes information about using C-SPOC to maintain LVM components.
- Chapter 24, Changing the Cluster Configuration, explains how to update and synchronize the cluster definition across all cluster nodes.
- Chapter 25, Verifying a Cluster Configuration, describes how to verify a cluster configuration, ensuring that all resources used by HACMP/ES are validly configured and that ownership and takeover of those resources are defined and in agreement across all nodes.
- Chapter 26, Saving and Restoring Cluster Configurations, explains how to use the cluster snapshot utility to save and restore cluster configurations.
- Chapter 27, Managing Users and Groups in a Cluster, explains how to use C-SPOC to manage user accounts and groups on all nodes in a cluster by executing a C-SPOC command on a single node.
- Chapter 28, Additional Tasks: NFS and Run-Time Parameters, describes how to ensure that NFS works properly on an HACMP/ES cluster and how to change a node's run-time parameters.
- Chapter 29, Troubleshooting HACMP/ES Clusters, describes how to diagnose a problem with an HACMP/ES cluster. It shows how to view cluster log files and get trace information on HACMP/ES daemons, and covers some common problems and solutions.
- Chapter 30, The Group Services Subsystem, describes the RSCT Groups Services components and operation.
- Chapter 31, The Event Management Subsystem, describes the RSCT Event Management components and operation.
- Chapter 32, The Topology Services Subsystem, describes the RSCT Topology Services components and operation.
- Appendix E, Script Utilities, describes the utilities called by the event and startup scripts supplied with HACMP/ES.
- Appendix F, RSCT Commands and Utilities, describes the common RSCT commands and utilities.
- Appendix G, RSCT Messages, lists the messages you might receive from the RSCT services.
- Appendix H, 7x24 Maintenance, provides important information for maintaining an HACMP/ES cluster on a 7x24 basis.
- Appendix I, VSM Graphical Configuration Application describes the sample program AIX Visual System Management.

Highlighting

The following highlighting conventions are used in this guide:

<i>Italic</i>	Identifies variables in command syntax, new terms and concepts, or indicates emphasis.
Bold	Identifies pathnames, commands, subroutines, keywords, files, structures, directories, and other items whose names are predefined by the system. Also identifies graphical objects such as buttons, labels, and icons that the user selects.
Monospace	Identifies examples of specific data values, examples of text similar to what you might see displayed, examples of program code similar to what you might write as a programmer, messages from the system, or information that you should actually type.

ISO 9000

ISO 9000 registered quality systems were used in the development and manufacturing of this product.

Related Publications

The following publications come with your HACMP/ES system. They provide additional information about the High Availability Cluster Multi-Processing for AIX (HACMP for AIX) software:

- *Release Notes* in `/usr/lpp/cluster/doc/release_notes` describe hardware and software requirements
- *HACMP for AIX, Version 4.4: Planning Guide*, order number 86 A2 55KX 02
- *HACMP for AIX, Version 4.4: Installation Guide*, order number 86 A2 56KX 02
- *HACMP for AIX, Version 4.4: Administration Guide*, order number 86 A2 57KX 02
- *HACMP for AIX, Version 4.4: Troubleshooting Guide*, order number 86 A2 58KX 02
- *HACMP for AIX, Version 4.4: Programming Locking Applications*, order number 86 A2 59KX 02
- *HACMP for AIX, Version 4.4: Programming Client Applications*, order number 86 A2 60KX 02
- *HACMP for AIX, Version 4.4: Master Index and Glossary*, order number 86 A2 65KX 02
- *HACMP for AIX, Version 4.4: Enhanced Scalability Installation and Administration Guide, Vol. 1*, order number 86 A2 62KX 02
- *HACMP for AIX, Version 4.4: Enhanced Scalability Installation and Administration Guide, Vol. 2*, order number 86 A2 89KX 01

The following publications provide information about the basic software for the IBM RS/6000 SP System:

- *IBM RS/6000 SP: Planning, Volume 2, Control Workstation and Software Environment*, Order Number GA22-7281
- *IBM Parallel System Support Programs for AIX: Installation and Migration Guide*, Order Number GA22-7347
- *IBM Parallel System Support Programs for AIX: Administration Guide*, Order Number SA22-7348
- *IBM Parallel System Support Programs for AIX: Managing Shared Disks*, Order Number SA22-7349
- *IBM Parallel System Support Programs for AIX: Diagnosis Guide*, Order Number SA22-7350
- *IBM Parallel System Support Programs for AIX: Command and Technical Reference Guide*, Order Number SA22-7351
- *IBM Parallel System Support Programs for AIX: Messages Reference*, Order Number SA22-7352
- *IBM Parallel System Support Programs for AIX: Performance Monitoring Guide and Reference*, Order Number SA22-7353
- *RS/6000 Cluster Technology (RSCT): Event Management Programming Guide and Reference*, Order Number SA22-7354
- *RS/6000 Cluster Technology (RSCT): Group Services Programming Guide and Reference*, Order Number SA22-7355

The following manuals provide information about the IBM 2105 Versatile Storage Server.

- *IBM Versatile Storage Server Introduction and Planning Guide*, Order Number GC26-7223-01
- *IBM Versatile Storage Server Host Systems Attachment Guide*, Order Number SC26-7225-00.
- *IBM Versatile Storage Server User's Guide*, Order Number SC26-7224-00.
- *IBM Versatile Storage Server SCSI Command Reference 2105 Model B09*, Order Number SC26-7226.

The AIX document set, as well as manuals accompanying machine and disk hardware, also provide relevant information. For the latest information on AIX and related products, see

http://www.rs6000.ibm.com/resource/aix_resource/pubs

Ordering Publications

To order additional copies of this guide, use Order Number 86 A2 62KX 02.

Part 1

Planning HACMP/ES Clusters

This part of the book introduces the HACMP/ES software and gives guidelines for planning the cluster hardware and software components for your environment.

Chapter 1, Building HACMP/ES Clusters

Chapter 2, Overview: Planning an HACMP/ES Cluster

Chapter 3, Initial Cluster Planning

Chapter 4, Planning Cluster Network Connectivity

Chapter 5, Planning Shared Disk Devices

Chapter 6, Planning Shared LVM Components

Chapter 7, Planning Resource Groups

Chapter 8, Cluster Events: Tailoring and Creating

Chapter 9, Planning for HACMP/ES Clients

Chapter 1 Building HACMP/ES Clusters

This chapter describes how to use the IBM High Availability Cluster Multi-Processing for AIX Enhanced Scalability (HACMP/ES) software to build highly available clusters on the IBM RS/6000 Scalable POWERParallel System (SP) or a combination of different SP systems and stand-alone IBM RS/6000 workstations. HACMP/ES clusters can contain up to 32 nodes, or 8 if concurrent access is configured.

Overview

This chapter:

- Defines availability
- Describes how the hardware and software in the AIX/SP environment provide a base level of availability
- Describes how the HACMP/ES software provides high availability
- Describes the new functionality provided in this version of HACMP/ES

Defining Availability

Availability is the degree to which the system can continue to let you do your work. Numerous factors—hardware, software, administrative, and environmental—affect a system's overall availability. These factors include:

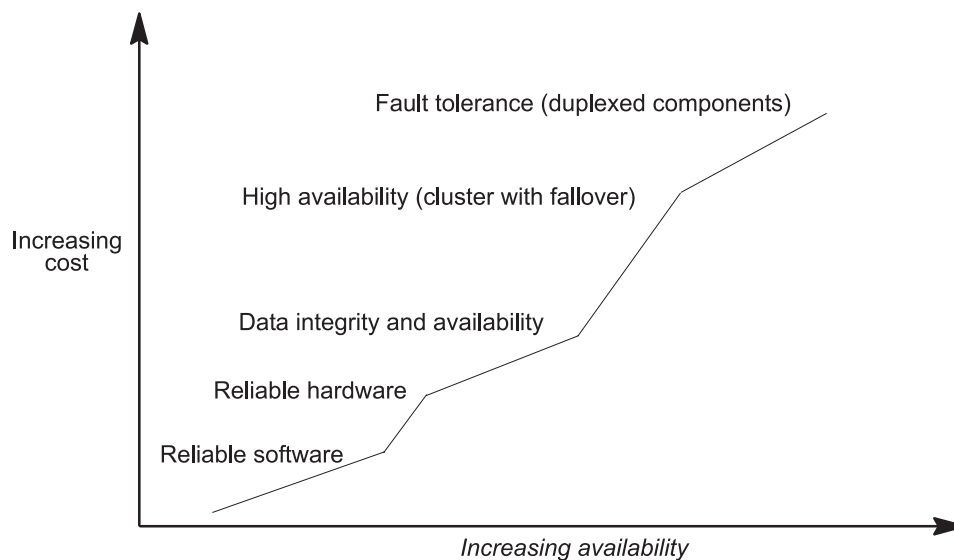
- *The reliability of the equipment itself.* Reliability is not only a function of hardware design and testing, but also of the environment in which it operates. The more hardware you have in a system, and the greater its complexity, the greater the chance of hardware failure.
- *The redundancy of components,* including processors, cards, peripherals, and networks. Redundant hardware provides a measure of availability because it shortens the time needed to replace a failed component.
- *The nature of the technology.* Solutions such as disk arrays provide a higher level of availability through technical design.
- *The software subsystems and applications* running on the hardware. As software continues to become more complicated and to require greater processor power and memory, opportunities for software failure increase. Rigorous testing and well-designed error handling schemes can influence availability.
- *Scheduled downtime* for maintenance, backups, and software revisions, and whether these tasks can be performed without disrupting the system's operations.
- *The thoroughness of manual procedures* for identifying and correcting problems.
- *Software that provides automatic fault detection, logging, and recovery* in lieu of slower, error-prone manual procedures. The system's ability to detect a failure automatically and take corrective actions also can improve system availability significantly.

The level of availability for a specific system, given all of the above, can still vary. You can implement software for automatic fault detection and recovery, but it may not be able to detect and recover from all possible failures. Further, you may need to detect and log soft-failure conditions so that a potential problem can be identified before it occurs. Also, the degree of automated recovery can have a profound impact on the length of time required to overcome a failure condition.

While you can establish manual procedures to duplicate or compensate for shortcomings in an automated solution, the speed and accuracy by which they are implemented is a critical factor. In general, an automated solution provides for less downtime, both through the speed by which an automated solution executes and through the elimination of a possible cause of manual error.

The Availability Costs and Benefits Continuum

The following figure shows the costs and benefits of availability technologies.



Cost and Benefits of Availability Technologies

As you can see, availability is not an all-or-nothing proposition. Think of availability as a continuum. Reliable hardware and software provide the base level of availability. Advanced features such as RAID devices provide an enhanced level of availability. High availability software provides near continuous access to data and applications. Specialized fault tolerant systems ensure the availability of the entire system.

High Availability with HACMP/ES

IBM's HACMP/ES software provides a low-cost commercial computing environment that ensures mission-critical applications can recover quickly from hardware and software failures. The HACMP/ES software is a high availability system that ensures that critical resources are available for processing. High availability combines custom software with industry-standard hardware to minimize downtime by quickly restoring services when a system, component, or application fails. While not instantaneous, the restoration of service is rapid, usually 30 to 300 seconds.

Contrast high availability with the traditional fault tolerant model. The fault tolerant or continuous availability model relies on specialized hardware to detect a hardware fault and instantaneously switch to a redundant hardware component—whether the failed component is a processor, memory board, power supply, I/O subsystem, or storage subsystem. Although this cutover is apparently seamless and offers non-stop service, a high premium is paid in both hardware cost and performance because the redundant components do no processing. More importantly, the fault tolerant model does not address software failures, by far the most common reason for downtime.

The distinguishing factor, then, between fault tolerance and high availability is this: *no service interruption* in a fault tolerant environment versus a *minimal service interruption* in a highly available environment. While high availability is not quite fault tolerance, many sites are willing to absorb a small amount of downtime rather than pay the much higher cost of providing fault tolerance.

Enhancing Availability with the SP Hardware

Building a highly available cluster begins with reliable hardware. Within the AIX environment, the SP and its supported disk subsystems provide a robust, stable platform for building highly available clusters.

RS/6000 SP System

The SP is a parallel processing machine that includes from two to 128 processors connected by a high-performance switch. The SP leverages the outstanding reliability provided by the RS/6000 series by including many standard RS/6000 hardware components in its design. The SP's architecture then extends this reliability by enabling processing to continue following the failure of certain components. This architecture allows a failed node to be repaired while processing continues on the healthy nodes. You can even plan and make hardware and software changes to an individual node while other nodes continue processing.

Disk Subsystems

The SP supports the disk subsystems used by the RS/6000. The disk subsystems most often shared as external disk storage in cluster configurations are:

- 7135 RAIDiant Array
- 7137 Disk Array
- 7131 and 7133 SSA serial disk subsystems
- IBM 2105 Versatile Storage Server models B09 and 100
- SCSI disks

IBM 7135 RAIDiant Disk Array Devices

You can use an IBM 7135-110 or 7135-210 RAIDiant Disk Array in HACMP/ES cluster configurations. The benefits of using an IBM 7135 RAIDiant Disk Array in an HACMP/ES cluster are its storage capacity, speed, and reliability features. The IBM 7135 RAIDiant Disk Array contains a group of disk drives that work together to provide enormous storage capacity (up to 135 GB of nonredundant storage) and higher I/O rates than single large drives.

RAID Levels

The IBM 7135 RAIDiant Disk Arrays support reliability features that provide data redundancy to prevent data loss if one of the disk drives in the array fails. As a RAID device, the array can provide data redundancy through RAID levels. The IBM 7135 RAIDiant Disk Arrays support RAID levels 0, 1, and 5. RAID level 3 can be used only with a raw disk.

In RAID level 0, data is striped across a bank of disks in the array to improve throughput. Because RAID level 0 does not provide data redundancy, it is not recommended for use in HACMP/ES clusters.

In RAID level 1, the IBM 7135 RAIDiant Disk Array provides data redundancy by maintaining multiple copies of the data on separate drives (mirroring).

In RAID level 5, the IBM 7135 RAIDiant Disk Array provides data redundancy by maintaining parity information that allows the data on a particular drive to be reconstructed if the drive fails.

All drives in the array can be hot-plugged. When you replace a failed drive, the IBM 7135 RAIDiant Disk Array reconstructs the data on the replacement drive automatically. Because of these reliability features, you should not define LVM mirrors in volume groups defined on an IBM 7135 RAIDiant Disk Array.

Dual Active Controllers

To eliminate adapters or array controllers as single points of failure in an HACMP/ES cluster, you can configure the IBM 7135 RAIDiant Disk Array with a second array controller that acts as a backup controller in the event of a fallover. This configuration requires that you configure each cluster node with two adapters, connecting each adapter to an array controller using a separate SCSI bus for each connection.

In this configuration, each adapter and array-controller combination defines a unique path from the node to the data on the disk array. The IBM 7135 RAIDiant Disk Array software manages data access through these paths. Both paths are active and can be used to access data on the disk array. If a component failure disables the current path, the disk array software automatically re-routes data transfers through the other path.

This dual-active, path-switching capability is independent of the capabilities of the HACMP/ES software, which provides protection from a node failure. When you configure the IBM 7135 RAIDiant Disk Array with multiple controllers and configure the nodes with multiple adapters and SCSI buses, the disk array software prevents a single adapter or controller failure from causing disks to become unavailable.

IBM 7137 Disk Array

The IBM 7137 disk array contains multiple SCSI-2 Differential disks. On the IBM 7137 array, you can group these disks together into multiple Logical Units (LUNs), with each LUN appearing to the host as a single SCSI device (hdisk).

IBM Serial Storage Architecture Disk Subsystem

You can use IBM 7133 and 7131-405 SSA disk subsystems as shared external disk storage devices in an HACMP/ES cluster configuration.

If you include SSA disks in a volume group that uses LVM mirroring, you can replace a failed drive without powering off the entire subsystem.

IBM 2105 Versatile Storage Server

The IBM 2105 Versatile Storage Server (VSS) provides multiple concurrent attachment and sharing of disk storage for a variety of open systems servers. RISC System/6000 processors can be attached, as well as other UNIX and non-UNIX platforms.

The VSS uses IBM SSA disk technology. Existing IBM 7133 SSA disk drawers can be used in the VSS. The RISC System/6000 attaches to the IBM 2105 Versatile Storage Server via SCSI-2 Differential Fast/Wide Adapter/A. A maximum of 64 open system servers can be attached to the VSS (16 SCSI channels with 4 adapters each).

There are many availability features included in the VSS. All storage is protected with RAID technology. RAID-5 techniques can be used to distribute parity across all disks in the array. *Sparing* is a function which allows you to assign a disk drive as a spare for availability. Predictive Failure Analysis techniques are utilized to predict errors *before* they affect data availability. Failover Protection enables one partition, or *storage cluster*, of the VSS to takeover for the other so that data access can continue.

The VSS includes other features such as a web-based management interface, dynamic storage allocation, and remote services support. For more information on VSS planning, general reference material, and attachment diagrams, see these URLs:

<http://www.storage.ibm.com/hardsoft/products/vss/books/vssrefinfo.htm>

<http://www.storage.ibm.com/hardsoft/products/vss/books/vsrlag.htm>

SCSI Disks

The benefits of the SCSI implementation are its low cost and minimal hardware overhead.

In an HACMP/ES cluster, shared SCSI disks are connected to the same SCSI bus for the nodes that share the devices. The disks are owned by only one node at a time (no concurrent access allowed). If the owner node fails, the cluster node with the next highest priority in the resource chain acquires ownership of the shared disks as part of failover processing. This ensures that the data stored on the disks remains accessible to client applications.

HACMP/ES Required and Supported Hardware

For a complete list of required and supported hardware, see Chapter 10, Overview: Installing and Configuring HACMP/ES.

Enhancing Availability with the AIX Software

The AIX operating system provides numerous features designed to increase system availability by lessening the impact of both planned (data backup, system administration) and unplanned (hardware or software failure) downtime. These features include:

- Journaled File System
- Disk mirroring
- Process control
- Error notification

Journalized File System

AIX's native file system, the Journalized File System (JFS), uses database journaling techniques to maintain its structural integrity. System or software failures cannot leave the file system in an unmanageable condition. When rebuilding the file system after a crash, AIX uses the JFS log to restore the file system to its last consistent state. Journaling thus provides faster recovery than the standard UNIX file system consistency check (fsck) utility.

Disk Mirroring

Disk mirroring software provides data integrity and on-line backup capability. It prevents data loss due to disk failure by maintaining up to three copies of data on separate disks so that data is still accessible after any single disk fails. Disk mirroring is transparent to the application. No application modification is necessary since no distinction between mirrored and conventional disks exists.

Process Control

The AIX System Resource Controller (SRC) monitors and controls key processes. The SRC can detect when a process terminates abnormally, log the termination, pass messages to a notification program, and restart the process on a backup processor.

Error Notification

The AIX Error Notification facility detects errors, such as network and disk adapter failures, and triggers a predefined response to the failure. (HACMP/ES builds on this AIX feature by providing error emulation functionality that allows you to test the predefined response without causing the error to occur, and an option to automatically configure notify methods for a set of device errors in one step.)

Using HACMP/ES to Ensure Total System Availability

The AIX features previously described eliminate specific components as single points of failure within a system. While they make a specific component more available, the system is still vulnerable to other component failures that can make it unavailable. For example, a disk array is itself a highly available component, but if the processor to which it is connected fails, the disk array also becomes unavailable. If you are doing mission-critical processing, you need something that ensures total system availability. The HACMP/ES software provides this reliability.

High Availability Clustering

HACMP/ES is built on the clustering model. A cluster is a group of systems that work together to provide fast, uninterrupted computing services.

In an HACMP/ES cluster, up to 32 SP nodes, RS/6000 standalones, or a combination of these cooperate to provide a set of services or resources to other entities. Clustering these servers to back up critical applications is a cost-effective high availability option. A business can use more of its computing power while ensuring that its critical applications resume running after a short interruption caused by a hardware or software failure.

HACMP/ES provides a highly available environment by identifying a set of resources essential to uninterrupted processing and by defining a protocol that nodes use to collaborate to ensure that these resources are available. HACMP/ES extends the clustering model by defining relationships among cooperating processors where one processor provides the service offered by a peer should the peer be unable to do so.

Automatic Fallover Restores Services

HACMP/ES uses a highly available server agent, called a recovery driver, or *cluster manager*, on each cluster node. The cluster manager is responsible for a number of tasks: monitoring local hardware and software subsystems, tracking the state of the cluster peers, and taking appropriate action in the event of a component failure. A cluster manager exchanges a heartbeat with its peers so that it can monitor the availability of the other servers in the cluster. If the heartbeat stops, the peer systems drive the recovery process. The peers take the necessary actions to get the critical applications up and running and to ensure that data has not been corrupted or lost.

Fallover vs. Fallback

It is important to keep in mind the difference between *fallover* and *fallback*. You will encounter these terms frequently in discussion of the various resource group policies.

Fallover

Fallover refers to the movement of a resource group from the node on which it currently resides (*owner* node) to another active node after its owner node experiences a failure. The new owner is not a reintegrating or joining node.

Fallback

Fallback refers to the movement of a resource group from its owner node specifically to a node that is joining or reintegrating into the cluster; a fallback occurs during a node up event.

Flexible Fallover Configurations

An HACMP/ES processor owns a set of resources: disks, volume groups, filesystems, networks, network addresses, and applications. When a processor fails or leaves the cluster, its resources are distributed among the surviving processors.

HACMP/ES's focus on resource ownership makes numerous cluster configurations possible, providing tremendous flexibility in defining the cluster environment to fit the particular needs of the application. HACMP/ES supports the following types of resource configurations:

- cascading, with or without the attribute cascading without fallback (CWOFF) enabled
- rotating
- concurrent

In a cascading configuration, you assign each node a takeover priority so that a given resource follows a defined fallover path that brings it back, whenever possible, to the highest-priority node for that resource. If it is important that specific nodes control certain resources, configure a cluster for cascading resources with the CWOFF flag set to **false**. This configuration allows you to keep the most crucial resources on the node with the greatest capability. Use cascading

resource groups with CWOFF set to **true** to avoid interruptions of service caused by fallbacks. Note that for cascading resource groups, a service IP label is optional; however, you *must* include it if you plan to use IP address takeover.

- *Cascading without Fallback* is a cascading resource group attribute which allows you to define fallback behavior. When the cascading without fallback flag is set to **false**, and a node of higher priority than that on which the resource group currently resides, joins or reintegrates into the cluster, the resource group falls back to the higher priority node. When the CWOFF flag is set to **true**, the resource group will not fallback to any node which joins or reintegrates into the cluster. A CWOFF resource group which falls over to another node when its owner node leaves the cluster does not return to the owner node when it rejoins the cluster.

If it is not of utmost importance *which* node controls a resource, and you want no interruption in service when a node rejoins the cluster, configure a cluster for *rotating* resources. In a rotating configuration, a node rejoining the cluster after a fallover does not reacquire the resources from the takeover node, so there is no accompanying interruption in service. Though similar in various ways to CWOFF, note that for a rotating resource group, (unlike CWOFF) you *must* define a service IP label. Also, while rotating groups tend to distribute themselves over various nodes, CWOFF groups may “clump” together with multiple CWOFF groups on the same node.

If you have applications that must run concurrently on several nodes, you can configure these resources for *concurrent* access by two or more cluster nodes. This provides constant shared disk access, with no time lost to fallovers or fallbacks. You can service nodeA, for example, while other nodes continue providing the service with no need for fallover when you shut down NodeA.

You will see more information on planning the resource configuration in Chapter 3, Initial Cluster Planning.

Clients

A total high availability solution must include the client machine that uses services provided by the servers. Clients can be divided into two categories, naive and intelligent. A naive client views the cluster as a single entity. If a server fails, the client must be restarted, or at least must reconnect to the server. An intelligent client, on the other hand, is cluster-aware. A cluster-aware client reacts appropriately in the face of server failure, connecting to an alternate server, perhaps masking the failure from the user. Such an intelligent client must have knowledge of the cluster state. HACMP/ES extends the cluster paradigm to clients by providing both dynamic cluster configuration reporting and notification of cluster state changes, such as changes in subsystems or node failure.

IP Address Takeover

The HACMP/ES IP address takeover facility transparently changes the network address of a redundant network adapter to the network address of the primary network adapter, should the primary adapter fail. This facility ensures that a processor has continuous access to a network. Other than for a brief delay, the fallover is usually transparent to clients. HACMP/ES also supports taking over the hardware address.

Applications

Again, the purpose of a highly available system is to ensure that critical services are accessible to users. Applications usually need no modification to run in the HACMP/ES environment. Any application that can be successfully restarted after an unexpected shutdown is a candidate for HACMP/ES.

For example, all commercial DBMS products checkpoint their state to disk in some sort of transaction journal. In the event of a server failure, the failover server restarts the DBMS, which reestablishes database consistency and then resumes processing.

If you use AIX Connections or AIX Fast Connect to share resources with non-AIX workstations, you can configure one of the other of them as an HACMP/ES resource, making it highly available in the event of node or adapter failure, and making its correct configuration verifiable with the **clverify** command.

As of version 4.4, HACMP/ES includes application monitoring capability, whereby you can define a monitor to detect the death of a process or to periodically poll the health of an application and take automatic action upon detection of a problem.

User-Defined Events

HACMP/ES provides recovery programs for all the HACMP/ES events. While you can use configuration definitions and shell scripts you customized for HACMP, you can also develop scripts for your individual recovery needs, telling HACMP/ES to react to conditions other than the ones predefined in the product. You can cause the system to respond with your specified script to any event the event manager can detect and monitor. See User-defined Events on page 8-3 for more information.

IBM RS/6000 Cluster Technology Availability Services

The IBM RS/6000 Cluster Technology (RSCT) high availability services provide greater scalability, notify distributed subsystems of software failure, and coordinate recovery and synchronization among all subsystems in the software stack.

Packaging these services with HACMP/ES makes it possible to run this software on all RS/6000s, not just on SP nodes.

RSCT Services include the following components:

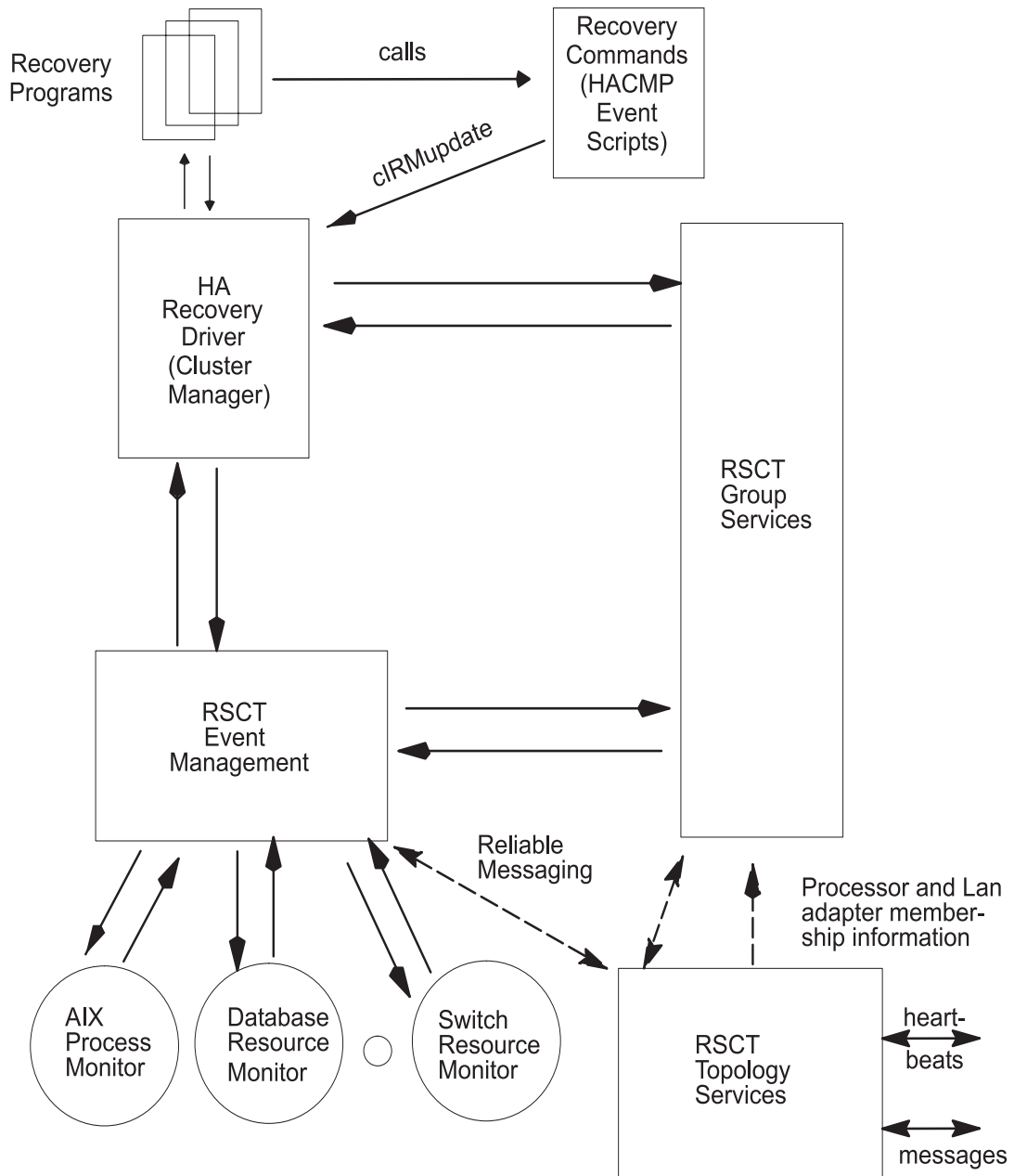
- *Event Manager*. A distributed subsystem providing a set of high availability services. It creates events by matching information about the state of system resources with information about resource conditions of interest to client programs. Client programs in turn can use event notifications to trigger recovery from system failures.
- *Group Services*. A system-wide, highly available facility for coordinating and monitoring changes to the state of an application running on a set of nodes. Group Services helps both in the design and implementation of highly available applications and in the consistent recovery of multiple applications. It accomplishes these two distinct tasks in an integrated framework.

Building HACMP/ES Clusters
 Using HACMP/ES to Ensure Total System Availability

- *Topology Services.* A facility for generating heartbeats over multiple networks and for providing information about adapter membership, node membership, and routing. Adapter and node membership provide indication of adapter and node failures respectively. Reliable Messaging uses the routing information to route messages between nodes around adapter failures.

See Part 4, *RSCT Services*, for more information on these services.

The following figure shows the main components that make up the HACMP/ES architecture.



**HACMP/ES is Comprised of IBM RSCT Availability Services
 and the HA Recovery Driver**

Restrictions on HACMP/ES

- HACMP/ES does not support the *forced down* option.
- HACMP/ES clusters with more than three nodes connected with non-IP networks (of the same type) must be connected in a “daisy chain” or point-to-point configuration. No HACMP/ES cluster node can be linked to more than two other nodes.
- HACMP/ES does not stop the joining of a node with an out-of-sync topology configuration. The administrator must synchronize topology and should verify a consistent configuration using the cluster verification utility.

New Features in HACMP/ES 4.4

HACMP/ES version 4.4 includes the following new or enhanced features:

HANFS Functionality Added to HACMP/ES 4.4

Prior to version 4.4, HACMP for AIX included a separate product subsystem called High Availability for Network File System for AIX (HANFS for AIX). HANFS for AIX provided reliable NFS server capability by allowing a backup processor to recover current NFS activity should the primary NFS server fail. The HANFS special functionality extended the HACMP architecture to include highly available modifications and locks on NFS filesystems. HANFS clusters could have a maximum of two nodes.

In version 4.4, the HANFS functionality has been added to the basic HACMP architecture. The following enhancements are included in version 4.4 of the HACMP/ES product subsystem:

- You can use the reliable NFS server capability that preserves locks and dupcache (2-node clusters only).
- You can now specify a network for NFS mounting.
- You can define NFS exports and mounts at the directory level.
- You can specify export options for NFS-exported directories and filesystems.

For more information on these added functionalities, refer to Chapter 18, Configuring an HACMP/ES Cluster, and in *Volume 2* of this manual, Chapter 28, Additional Tasks: NFS and Run-Time Parameters.

Other New or Enhanced Features

- **Application monitoring functionality.** HACMP/ES can now monitor one or more applications, defined through the SMIT interface, and take user-specified actions upon detection of process death or application failure.
- **Cluster monitoring with Tivoli.** You can now monitor the state of HACMP/ES cluster components through your Tivoli management interface.
- **Cascading without fallback,** a new resource group attribute (for cascading only) that combines some of the advantages of cascading and rotating resource policies.
- **Greater and easier control of tuning parameters.** HACMP/ES 4.4 gives you enhanced SMIT options for tuning parameters such as failure detection rate, I/O pacing, and synced frequency.

- **Enhanced clverify functionality.** The **clverify** utility now checks for errors such as invalid characters in names of nodes and other components, multiple RS232 serial networks on the same tty device, and more than two non-IP networks of one type per node. In addition, if the cluster is not active, you can now opt to save time by skipping verification during cluster synchronization.
- **TaskGuide improvements.** The TaskGuide for creating shared volume groups, introduced in version 4.3.0, now creates a JFS log automatically after creating the shared volume group, and also displays the physical location of available disks.
- **Improved conversion utilities.** Enhancements to the **cl_convert** and **clconvert_snapshot** utilities make it easier to migrate to HACMP/ES 4.4 from earlier versions of HACMP/ES or HACMP.
- **New documentation on 7x24 support.** The HACMP/ES 4.4 documentation includes a new appendix containing additional information on issues specific to maintaining a cluster on a 7x24 basis.

Chapter 2 Overview: Planning an HACMP/ES Cluster

This chapter provides an overview of the recommended planning process.

Design Goal: Eliminating Single Points of Failure

The HACMP/ES software provides numerous facilities you can use to build highly available clusters. Designing the cluster that provides the best solution for your organization requires careful and thoughtful planning. In fact, adequate planning is the key to building a successful HACMP/ES cluster. A well-planned cluster is easier to install, provides higher availability, performs better, and requires less maintenance than a poorly planned cluster.

Your major goal throughout the planning process is to eliminate single points of failure. A *single point of failure* exists when a critical cluster function is provided by a single component. If that component fails, the cluster has no other way of providing that function, and the service dependent on that component becomes unavailable.

For example, if all the data for a critical application resides on a single disk, and that disk fails, that disk is a single point of failure for the entire cluster. Clients cannot access that application until the data on the disk is restored. Likewise, if dynamic application data is stored on internal disks rather than on external disks, it is not possible to recover an application by having another cluster take over the external disks. Therefore, identifying necessary logical components required by an application, such as file systems and directories (which could contain application data and configuration variables), is an important prerequisite for planning a successful cluster.

Realize that, while your goal is to eliminate all single points of failure, you may have to make some compromises. There is usually a cost associated with eliminating a single point of failure. For example, purchasing an additional hardware device to serve as backup for the primary device increases cost. The cost of eliminating a single point of failure should be compared against the cost of losing services should that component fail. Again, the purpose of the HACMP/ES for AIX software is to provide a cost-effective, highly available computing platform that can grow to meet future processing demands.

Important: HACMP/ES for AIX is designed to recover from a *single* hardware or software failure. It may not be able to handle *multiple* failures, depending on the sequence of failures. For example, the default event scripts cannot do an adapter swap after an IP address takeover (IPAT) has occurred if only one standby adapter exists for that network.

Planning Guidelines

To be highly available, all cluster resources associated with a critical application should be without single points of failure. As you design an HACMP/ES cluster, your goal is to identify and address all potential single points of failure. Questions to ask yourself include:

- What services are required to be highly available? What is the priority of these services?
- What is the cost of a failure compared to the necessary hardware to eliminate the possibility of this failure?

Overview: Planning an HACMP/ES Cluster
Design Goal: Eliminating Single Points of Failure

- What is the required availability of these services? Do they need to be available 24 hours a day, seven days a week? Or is eight hours a day, five days a week sufficient?
- What could happen to disrupt the availability of these services?
- What is the allotted time for replacing a failed resource? What is an acceptable degree of performance degradation while operating after a failure?
- Which failures are detected as cluster events? Which failures need to have custom code written to detect the failure and trigger a cluster event?
- What is the skill level of the group implementing the cluster? The group maintaining the cluster?

To plan, implement, and maintain a successful HACMP/ES cluster requires continuing communication among many groups within your organization. Ideally, you should assemble the following representatives (as applicable) to aid in HACMP/ES planning sessions:

- Network administration
- System administration
- Database administration
- Application programming
- Support
- End users

Eliminating Cluster Objects as Single Points of Failure

The table below summarizes potential single points of failure within an HACMP/ES cluster and describes how to eliminate them.

Cluster Object	Eliminated as Single Point of Failure By...
Node	Using multiple nodes
Power source	Using multiple circuits or uninterruptable power supplies
Network adapter	Using redundant network adapters
Network	Using multiple networks to connect nodes
TCP/IP subsystem	Using serial networks to connect adjoining nodes and clients
Disk adapter	Using redundant disk adapters
Controller	Using redundant disk controllers
Disk	Using redundant hardware and disk mirroring
Application	Assigning a node for application takeover; configuring an application monitor

See Chapter 1, *Building HACMP/ES Clusters*, for a discussion on eliminating cluster objects as single points of failure. The following chapters provide more information about how to plan for elimination of these single points of failure.

The Planning Process

This section describes the recommended steps for planning an HACMP/ES cluster. As you plan a cluster, be sure to plan for application servers and resource groups within the cluster, and plan to tailor event processing to allow the cluster to handle special failure situations.

Using the Planning Worksheets

At each stage of the cluster planning process, you are provided with two types of worksheets to aid in your planning. The traditional paper worksheets, which you fill out by hand and have physically nearby to refer to as you configure your cluster, are located in this book, in Appendix A, Planning Worksheets.

You can also use the web-based online planning worksheet program, which is located in the `/usr/es/lpp/cluster/samples` directory along with your other HACMP/ES filesets. The online worksheets allow you to enter data as you plan. At the end of the planning process, you can write all your entered configuration data to an AIX file and transfer it to your cluster. For more information, see Appendix B, Using the Online Cluster Planning Worksheet Program.

Step 1: Planning for Highly Available Applications

In this step you plan the core of the cluster—the applications to be made highly available, the types of resources they require, the number of nodes, shared IP addresses, and a mode for sharing disks (non-concurrent or concurrent access). Your goal is to develop a high-level view of the system that serves as a starting point for the cluster design. After making these initial decisions you record them in the Application Planning worksheets and start to draw a diagram of the cluster. Chapter 3, Initial Cluster Planning, describes this step of the planning process.

Step 2: Planning Cluster Network Connectivity

In this step you plan the networks that connect the nodes in your system. You first examine issues relating to TCP/IP and serial networks in an HACMP/ES environment. After deciding on the networks you will use, complete the Network Planning worksheets and add the networks to the cluster diagram. Chapter 4, Planning Cluster Network Connectivity, describes this step of the planning process.

Step 3: Planning Shared Disk Devices

In this step you plan the shared disk devices for the cluster. You decide which disk storage technologies you will use in your cluster, and examine issues relating to those technologies in the HACMP/ES environment. Complete the disk worksheets and add the shared disk configuration to your diagram. Chapter 5, Planning Shared Disk Devices, describes this step of the planning process.

Step 4: Planning Shared LVM Components

In this step you plan the shared volume groups for the cluster. You first examine issues relating to LVM components in an HACMP/ES environment, and then you fill out worksheets describing physical and logical storage. Chapter 6, Planning Shared LVM Components, describes this step of the planning process.

Step 5: Planning Resource Groups

Planning resource groups pulls together all the information you have generated in the previous steps. Complete the resource group planning worksheets. Chapter 7, Planning Resource Groups, describes this step of the planning process.

Step 6: Tailoring Cluster Event Processing

In this step you tailor the event processing for your cluster. Chapter 8, Cluster Events: Tailoring and Creating, describes this step of the planning process.

Step 7: Planning HACMP/ES Clients

In this step you examine issues relating to HACMP/ES clients. Chapter 9, Planning for HACMP/ES Clients, describes this step of the planning process.

Step 8: Installing an HACMP/ES Cluster

After completing the planning steps, you are ready to install the cluster. Use the planning diagrams and worksheets you completed during the planning process to guide you through the installation process. See Chapter 14, Installing the HACMP/ES Software, for complete instructions.

Chapter 3 Initial Cluster Planning

This chapter describes the initial steps you take to plan an HACMP/ES cluster to make applications highly available, including completing the planning worksheets.

Prerequisites

It is essential that you understand the concepts and terminology necessary for planning an HACMP/ES cluster. Read Chapter 1, Building HACMP/ES Clusters, before beginning the planning process. The planning steps in this chapter assume a thorough understanding of the information presented in that chapter.

Overview

The central purpose for combining nodes in a cluster is to provide a highly available environment for mission-critical applications. In many organizations, these applications must remain available at all times. For example, an HACMP/ES cluster could run a database server program which services client applications. The clients send queries to the server program which responds to their requests by accessing a database, stored on a shared external disk.

The table below summarizes the potential single points of failure addressed in this chapter.

Cluster Network Object	Eliminated as a Single Point of Failure by...
Node	Using multiple nodes, with IP address takeover
Application	Assigning a node for application takeover and configuring an application monitor

Planning Worksheets—Paper and Online

Throughout the planning chapters in this guide, you are encouraged to fill out the planning worksheets provided in Appendix A, Planning Worksheets.

In addition, you have the option of using the web-based online planning “worksheets,” a series of panels in which you enter configuration data as you plan your cluster. At the end of the online planning process, you can configure your cluster by transferring the data from a PC-based system to your AIX cluster nodes. Instructions for the online planning worksheets are located in Appendix B, Using the Online Cluster Planning Worksheet Program.

Initial Cluster Planning

For each critical application you need to be aware of the resources required by the application, including its processing and data storage requirements. For example, when you plan the size of your cluster, you must include enough nodes to handle the processing requirements of your application after a node fails.

You can create HACMP/ES clusters that include up to 32 nodes. Keep the following considerations in mind when determining the number of cluster nodes:

- An HACMP/ES cluster can be made up of SP nodes on a single system, or any combination of nodes from different SP systems and RS/6000 standalone workstations.
- An SP node is an SP thin, high, or wide node that runs the HACMP/ES server software. It can also run the HACMP/ES client software.
- Nodes that have entirely separate functions and do not share resources should not be combined together in a single cluster. Instead, create several smaller clusters on the SP. Smaller clusters are easier to design, implement, and maintain.
- For performance reasons, it may be desirable to use multiple nodes to support the same application. To provide mutual takeover services, the application must be designed in a manner which allows multiple instances of the application to run on the same node.

For example, if an application requires that the dynamic data reside in a directory called */data*, chances are that the application cannot support multiple instances on the same processor. For such an application (running in a non-concurrent environment), try to partition the data so that multiple instances of the application can run—each accessing a unique database.

Furthermore, if the application supports configuration files that enable the administrator to specify that the dynamic data for *instance1* of the application resides in the *data1* directory, *instance2* resides in the *data2* directory, and so on, then multiple instances of the application are probably supported.

- In certain configurations, including additional nodes in the cluster design can increase the level of availability provided by the cluster; it also gives you more flexibility in planning node fallover and reintegration.

Application Planning

You must also be aware of the data resources of your application and their location within the cluster. You provide a solution that enables them to be handled correctly should a node fail. To prevent a failure, you must have a thorough understanding of the application and how it behaves in a single-node and multi-node environment. Do not make assumptions about the application's performance under adverse conditions.

With the application monitoring feature, you can direct HACMP/ES to monitor the death of a process or more subtle problems affecting an application, and automatically attempt to restart the application and take appropriate action (notification or fallover) if restart attempts fail.

In this chapter, you record all the key information about your application or applications. You record this information in an *Application Worksheet* and make the initial steps to drawing your cluster diagram.

Keep in mind the following guidelines to ensure that your applications are serviced correctly within an HACMP/ES cluster environment:

- Lay out the application and its data so that only the data resides on shared external disks. This arrangement not only prevents software license violations, but it also simplifies failure recovery.
- Write robust scripts to both start up and shut down the application on the cluster nodes. The startup script especially must be able to recover the application from an abnormal termination, such as a power failure. You should verify that it runs properly in a single-node environment before including the HACMP/ES software. Be sure to include the start and stop resources on both the *Application Worksheet* and the *Application Server Worksheet* in Appendix A, Planning Worksheets. You will use this information as you install the HACMP/ES software.
- Some vendors require a unique license for each processor that runs an application, which means that you must license-protect the application by incorporating processor-specific information into the application when it is installed. As a result, it is possible that even though the HACMP/ES software processes a node failure correctly, it is unable to restart the application on the fallover node because of a restriction on the number of licenses available within the cluster for that application. To avoid this problem, be sure that you have a license for each system unit in the cluster that may potentially run an application.
- Verify that the application executes successfully in a single-node environment. Debugging an application in a cluster is more difficult than debugging it on a single processor.

For further discussion of what types of applications work best under HACMP/ES and some strategies that can help keep your applications highly available, see Appendix C, Applications and HACMP.

Application Servers

To put the application under HACMP/ES control, you create an *application server* resource that associates a user-defined name with the names of specially written scripts to start and stop the application. By defining an application server, HACMP/ES can start another instance of the application on the takeover node when a fallover occurs. An application server can also be monitored with the application monitoring feature.

Once you define the application server, you can add it to a *resource group*. A resource group is a set of resources that you define so the HACMP/ES software can treat them as a single unit. You will find full instructions for adding your application servers and other resources to resource groups, and configuring application monitors, in Chapter 18, Configuring an HACMP/ES Cluster.

Applications Integrated with HACMP/ES

AIX Fast Connect, AIX Connections, and CS/AIX network operating softwares are already integrated with HACMP/ES and can be configured directly as highly available resources, without application servers or additional scripts. In addition, the integration of these applications means the **clverify** utility verifies the correctness and consistency of your AIX Fast Connect, AIX Connections, or CS/AIX configuration.

Later sections in this chapter describe how these applications work in HACMP/ES and how to plan for configuring them.

Application Monitoring

HACMP/ES version 4.4 can monitor applications that are defined to application servers, in one of two ways:

- *Process monitoring* detects the death of a process, using RSCT event management capability.
- *User-defined monitoring* monitors the health of an application based on a monitor method that you define.

Instructions later in this chapter help you to decide which method is appropriate for your applications, and to fill out the *Application Monitoring* worksheet(s).

Also refer to the full instructions for defining application monitors using the SMIT interface in Chapter 18, Configuring an HACMP/ES Cluster.

Application and Application Server Worksheets

In an HACMP/ES environment, the critical applications can themselves be a single point of failure. To ensure the availability of these applications, you create an *application server* cluster resource under HACMP/ES control. The application server associates a user-defined name with the names of scripts created to start and stop the application. When you define an application server, HACMP/ES can start another instance of the application when needed.

Completing the Application and Application Server Worksheets

To help you plan the applications for your cluster, you will now complete the following worksheets:

- *Application Worksheet*
- *Application Server Worksheet*
- *Application Monitoring (Process or User-Defined) Worksheet*

Photocopy these worksheets from Appendix A before completing the following procedure. You will need a copy of the worksheet for each application in the cluster.

You can use the online worksheet program instead of, or in addition to, filling out the paper worksheets. However, the online worksheet program does not include entries for application monitoring.

See Appendix B, Using the Online Cluster Planning Worksheet Program for more information.

Completing the *Application Worksheet*

Complete an *Application Worksheet* for each application you want to make highly available in your cluster.

The following sections describe how to fill in the various fields in this worksheet.

Record Key Information about the Application

To complete the application planning worksheet, perform the following:

1. Assign a name to the application and record it in the application name in the **Application Name** field.
2. Enter information describing the application's executable and configuration files under the **Directory/Path, Filesystem, Location, and Sharing** columns. Be sure to enter the full path name of each file.

Note: You can store the filesystem for either the executable or configuration files on either an internal or external disk device. Different situations may require you to do it one way or the other. Be aware, if you store the filesystem on the internal device, that the device will not be accessible to other nodes during a resource takeover.

3. Enter information describing the application's data and log files under the appropriate columns listed in Step 2. Data and log files can be stored in a file system (or on a logical device) and must be stored externally if a resource takeover is to occur successfully.
4. Enter in the **Normal Start Command/Procedures** field the names of the start command/script you created to start the application after a resource takeover.
5. Enter in the **Verification Commands/Procedures** field the names of commands or procedures to use to verify that the normal start command/script executed successfully.

Assigning a Name and ID to Your Cluster

Assign a unique name and ID to your cluster and record them in the **Cluster Name** and **Cluster ID** fields. Make sure that the cluster name and ID do not conflict with the names and IDs of other clusters at your site.

Cluster ID The cluster ID can be any positive integer less than 99,999. The cluster ID must be unique for each cluster on the network.

Cluster Name The cluster name is an arbitrary string of no more than 31 characters (alphanumeric and underscores only).

Defining the Node Relationship of Your Cluster

Define the relationship of the nodes in your cluster (their resource group type) and record it in the **Node Relationship** field.

Fallover vs. Fallback

It is important to understand the difference between *fallover* and *fallback*. You will encounter these terms frequently in discussion of the various resource group policies. **Fallover** refers to the movement of a resource group from the node on which it currently resides (*owner* node) to another node, after its owner node experiences a failure. The new owner node was active at the time of the original owner node failure, and is not a reintegrating or joining node. **Fallback** refers to the movement of a resource group from its owner node specifically to a node which is joining or reintegrating into the cluster; a fallback occurs during a `node_up` event.

Resource Group Types

Resource group types can be cascading, rotating, or concurrent.

- *Cascading*, where a resource group is taken over by one or more nodes in a resource chain according to the takeover priority assigned to each node. The available node within the cluster with the highest priority will own the resource group. Depending on the order in which nodes are brought up in the cluster, the resource group may cascade from one node to another until reaching the node that has been assigned the highest priority for that resource group.
 - *Cascading without Fallback (CWOFF)* is a cascading resource group attribute which defines fallback behavior. When CWOFF is set to **false**, normal cascading behavior occurs, that is, a resource group falls back to any higher priority node when such a node joins or reintegrates into the cluster, causing an interruption in service. When CWOFF is set to **true**, the resource group will not fallback to any node which joins or reintegrates into the cluster, thus avoiding the interruption in service caused by fallback.
- *Rotating*, where a resource group is associated with a group of nodes and rotates among these nodes. When a node fails, the highest priority node on its boot address will acquire the resource group. When a node rejoins the cluster, however, it does not reacquire resource groups; instead, it rejoins the cluster as a standby node. Rotating groups share some similarities with Cascading without Fallback groups, yet unlike CWOFF, rotating groups require the use of IP address takeover. Furthermore, while CWOFF groups tend to “clump” on one node, rotating groups distribute over the nodes in a cluster; if several rotating groups share a network, only one of these resource groups can be up on a given node at any time
- *Concurrent*, where a resource can be accessed by more than one node at the same time. When a node fails, the other nodes in the concurrent configuration still have access; no fallover or fallback occurs.

Keep the following considerations in mind when deciding which resource group type to assign:

- If specifying a preferred node for a critical application is essential, and does not require an IP address, a *cascading* configuration may be the best resource group choice. Using cascading resources ensures that an application is owned by a specified preferred node whenever that node is active in the cluster. This ownership allows the node to cache data the application frequently uses, thereby improving the node’s performance for providing data to the application.

If the active node fails, its resources will be taken over by the available node with the highest priority in the resource chain. Note, however, that when the failed node reintegrates into the cluster, it temporarily disrupts application availability as it takes back its resources.

Cascading configurations do not require IP addresses unless you plan on using IP address takeover. (See Chapter 4, Planning Cluster Network Connectivity, for more information on configuring IP address takeover.)

- Use cascading resource groups with Cascading without Fallback set to **false** when you have a strong preference for which cluster node you want to control a resource group. For example, you may want the cluster node with the highest processing capabilities to control the resource group. Use cascading resource groups with CWOFF set to **true** to avoid interruptions of service caused by fallbacks.
- If it is important to avoid downtime associated with fallbacks as well as to keep resources distributed among more than one node, a *rotating* configuration may be the best choice. Application availability is not disrupted during node reintegration because the reintegrating node rejoins the cluster as a standby node only and does not attempt to reacquire its resources.

Note that rotating resource groups require that you define a service IP label.

- If it is crucial to have constant shared access with no time lost to fallovers or fallbacks, you may want to choose a *concurrent* configuration. A concurrent cluster can have up to eight nodes.

To define a concurrent configuration, you must have the HACMP/ES Concurrent Resource Manager (ESCRM) software installed.

Assigning Each Cluster Node a Name and a Role in the Fallover Strategy

Assign each node a unique name and record it in the Node fields under the **Fallover Strategy** heading. The node name is an arbitrary string of no more than 31 characters (alphanumeric characters and underscores only).

Under each node, record its role in the takeover strategy. For cascading configurations, use the letter P to indicate the node that is the primary owner of the application resource group. Use the letter T to indicate a takeover node, including a number to indicate its priority in the takeover chain. The priority is determined by the order in which you list the nodes in the resource chain.

While each application can be assigned to only one resource group, a single node can support multiple resource groups of different types.

Enter any takeover caveats that may be associated with a particular node name in fields under the **Node Reintegration/Takeover Caveats** heading.

Completing the *Application Server Worksheet*

To help you plan the application servers for your cluster, you will now complete an *Application Server Worksheet* referenced in Appendix A. Photocopy this worksheet before completing the following procedure.

Enter the cluster ID in the Cluster ID field and the cluster name in the **Cluster Name** field. You determined these values while completing the *Application Worksheet*.

For each application server you define, fill in the following fields:

Initial Cluster Planning

Application and Application Server Worksheets

- Record the name of the application in the **Application** field. You assigned a name to the application in the *Application Worksheet*.
- Assign a symbolic name that identifies the server and record it in the **Server Name** field. For example, you could name the application server for the customer database application *custdata*. The name must be no more than 31 characters, and can contain alphanumeric and underscore (`_`) characters only.
- Record the full pathname of the user-defined script that starts the application server in the **Start Script** field. This information was recorded in the *Application Worksheet*. Be sure to include the script's arguments, if necessary. The script is called by the cluster event scripts. For example, you could name the start script and specify its arguments for starting the *custdata* application server as follows:

```
    /usr/es/sbin/cluster/utils/start_custdata -d mydir -a jim_svc
```

where the **-d** option specifies the name of the directory for storing images, and the **-a** option specifies the service IP address (label) for the server running the demo.
- Record the full pathname of the user-defined script that stops the server in the **Stop Script** field. This script is called by the cluster event scripts. For example, you could name the stop script for the *custdata* application server `/usr/es/sbin/cluster/utils/stop_custdata`.

Completing the *Application Monitoring Worksheet*

To help prepare for defining application monitors for application servers, you should complete one or more *Application Monitoring Worksheets (Process or User-Defined)*.

Completing the Application Monitor (Process) Worksheet

To plan the configuration of a process monitor, photocopy the worksheet for each application server you plan to monitor, and fill in the worksheet(s) as follows:

1. Fill in the Cluster ID and Cluster Name fields.
2. Specify the application server name for which you are configuring a process monitor.
3. Check whether this application can be monitored with a process monitor. Shell scripts, for example, cannot. See the section *Configuring Application Monitoring* on page 18-17 for more information.

If the answer is yes, proceed to step 4. If no, proceed to the instructions for the User-Defined Application Monitor worksheet.

4. Indicate the name(s) of one or more processes to be monitored.

Note: Be careful when listing process names. It is very important that the names are correct when you enter them in SMIT as you configure the application monitor. For more information, see the section *Identifying Correct Process Names* on page 18-19.

5. Specify the user id of the owner of the processes specified in step 4, for example *root*. Note that the process owner must own all processes to be monitored.
6. Specify how many instances of the application to monitor. The default is **1** instance.

Note: This number *must* be **1** if you have specified more than one process to monitor.

7. Specify the time (in seconds) to wait before beginning monitoring. For instance, with a database application, you may wish to delay monitoring until after the start script and initial database search have been completed. You may need to experiment with this value to balance performance with reliability.

Note: In most circumstances, this value should *not* be zero. A minimum of 2-3 seconds is recommended to allow HACMP and user applications to quiesce before beginning application monitoring.

8. Specify the restart count, denoting the number of times to attempt to restart the application before taking any other actions. The default is **3**.

Note: Make sure you enter a Restart Method (see step 13) if your Restart Count is any non-zero value.

9. Specify the interval (in seconds) that the application must remain stable before resetting the restart count. This interval becomes important if a number of failures occur over a period of time. Resetting the count to zero at the proper time keeps a later failure from being counted as the last failure from the previous problem, when it should be counted as the first of a new problem.

Do not set this to be shorter than (Restart Count) x (Stabilization Interval). The default is 10% longer than that value. If it is too short, the count will be reset to zero repeatedly, and the specified failure action will never occur.

10. Specify the action to be taken if the application cannot be restarted within the restart count. The default choice is **notify**, which runs an event to inform the cluster of the failure. You can also specify **fallover**, in which case the resource group containing the failed application moves over to the cluster node with the next-highest priority for that resource group.

Note: Keep in mind that if you choose the **fallover** option of application monitoring, which may cause resource groups to migrate from their original owner node, the possibility exists that while the highest priority node is up, the resource group remains down. Unless you bring the resource group up manually, it will remain in an inactive state. See Common Problems and Solutions on page 29-24 for more information.

11. (Optional) Define a notify method that will run when the application fails. This user-defined method, typically a shell script, runs during the restart process and during notify activity.

12. (Optional) Specify an application cleanup script to be invoked when a failed application is detected, before invoking the restart method. The default is the application server stop script you must define when you set up the application server.

Note: Since the application is already stopped when this script is called, the server stop script may fail. For more information on writing correct stop scripts, see Appendix C, Applications and HACMP.

13. Specify a restart method if desired. (This is required if Restart Count is not zero.) The default restart method is the application server start script you define when the application server is set up.

Completing the Application Monitor (User-Defined) Worksheet

If you plan to set up a custom monitor method, complete this worksheet for each user-defined application monitor you plan to configure, as follows:

1. Fill in the Cluster ID and Cluster Name fields.
2. Fill in the name of the application server.
3. Specify a script or executable for custom monitoring of the health of the specified application. You must not leave this field blank when you configure the monitor in SMIT. The monitor method must return a zero value if the application is healthy, and a non-zero value if a problem is detected. See the note below regarding defining a monitor method.
4. Specify the polling interval (in seconds) for how often the monitor method is to be run. If the monitor does not respond within this interval, it is considered “hung.”
5. Specify a signal to kill the user-defined monitor method if it does not return within the monitor interval. The default signal is **kill -9**.
6. Specify the time (in seconds) to wait before beginning monitoring. For instance, with a database application, you may wish to delay monitoring until after the start script and initial database search have been completed. You may need to experiment with this value to balance performance with reliability.

Note: In most circumstances, this value should *not* be zero. A minimum of 2-3 seconds is recommended to allow HACMP and user applications to quiesce before beginning application monitoring.

7. Specify the restart count, denoting the number of times to attempt to restart the application before taking any other actions. The default is **3**.
8. Specify the interval (in seconds) that the application must remain stable before resetting the restart count. This interval becomes important if a number of failures occur over a period of time. Resetting the count to zero at the proper time keeps a later failure from being counted as the last failure from the previous problem, when it should be counted as the first of a new problem.

Do not set this to be shorter than (Restart Count) x (Stabilization Interval + Monitor Interval). The default is 10% longer than that value. If it is too short, the count will be reset to zero repeatedly, and the specified failure action will never occur.

9. Specify the action to be taken if the application cannot be restarted within the restart count. You can keep the default choice **notify**, which runs an event to inform the cluster of the failure, or choose **fallover**, in which case the resource group containing the failed application moves over to the cluster node with the next-highest priority for that resource group.

Note: Keep in mind that if you choose the **fallover** option of application monitoring, which may cause resource groups to migrate from their original owner node, the possibility exists that while the highest priority node is up, the resource group remains down. Unless you bring the resource group up manually, it will remain in an inactive state. See Common Problems and Solutions on page 29-24 for more information.

10. (Optional.) Define a notify method that will run when the application fails. This user-defined method runs during the restart process and during a server_down event.
11. (Optional) Specify an application cleanup script to be invoked when a failed application is detected, before invoking the restart method. The default is the application server stop script defined when the application server was set up.

Note: The application may be already stopped when this script is called, and the server stop script may fail. For more information on writing correct stop scripts, see Appendix C, Applications and HACMP.

12. (Required if Restart Count is not zero.) The default restart method is the application server start script you define when you set up the application server. You can specify a different method here if desired.

Notes on Defining a Monitor Method

When devising your custom monitor method, keep the following points in mind:

- The monitor method must be an executable program (it can be a shell script) that tests the application and exits, returning an integer value that indicates the application's status. The return value must be zero if the application is healthy, and must be a non-zero value if the application has failed.
- HACMP will not pass arguments to the monitor method.
- The monitor method logs messages to `/tmp/clappmond.<application monitor name>.monitor.log` by printing messages to the standard output (`stdout`) file. The monitor log file is overwritten each time the application monitor runs.
- Since the monitor method is set up to be killed if it does not return within the specified polling interval, do not make the method overly complicated.

You should test your monitor method under different workloads to arrive at the best polling interval value.

Planning for Applications Integrated with HACMP/ES

Some applications do not require application servers, because they are already integrated with HACMP/ES. You do not need to write additional scripts or create an application server for these to be made highly available under HACMP/ES.

- **AIX Connections** software enables you to share files, printers, applications, and other resources between AIX workstations and PC and Mac clients. AIX Connections allows you to take advantage of AIX's multi-user and multi-tasking facilities, scalability, file and record locking features, and other security features with clients on other platforms. The AIX Connections application is integrated with HACMP/ES so that you can configure it as a resource in your HACMP/ES cluster, making the protocols handled by AIX Connections—IPX/SPX, Net BEUI, and AppleTalk—highly available in the event of node or adapter failure. For more information, see Planning for AIX Connections on page 3-14.

Initial Cluster Planning

Planning for Applications Integrated with HACMP/ES

- **AIX Fast Connect** allows client PCs running Windows, DOS, and OS/2 operating systems to request files and print services from an AIX server. Fast Connect supports the transport protocol NetBIOS over TCP/IP. You can configure AIX Fast Connect resources using the SMIT interface. See Planning for AIX Fast Connect on page 3-12.
- **Communications Server for AIX (CS/AIX)** is also integrated with HACMP/ES, allowing you to designate Data Link Control (DLC) profiles and their associated objects as highly available resources. See Planning for Communications Server for AIX (CS/AIX) on page 3-16 for more information.

In addition, the integration of these applications means the **clverify** utility verifies the correctness and consistency of your AIX Fast Connect, AIX Connections, or CS/AIX configuration.

Planning for AIX Fast Connect

AIX Fast Connect allows client PCs running Windows, DOS, and OS/2 operating systems to request files and print services from an AIX server. Fast Connect supports the transport protocol NetBIOS over TCP/IP. You can configure AIX Fast Connect resources using the SMIT interface.

The Fast Connect application is integrated with HACMP/ES already so you can configure Fast Connect services, via the SMIT interface, as highly available resources in resource groups. HACMP/ES can then stop and start the Fast Connect resources when failover, recovery, and dynamic resource group migrations occur. This application does not need to be associated with application servers or special scripts.

Converting from AIX Connections to AIX Fast Connect

You cannot have both AIX Fast Connect and AIX Connections configured in the same resource group at the same time. Therefore, if you previously configured the AIX Connections application as a highly available resource, and you now wish to switch to AIX Fast Connect, you should take care to examine your AIX Connections planning and configuration information before removing it from the resource group.

Remember these points when planning for conversion from AIX Connections to Fast Connect:

- Keep in mind that AIX Fast Connect does not handle the AppleTalk and NetWare protocols that AIX Connections is able to handle. Fast Connect is primarily for connecting with clients running Windows operating systems. Fast Connect uses NetBIOS over TCP/IP.
- You will need to unconfigure any AIX Connections services before configuring AIX Fast Connect services as resources.
- Take care to note the AIX Connections services configuration, so you can make sure that AIX Fast Connect connects you to all of the files and print queues you have been connected to with AIX Connections.

For additional details and instructions, see the section Converting from AIX Connections to AIX Fast Connect on page 24-26 in *Volume 2*.

Planning Considerations for Fast Connect

When planning for configuration of Fast Connect as a cluster resource in HACMP/ES, keep the following points in mind:

- Install the Fast Connect Server on all nodes in the cluster.
- If Fast Connect printshares are to be highly available, the AIX print queue names must match for every node in the cluster.
- For cascading and rotating resource groups, assign the *same* netBIOS name to each node when the Fast Connect Server is installed. This action will minimize the steps needed for the client to connect to the server after failover.

Note: Only one instance of a netBIOS name can be active at one time. For that reason, remember not to activate Fast Connect servers that are under HACMP/ES control.

- For concurrently configured resource groups, assign *different* netBIOS names across nodes.
- In concurrent configurations, you should define a second, non-concurrent, resource group to control any filesystem that must be available for the Fast Connect nodes. Having a second resource group configured in a concurrent cluster keeps the AIX filesystems used by Fast Connect cross-mountable and highly available in the event of a node failure.
- As stated in the previous section, you cannot configure both Fast Connect and AIX Connections in the same resource group or on the same node.
- Fast Connect cannot be configured in a mutual takeover configuration. In other words, a node cannot participate in two Fast Connect resource groups at the same time.

Fast Connect as a Highly Available Resource

When configured as part of a resource group, AIX Fast Connect resources are handled by HACMP/ES as follows.

Start/Stop of Fast Connect

When a Fast Connect server has resources configured in HACMP/ES, HACMP/ES starts and stops the server during failover, recovery, and resource group reconfiguration or migration.

Note: The Fast Connect server must be stopped on all nodes when bringing up the cluster. This ensures that HACMP/ES will start the Fast Connect server and handle its resources properly.

Node Failure

When a node that owns Fast Connect resources fails, the resources become available on the takeover node. When the failed node rejoins the cluster, the resources are again available on the original node (as long as the resource policy is such that the failed node reacquires its resources).

There is no need for clients to reestablish a connection to access the Fast Connect Server after failover, as long as IP and hardware address takeover are configured and occur, and users have configured their Fast Connect server with the same NetBIOS name on all nodes (for non-concurrent resources groups).

Note: For switched networks and for clients not running Clinfo, you may need to take additional steps to ensure client connections after failover. See Chapter 9, Planning for HACMP/ES Clients for more information.

Initial Cluster Planning

Planning for Applications Integrated with HACMP/ES

Adapter Failure

When a service adapter running the transport protocol needed by the Fast Connect server fails, HACMP/ES performs an adapter swap as usual, and Fast Connect establishes a connection with the new adapter. After an adapter failure, clients are temporarily unable to access shared resources such as files and printers; after the adapter swap is complete, clients can again access their resources.

Completing the Fast Connect Worksheet

Now fill out the Fast Connect worksheet to identify the resources you will enter in SMIT when you configure Fast Connect as a resource. The worksheets are located in Appendix A, Planning Worksheets. Make photocopies as needed.

To complete the planning worksheet for Fast Connect:

1. Record the cluster ID in the Cluster ID field.
2. Record the cluster name in the Cluster Name field.
3. Record the name of the resource group that will contain the Fast Connect Resources.
4. Record the nodes participating in the resource group.
5. Record the Fast Connect Resources to be made highly available. These resources will be chosen from the SMIT picklist when you configure your resources.
6. Record the filesystems that contain the files or directories that you want Fast Connect to share. Be sure to specify these in the Filesystems SMIT field when you configure the resource group.

See the instructions on using SMIT to configure Fast Connect services as resources in Chapter 18, Configuring an HACMP/ES Cluster.

Planning for AIX Connections

AIX Connections software enables you to share files, printers, applications, and other resources between AIX workstations and PC and Mac clients. AIX Connections allows you to take advantage of AIX's multi-user and multi-tasking facilities, scalability, file and record locking features, and other security features with clients on other platforms. The AIX Connections application is integrated with HACMP/ES so that you can configure it as a resource in your HACMP/ES cluster, making the protocols handled by AIX Connections—IPX/SPX, Net BEUI, and AppleTalk—highly available in the event of node or adapter failure.

AIX Connections Realms and Services

Three realms are available through AIX Connections:

- **NB for NetBIOS clients** offers services over two transport protocols, TCP/IP and NetBEUI. This lets your AIX workstation running Connections provide file, print, terminal emulation, and other network services to client PCs running DOS, OS/2, Windows, or Windows NT.
- **NW for NetWare clients** offers services over IPX/SPX, so your AIX workstation can provide services to NetWare-compatible client PCs.
- **AT for AppleTalk clients** offers services over AppleTalk, so your AIX workstation can act as a Macintosh AppleShare server and provide services to Macintosh clients on an Ethernet or token ring network.

Planning Notes for AIX Connections

- In order to configure AIX Connections services in HACMP/ES, you must copy the AIX Connections installation information to all nodes on which the program might be active.
- You must configure a realm's AIX Connections to use the service adapter, not the standby.

AIX Connections as a Highly Available Resource

When AIX Connections services are configured as resources, HACMP/ES handles the protocols during the following events:

Start-up

HACMP/ES starts the protocols handled by AIX Connections when it starts the resource groups.

Note: If you are running NetBIOS and it is attached to other non-HACMP/ES applications, you need to restart those applications after failover, startup, and dynamic reconfiguration events involving an AIX Connections resource group.

Node Failure and Recovery

In the event of a node failure and node start, HACMP/ES handles the AIX Connections services listed in the resource groups just like all other resources so listed. It starts and stops them just as it does volume groups and filesystems.

After resource group takeover for node failure (not adapter failure), AIX Connections services revert to allowing new connections so that HACMP/ES can swap adapters and reconnect users to the services. This allows the reconnection to take place without interruption, and without users even noticing.

Note: If you have services previously set to *reject* new connections, be aware that they are not automatically reset this way on takeover. You must reset these manually.

You can include a command to reject new connections in the `start_server` script if you do not wish to do the manual reset after failover. However, be aware that when services are permanently set to reject new connections, users will experience an interruption in service as they will not automatically connect to the server.

Adapter Failure

In the event of an adapter failure, HACMP/ES moves all AIX Connections protocols in use on the failed adapter to a currently active standby adapter if one is available. See Chapter 11, Checking Installed Hardware for a discussion of adapter failure considerations.

Completing the AIX Connections Worksheet

Now fill out the AIX Connections worksheet to identify the realm/service pairs you will enter in SMIT when you configure AIX Connections as a resource.

In the AIX Connections worksheet:

1. Record the cluster ID in the Cluster ID field.

Initial Cluster Planning

Planning for Applications Integrated with HACMP/ES

2. Record the cluster name in the Cluster Name field.
3. Record the name of the resource group that will contain the AIX Connections services.
4. Record the nodes participating in this resource group.
5. Record the AIX Connections realm/service pairs to be made highly available, based on the following:
 - A *realm* is one of the following, as described earlier: **NB**, **NW**, **AT**
 - A *service* is one of the following: **file**, **print**, **term**, **nvt**, and **atlw**.

You assign a name to the AIX Connections service and then specify realm/service pairs using the format

```
<realm>/<service_name>%<service_type>
```

For example, to specify the NetBIOS realm for a file you have named *solenb*, your realm/service pair would be in the following format:

```
NB/solenb%file
```

For full instructions on configuring AIX Connections resources in SMIT, see Chapter 18, Configuring an HACMP/ES Cluster.

Planning for Communications Server for AIX (CS/AIX)

CS/AIX is a set of communications protocols that enables an AIX computer to participate in an SNA network that includes mainframes, PCs and other workstations. It is typically associated with legacy computer environments, given that a mainframe connection usually exists. CS/AIX is supported over a number of network types, including Token Ring, Ethernet and FDDI.

HACMP/ES enables you to designate CS/AIX Data Link Control (DLC) profiles as highly available resources. In addition, you can specify associated CS/AIX objects, such as ports, link stations and applications, as highly available.

You configure the highly available CS/AIX communication link(s) through the HACMP/ES SMIT interface. In a non-HACMP/ES environment, you lose your SNA network connection if the network adapter which the DLC profiles are associated with fails, or the node on which the CS/AIX software is running fails. HACMP/ES enables you to place the CS/AIX DLC profiles in a highly available resource group. Once you add CS/AIX DLC profiles to a resource group, takeover and recovery will happen automatically if a network adapter or node fails. See Chapter 18, Configuring an HACMP/ES Cluster, for more detail on the events that occur during adapter and node failover and recovery for highly available CS/AIX communications links.

Planning Considerations for CS/AIX

Keep the following points in mind when planning highly available CS/AIX communication links:

- This feature is supported with the following two CS/AIX products: Communications Server for AIX Version 4.2 and eNetwork Communications Server for AIX Version 5.0.
- HACMP/ES CS/AIX communications links are supported over Token Ring, Standard Ethernet and FDDI networks.
- HACMP/ES supports LU 2 and LU 6.2 CS/AIX logical unit protocols.
- HACMP/ES requires that CS/AIX be installed on all nodes where takeover might occur. You must also define CS/AIX configuration information on all nodes where resource groups containing CS/AIX configurations might become active.

Completing the CS/AIX Communications Links Worksheet

To help you plan the CS/AIX communications links for your cluster, you will now complete the CS/AIX communications links worksheet in Appendix A, Planning Worksheets. Photocopy this worksheet for additional CS/AIX communications links as needed.

Do the following for each CS/AIX communication link in your cluster:

1. Enter the cluster ID in the **Cluster ID** field.
2. Enter the cluster name in the **Cluster Name** field.
3. Enter the communications link name in the **Communications Link Name** field.
4. Enter the resource group in which the communications link is defined in the **Resource Group** field.
5. Enter nodes participating in the resource group in the **Nodes** field.
6. Enter the DLC name in the **DLC Name** field. This is the name of an existing CS/AIX DLC profile to be made highly available.
7. Enter the names of any CS/AIX ports to be started automatically in the **Port** field.
8. Enter the names of the CS/AIX link stations in the **Link Station** field. This field is only available if you are using CS/AIX Version 5.0.
9. Enter the name of an application start script in the **Service** field. This start script starts any application layer processes that use the communication link. This field is optional.

For full instructions on configuring CS/AIX Communications link resources in SMIT, see Chapter 18, Configuring an HACMP/ES Cluster.

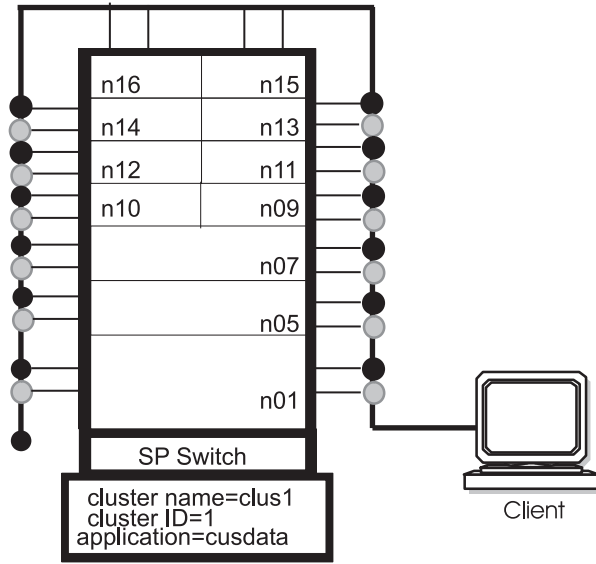
Starting to Draw a Cluster Diagram

Next, you start drawing a diagram of your cluster. The cluster diagram combines the information from each step in the planning process into one drawing that shows the cluster's function and structure.

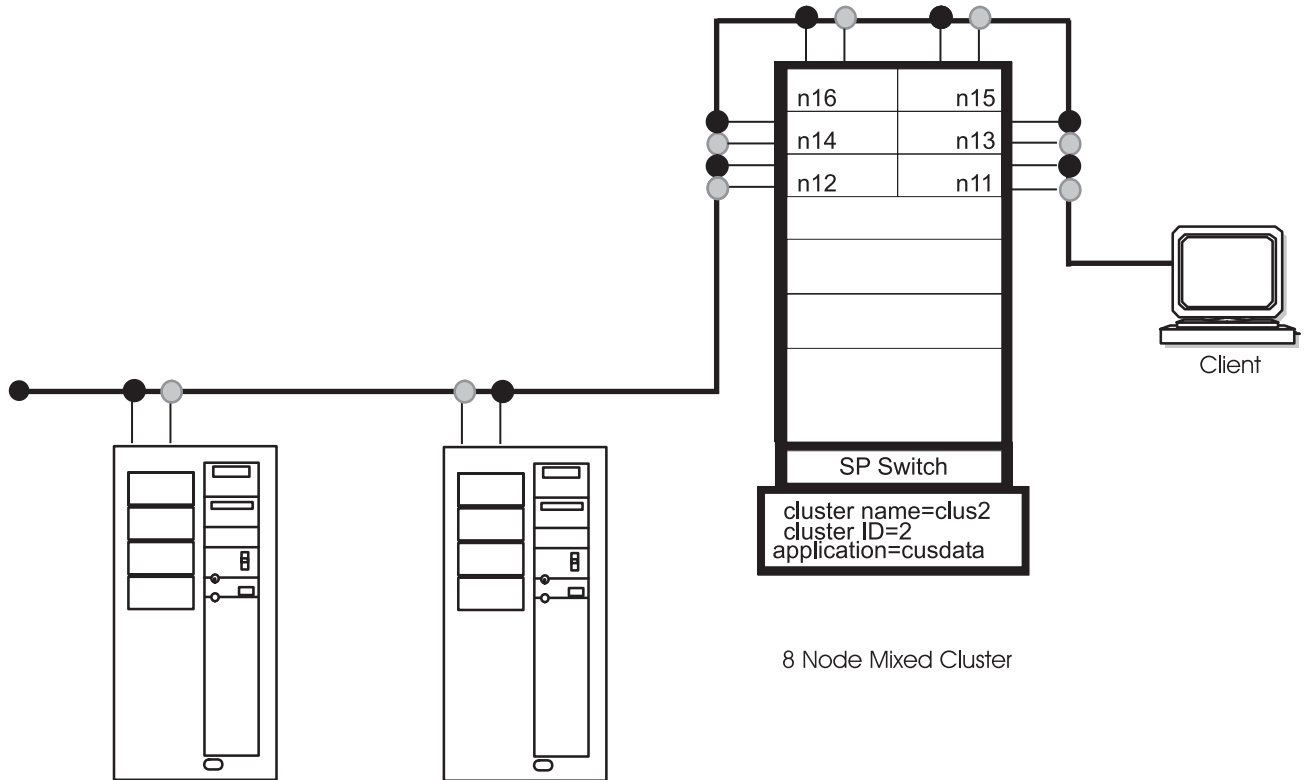
The Initial Cluster Diagram illustration shows a mixed cluster, including a rear view of the SP. The SP diagram uses rectangular boxes to represent the slots supported by the nodes. If your cluster uses thin nodes, darken the outline of the nodes and include two nodes to a drawer. For wide nodes, use the entire drawer. For high nodes, use the equivalent of two wide nodes. Keep in mind that each thin node contains an integrated Ethernet connection. Also indicate if your cluster includes an SP Switch.

Begin drawing this diagram by identifying the cluster name and ID and the applications that are being made highly available. Next, darken the outline of the nodes that will make up the cluster. Include the name of each node. In subsequent chapters, you add information about networks and disk storage subsystems to the diagram.

Initial Cluster Planning
Starting to Draw a Cluster Diagram



11 Node SP Cluster



8 Node Mixed Cluster

Initial Diagram of Two Types of HACMP/ES Clusters

Where You Go From Here

Next you will design the network connectivity of your cluster, described in Chapter 4, Planning Cluster Network Connectivity.

Initial Cluster Planning
Where You Go From Here

Chapter 4 Planning Cluster Network Connectivity

This chapter describes planning the network support for an HACMP/ES cluster.

Prerequisites

In Chapter 3, Initial Cluster Planning, you began planning your cluster, identifying the number of nodes and the key applications you want to make highly available. You started drawing the cluster diagram. This diagram is the starting point for the follow-on planning you will do in this chapter. Also, by now you should have decided whether or not the cluster will use IP address takeover to maintain specific IP addresses.

Overview

In this chapter, you plan the networking support for the cluster. Your primary goal is to use redundancy to design a cluster topology that eliminates network components as potential single points of failure. The following table lists these network components with recommended solutions.

Cluster Network Object	Eliminated as a Single Point of Failure by...
Network	Using multiple networks to connect to nodes
TCP/IP subsystem	Using a serial network to back up TCP/IP
Network adapter	Using redundant network adapters

Working through this chapter entails the following planning tasks:

- Designing the cluster network topology, that is, the combination of networks and point-to-point links that connect your cluster nodes and the number of connections each node has to a single network.
- Completing the network and network adapter planning worksheets. As you fill out these planning worksheets:
 - Assign each network a name, type, attribute (serial, public, or private), and netmask.
 - List the nodes connected by the network.
 - Assign each network adapter a label, function, and IP address.

Note: If you are using the online planning worksheets, you may not be filling out the paper worksheets, but you should still read through the conceptual information in this chapter, in order to make the best planning decisions.

- Add networking to the cluster diagram.

This chapter also includes detailed information about setting up IP address takeover and hardware address swapping.

Designing the Network Topology

HACMP/ES requires that each node in the cluster have a direct network connection with every other node. The software uses these network connections to pass heartbeat messages among the cluster nodes to determine the state of all cluster nodes.

The HACMP/ES software supports an unlimited number of TCP/IP network adapters on each node. Therefore, you have a great deal of flexibility in designing a network configuration. The combination of networks and point-to-point connections that link cluster nodes and clients is called the cluster *network topology*.

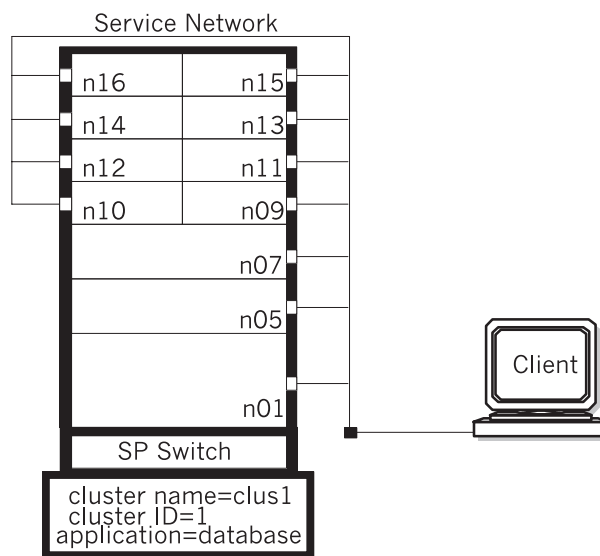
When designing your network topology, keep in mind the following goals:

- Eliminating the network as a single point of failure
- Eliminating the TCP/IP subsystem as a single point of failure
- Eliminating the network adapter as a single point of failure.

The following sections describe recommended approaches for addressing these goals in an HACMP/ES cluster.

Eliminating the Network as a Single Point of Failure

In a single-network setup, each node in the cluster is connected to just one network and has only one service adapter available to clients. In this setup, the service adapter on any node (and the network itself) is a single point of failure. The following diagram shows a single-network configuration.



Single-Network, Single-Adapter Setup

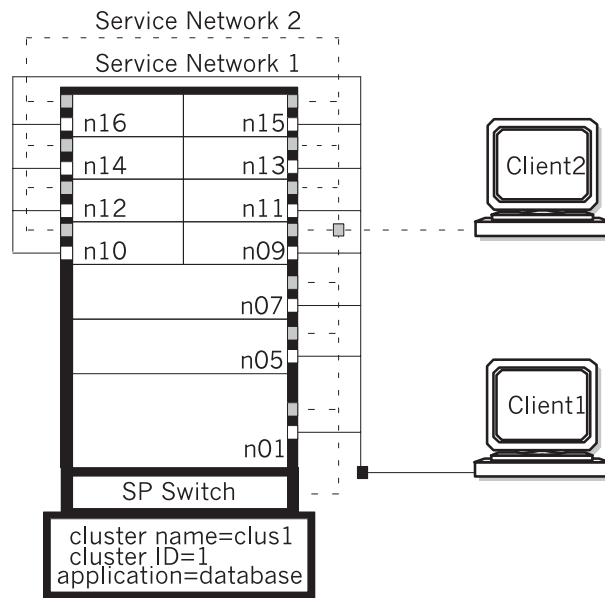
Configuring Multiple Networks

To eliminate the network as a single point of failure, configure a multiple-network setup so that HACMP/ES has multiple paths among cluster nodes. (If you use the SP Switch network, you MUST configure another network. See page 4-5 for more information on the SP Switch.) If one network fails, the remaining networks can still function, connecting nodes and providing access for clients. The more networks you can configure to carry heartbeats and other information among cluster nodes, the greater the degree of system availability.

In some recovery situations, a node connected to two networks may route network packets between networks. In normal cluster activity, however, each network is completely separate—both logically and physically. Logical and physical separation is necessary to eliminate the network as a single point of failure.

Note: Keep in mind that a client, unless it is connected to more than one network, is susceptible to network failure.

The following diagram illustrates a dual-network setup with more than one path to each cluster node.

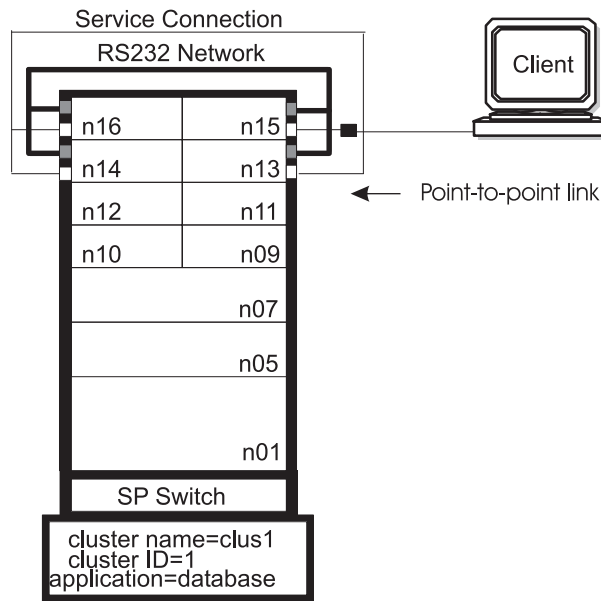


Dual-Network Setup

Configuring Point-to-Point Connections

You can also increase availability by configuring point-to-point connections that directly link cluster nodes.

The following diagram shows a cluster consisting of four thin nodes and a client. A single public network connects the nodes and the client, and the nodes are linked point-to-point by a private RS232 connection that provides an alternate path for heartbeats should the public network fail.



A Point-to-Point Connection

The above figure shows four RS-232 networks.

Configuring Global Networks

You can group multiple HACMP/ES networks of the same type under one global network name. This reduces the probability of network partitions that can cause the cluster nodes on one side of the partition to go down. You should always configure a global network when SP administrative ethernet adapters are included in the HACMP/ES configuration.

On large SP systems, each SP frame of nodes is usually set up as a different subnet on the SP administrative ethernet. Each of these subnets is then defined as a different HACMP/ES network. Defining a global network that includes all these SP administrative ethernets will avoid network partitions.

Supported Network Types

The HACMP/ES software works with the following TCP/IP-based networks:

- Ethernet
- Token-Ring
- Fiber Distributed Data Interchange (FDDI)
- ATM and ATM LAN Emulation.

You cannot use SOCC or SLIP networks in an HACMP/ES cluster.

In addition, there are three SP-specific networks:

- **SP Ethernet**—Every SP frame includes an Ethernet network that connects all SP nodes to the Control Workstation. This network is used for software installation and other administrative purposes; it is available to HACMP/ES and recommended for use as a heartbeat network.
- **SP Switch**—An SP frame can optionally contain a switch network that connects all SP nodes, providing a very, high-speed data connection. The SP Switch supports the TCP/IP protocol and is available to HACMP/ES and client traffic. This network is labeled HPS (however, the old HPS switch is not supported).

Note: Do not use the SP Switch network as your only network. When using the SP Switch, you **MUST** have an additional network defined to HACMP/ES. If not, you will encounter errors when synchronizing cluster topology. The switch uses IP aliasing, so defining the switch network does not update the CuAt ODM database with an HACMP-defined adapter. During synchronization, HACMP/ES looks for entries for adapters in the CuAt ODM database. If it finds none, synchronization will fail.

- **SP Serial network**—Every SP frame includes a serial network that connects all SP nodes to the Control Workstation. This network is used for administrative purposes and is not available for use by HACMP/ES or external client traffic.

Network Planning Considerations

When planning to use multiple networks in your cluster, keep in mind the following:

- Do not route client traffic over the SP Ethernet. This network is used by the SP for administrative traffic, such as netinstalls, which could be affected by heavy client traffic.
- Applications that make use of the SP Switch network must be “tolerant” of brief network outages during switch initialization or switch faults. In general, most TCP/IP applications are not affected by switch initialization. There is no impact to HACMP/ES other than it noticing this behavior.
- The SP Switch is a highly available network. All nodes have four paths to each other through the switch network and fault isolation and recovery is automatic.

Handling Network Failure

Rare failures may result in an SP Switch outage on one or more nodes in the cluster. HACMP/ES distinguishes between two types of network failure, *local* and *global*. A local network failure occurs when a node can no longer communicate over a particular network, but the network is still in use by other nodes. This may happen, for instance, if the cable between one node and the SP switch breaks. A global network failure occurs when all nodes lose the ability to communicate over a network, for example, if the SP Switch itself were to be powered off.

To detect failure and invoke user-defined scripts to verify the failure and recover:

- Use the HACMP/ES network-related event scripts
- Use the AIX error notification strategies and run a script to perform a corrective action

If one of these methods detects a local network failure, the script could issue the **clstop -grsy** command, which would promote the network failure to a node failure. The resources and workload would then fallover to a surviving node.

If one of these methods detects a global network failure, the script could move resources and workload to a reconfigured backup network.

Warning: Do not promote a global network_down failure to node failure. A node failure causes all the nodes to be powered off.

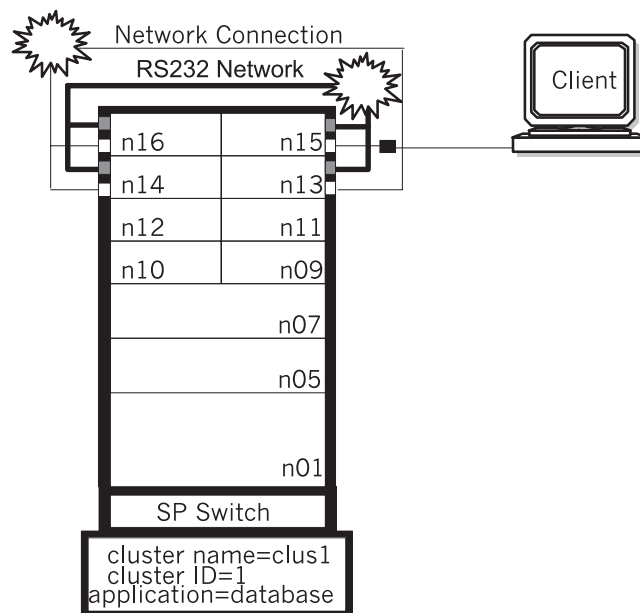
Special SP Switch Failure Considerations

In the event of a SP Switch power off, to the error notification subsystem, the failure looks like an HPS_FAULT9_ERR followed by an HPS_FAULT6_ERR (fault service daemon terminated).

Note that in order to recover from a major switch failure (power off, for example), you must issue **Eclock** and **Estart** commands to bring the switch back on-line. The **Eclock** command runs **rc.switch**, which deletes the aliases HACMP/ES needs for switch IP address takeover. Therefore, it is recommended that if you are configuring IPAT on the SP Switch, you create an event script for either the network_down or the network_down_complete event to add back the aliases for css0.

Eliminating the TCP/IP Subsystem as a Single Point of Failure

Node isolation occurs when all networks connecting two or more parts of the cluster fail, and there is no global network defined. Each group of cluster nodes (one or more) is completely isolated from the other groups. A cluster in which certain groups of nodes are unable to communicate with nodes is considered a *partitioned cluster*.



A Partitioned Cluster

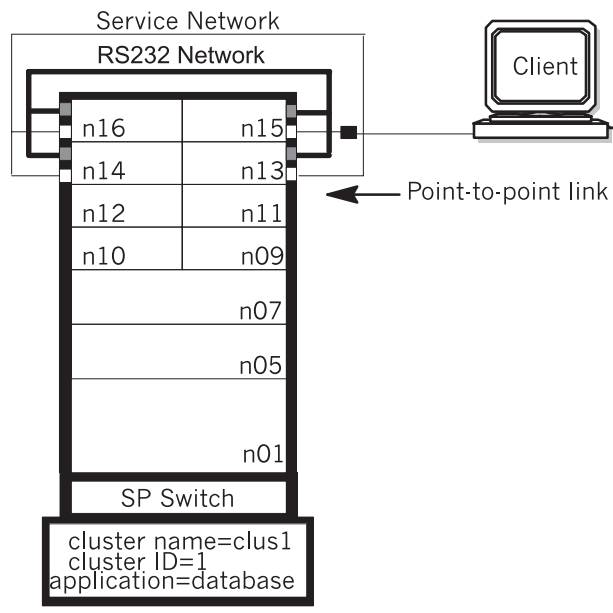
The problem with a partitioned cluster is that each node on one side of the partition interprets the absence of heartbeats from the nodes on the other side of the partition to mean that those nodes have failed; they then generate node failure events for the assumed failed nodes. As the events scripts run, nodes on each side of the cluster (if so configured) try to take over resources from nodes that are still active and that logically own specific resources. Attempted takeovers can cause unpredictable results in the cluster—for example, data corruption resulting from a disk reset.

To guard against TCP/IP subsystem failure and to prevent partitioned clusters, we strongly recommend that each cluster node be connected to its neighboring node by a point-to-point serial network to form a logical “ring.” This logical ring of serial networks reduces the chance of node isolation by allowing neighboring nodes to communicate even when all TCP/IP-based networks fail.

It is important to understand that the serial network does not carry TCP/IP communication between nodes; it only allows nodes to exchange heartbeats and control messages so that Cluster Managers have accurate information about the status of peer nodes.

It is strongly recommended (but not required) that a non-IP network be present between nodes that share resources, to eliminate TCP/IP (inetd) on one node as a single point of failure. At present, target mode SCSI (tm SCSI), target mode SSA (tm SSA), or serial (tty) networks are supported by HACMP/ES.

The following diagram shows a cluster consisting of a single public network connecting the nodes and the client, and the nodes are linked point-to-point by a series of point-to-point RS232 serial connections that provides an alternate path for cluster traffic should the public network fail.



Point-to-Point Network Configuration

Serial Network Planning Considerations

When planning a serial network, keep the following in mind:

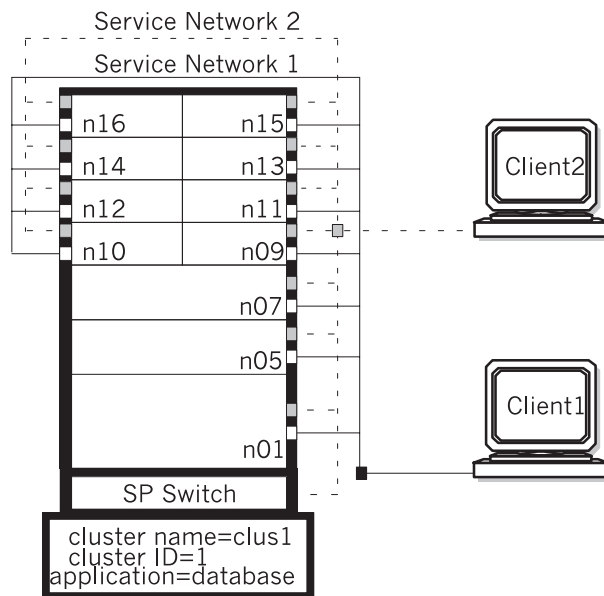
- On the SP thin or wide nodes there are no serial ports available. Therefore, any HACMP/ES configurations that require a tty network need to make use of a serial adapter card (8-port async EIA-232 adapter, FC/2930), available on the SP as an RPQ.
- The 7013-S70, 7015-S70, and 7017-S70 do not support the use of native serial ports in an HACMP/ES RS232 serial network. Configuration of an RS232 serial network in an S70 system requires a PCI multi-port Async card.
- For IBM 7135 disk subsystem and SCSI configurations, tm SCSI or tm SSA can be used with I/O pacing to provide the serial network point-to-point connection.

For information on target mode connections, see Chapter 11, Checking Installed Hardware.

Eliminating Network Adapters as a Single Point of Failure

A network adapter connects a node to a network. When configured with a single adapter per network, the adapter becomes a potential single point of failure. To remedy this problem, configure a node with at least two network adapters for each network to which it connects.

In the following figure, each cluster node has two connections to each network.



Dual-Network, Dual-Adapter Configuration

Network Adapter Functions

When a node is configured with multiple connections to a single network, the adapters serve different functions: one is called the *service* adapter and the other adapter is called the *standby* adapter.

Service Adapter

The service adapter is the primary connection between the node and the network. A node has one service adapter for each physical network to which it connects. The service adapter is used for general TCP/IP heartbeat traffic and is the address the Cluster Information Program (Cinfo) makes known to application programs that want to use cluster services.

Note: The SP switch adapter `css0` base IP address can be configured as a service adapter, as long as IP Address Takeover is not configured for the switch. The base address cannot be modified. See the sections later in this chapter on planning for the SP Switch network.

Note: In configurations using the Classical IP form of the ATM protocol (i.e. *not* ATM LAN Emulation), a maximum of 7 service adapters per cluster is allowed if hardware address swapping is enabled.

Standby Adapter

A standby adapter backs up a service adapter. All client traffic is carried over the service adapter; standby adapters are hidden from client applications and carry only internal HACMP/ES traffic. If a service adapter fails, HACMP/ES swaps the standby adapter's address with the service adapter's address. Using a standby adapter eliminates a network adapter as a single point of failure. A node can have no standby adapter, or it can have from one to seven standby adapters for each network to which it connects. Your software configuration and hardware constraints determine the actual number of standby adapters that a node can support.

Note: In an IP address takeover configuration using the SP switch, standby adapters are not used.

Boot Adapter (Label)—Assigned for IP Address Takeover

IP address takeover is an HACMP/ES facility that allows one node to acquire the network address of another node in the cluster. To enable IP address takeover, a second network address, called the *boot adapter label* (address) must be assigned to the service adapter on each cluster node on which IP address takeover might occur. (For information about defining a boot address for each service adapter on which IP address takeover might occur, see the section Adding a Network Adapter on page 24-6.)

Nodes use the boot label after a system reboot and before the HACMP/ES software is started. When the HACMP/ES software is started on a node, the node's service adapter is reconfigured to use the service label (address) instead of the boot label. If the node should fail, a takeover node acquires the failed node's service address on its standby adapter, making the failure transparent to clients using that specific service address.

For example, if Node A fails, in IP Address Takeover Node B acquires Node A's service address and services client requests directed to that address. Later, when Node A is restarted, it comes up on its boot address and attempts to reintegrate into the cluster on its service address

by requesting that Node B release Node A's service address. When Node B releases the requested address, Node A reclaims it and reintegrates into the cluster. Reintegration, however, would fail if Node A had not been configured to boot using its boot address.

Note: In configurations using rotating resources, the service adapter on the standby node remains on its boot address until it assumes the shared IP address. Consequently, Clinfo makes known the boot address for this adapter.

It is important to realize that the boot address does not use a separate physical adapter, but instead is a second name and IP address associated with a service adapter. All cluster nodes must have this entry in the local `/etc/hosts` file and, if applicable, in the `nameserver` configuration.

Boot/Service/Standby Address Requirements for Resource Groups

The following charts specify the boot, service, and standby address requirements for each resource group configuration in a two- and three-node cluster.

Two-Node Cluster Address Requirements

Resource Group Type	Boot	Service	Standby
Cascading without IP address takeover <i>(same for Cascading without Fallback)</i>	none required	2	2
Cascading with IP address takeover <i>(same for Cascading without Fallback)</i>	1 per highly available service address	1 per highly available client connection	2
Rotating	2	1 per node per network minus 1	2
Concurrent	not supported	not supported	not supported

Three-Node Cluster Address Requirements

Resource Group Type	Boot	Service	Standby
Cascading without IP address takeover <i>(same for Cascading without Fallback)</i>	none required	3	3
Cascading with IP address takeover <i>(same for Cascading without Fallback)</i>	1 per highly available service address	1 per highly available client connection	3
Rotating	3	1 per node per network minus 1	3
Concurrent	not supported	not supported	not supported

IP Address Takeover Not Supported on the SP Administrative Ethernet

Do not configure IP Address Takeover on the SP Ethernet network adapter. The SP software assumes that an IP address on the SP administrative Ethernet (en0 adapter) is associated with a specific node. You should configure this adapter so the network will be monitored by HACMP/ES (this is done by defining the SP ethernet IP address as a service address to HACMP). In addition, no owned or takeover resources can be associated with this adapter.

Planning for the SP Switch Network

There are a number of considerations specific to the SP Switch network, especially regarding IP Address Takeover configuration. These issues are discussed in the following sections.

Configuring the SP Switch Base IP Address as Service Adapter

The SP Switch adapter css0 base IP address can be used as a service adapter as long as IP Address Takeover is *not* configured for the switch. It must not be used in an IP Address Takeover configuration.

The base address cannot be modified.

SP Switch EPrimary Management

HACMP/ES 4.2.1 and 4.2.2 allowed either the HPS (older version) or the SP (newer version) of the switch. HACMP/ES 4.3.0 and later versions support only the SP switch.

The Eprimary node is the designated node for the switch initialization and recovery.

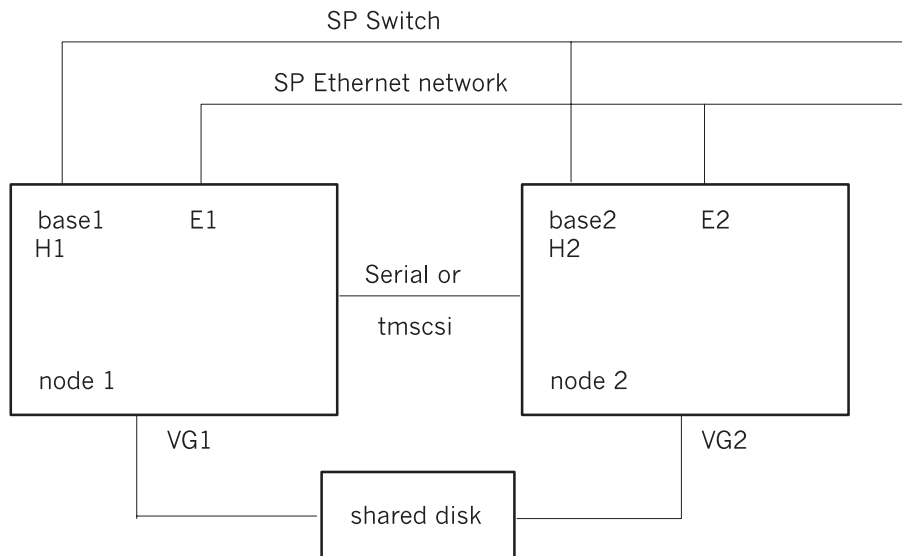
The SP switch cannot be managed by HACMP/ES. The SP switch can configure a secondary Eprimary and reassign the Eprimary automatically on failure. This is handled by the SP software, outside of HACMP/ES.

Handling SP Switch Adapter Failure

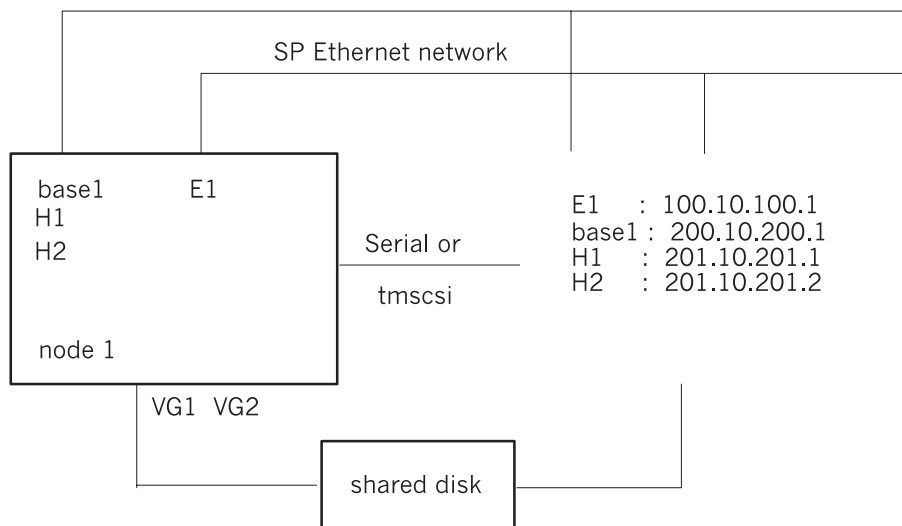
You cannot configure a standby adapter for the SP Switch Network. To eliminate this as a cluster single point of failure, use the AIX Error Notification facility to promote an SP Switch adapter failure to a node failure.

Planning for IP Address Takeover with the SP Switch Network

The SP Switch is the first network to make use of IP address aliasing with HACMP/ES to permit IP address takeover (IPAT). See the figure below for a general illustration of IP address takeover on the SP Switch. In the figure, node 1 and node 2 are SP nodes. E1 and E2 are administrative ethernet IP addresses. The base1 and base2 labels reflect SP Switch base IP addresses. H1 and H2 are SP Switch alias HACMP/ES service addresses. VG1 and VG2 are volume groups.



Cluster after node 2 fails:



Sample Two-Node Cluster Configuration on the SP Machine.

Note: The HACMP/ES boot addresses are not included in the above figure. The boot addresses are also aliases; they are not the same as SP switch base IP addresses.

IP Address Takeover Considerations with the SP Switch

Take these considerations into account when planning IP address takeover on the SP switch:

- To use IP Address Takeover on an SP Switch network, you must enable Address Resolution Protocol (ARP) for the SP Switch network. You can configure this with an SP customize operation, or during initial SP setup. The network type is **hps**.
See the section on ARP below for more information.
- Standby adapters/addresses are not used for SP Switch IP address takeover.
- HACMP/ES boot and service addresses are alias addresses on the SP Switch `css0` interface. The `css0` interface can have more than one alias IP address; therefore, it can support IP takeover addresses. At present, only one boot address can be defined per SP Switch `css0` interface, and up to seven takeover addresses. These HACMP/ES “alias HPS service addresses” appear as “ifconfig alias” addresses on the `css0` interface.
- SP switch boot and service addresses must be different from the `css0` base IP address in order to use IP address takeover.
- The SP Switch alias addresses for IP address takeover can be configured as a part of a cascading or rotating resource group.

Note: In the case of a major SP Switch failure, the aliases HACMP/ES needs for switch IP address takeover may be deleted when the **Eclock** command runs **rc.switch**. For this reason, if you are configuring IPAT with the SP Switch, you should create an event script for either the `network_down` or the `network_down_complete` event to add back the aliases for `css0`.

SP Switch Address Resolution Protocol (ARP)

If your SP nodes are already installed and the switch network is up on all nodes, you can verify whether ARP is enabled. On the control workstation, enter the following command:

```
dsh -av "/usr/lpp/ssp/css/ifconfig css0"
```

If NOARP appears as output from any of the nodes, you must enable ARP to use IP takeover on the SP Switch. ARP must be enabled on all SP nodes connected to the SP Switch.

Warning: Before you perform the following steps, be sure to back up **CuAt**. If user error causes **CuAt** to become corrupt, the SP nodes may be corrupted and will have to be re-installed. You will need to copy your backup of **CuAt** to `/etc/objrepos/CuAt` prior to rebooting the system. Be careful! If you feel this is too risky, customize the nodes to turn ARP on (see the *SP Administration Guide* for help with this procedure).

To enable ARP on all the nodes, follow these steps carefully. Enter all commands from the control workstation. Ensure all nodes are up. The quotation marks shown in the commands must be typed.

1. Create a copy of the **CuAt** file on all nodes:

```
dsh -av "cp /etc/objrepos/CuAt /etc/objrepos/CuAt.save"  
dsh -av "odmget -q 'name=css and attribute=arp_enabled' CuAt |  
sed s/no/yes/ > /tmp/arpon.data"  
dsh -av "odmchange -o CuAt -q'name=css and attribute=arp_enabled'  
/tmp/arpon.data"
```

2. Verify that the previous commands worked:

```
dsh -av "odmget -a 'name=css and name=arp_enabled' CuAt | grep value"  
You should see an entry reporting "value=yes" from every node.
```

3. Remove the temporary file from all nodes:

```
dsh -av rm /tmp/arpon.data
```

4. Shut down and reboot the nodes:

```
dsh -av "shutdown -Fr"
```

Using HACMP/ES with NIS and DNS

HACMP/ES facilitates communication between the nodes of a cluster so that each node can determine if the designated services and resources are available. Resources can include, but are not limited to, IP addresses and names and storage disks.

Subnetting the service and standby adapters (using TCP/IP) and having at least two separate networks (Ethernet and RS232 for example) permits HACMP/ES to determine if a communication problem exists in the adapter, the network, or another common component such as TCP/IP software itself. The ability to subnet the adapters insures that HACMP/ES can direct the keepalive traffic from service to standby, standby to service adapters, standby to standby, and so on. For example, if there are two nodes in a cluster, and both standby and service adapters of Node A can receive and send to the standby adapter of node B, but cannot communicate with the service adapter of Node A, then it can be assumed that the service adapter is not working properly. HACMP/ES will recognize this and perform the swap between the service and standby adapters.

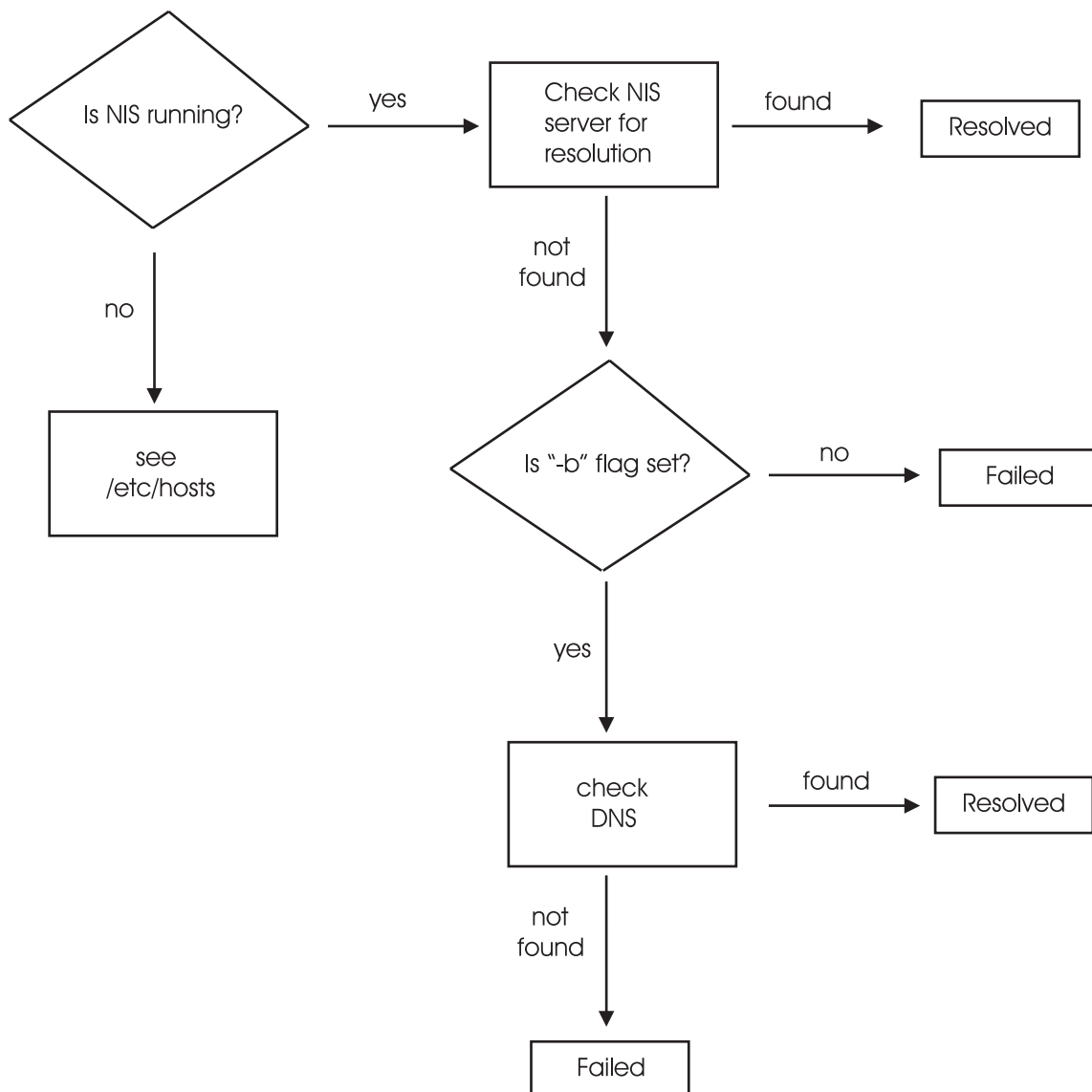
Some of the commands used to perform the swap require IP lookup. This defaults to a nameserver for resolution if NIS or DNS is operational. If the nameserver was accessed via the adapter that is down, the request will timeout. To insure that the cluster event (in this case an adapter swap) completes successfully and quickly, HACMP/ES disables NIS or DNS hostname resolution. It is therefore required that the nodes participating in the cluster have entries in the **/etc/hosts** file.

How HACMP/ES Enables and Disables Nameserving

This section provides some additional details on the logic a system uses to perform hostname resolution and how HACMP/ES enables and disables DNS and NIS.

Note: Using HACMP/ES with nameserving requires that the “Host uses NIS or Name Server” Run Time Parameter be set to **true**. Refer to Changing a Node’s Run-Time Parameters on page 28-5 in *Volume 2* for more information.

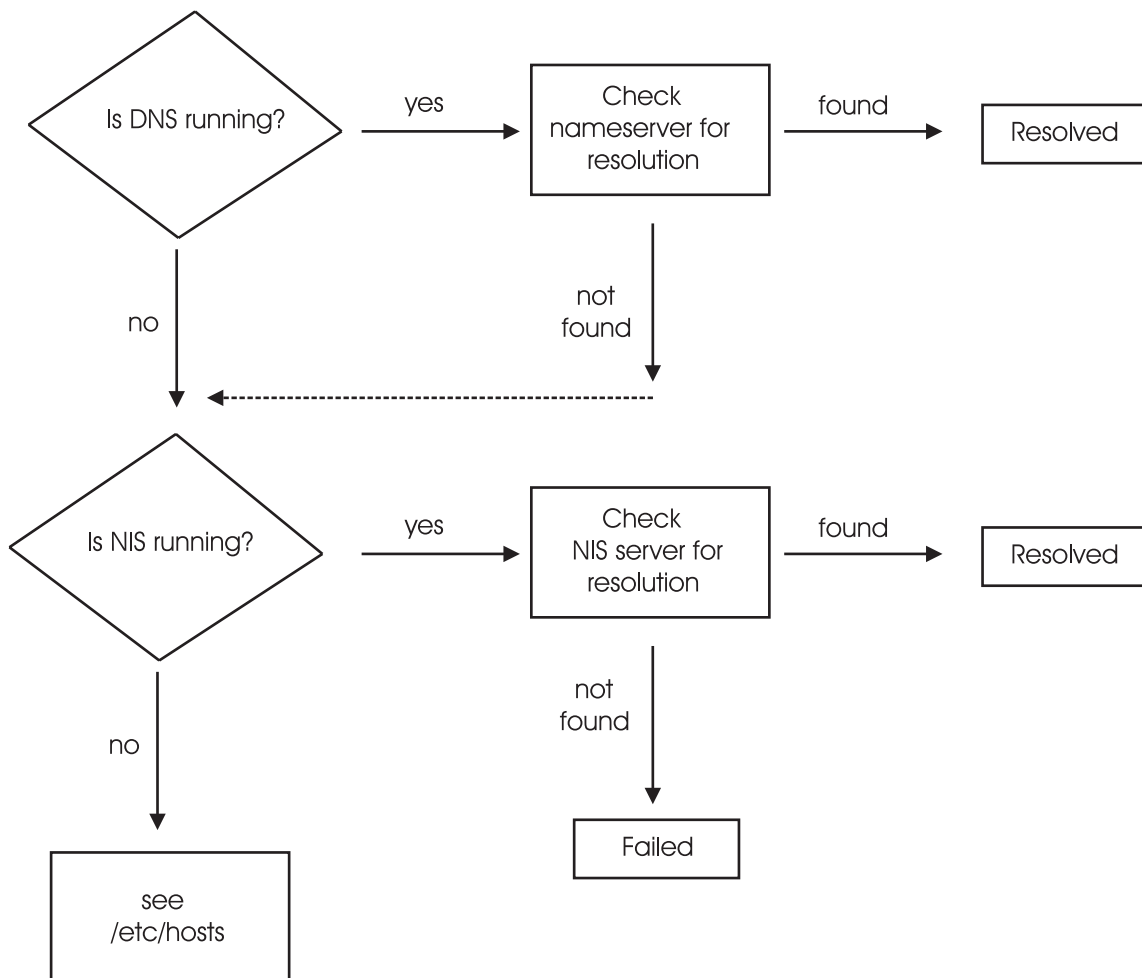
If a node is using either domain nameserving or NIS, the hostname will normally be resolved by contacting a suitable server. This will, at best, cause a time delay, and at worst, never return a response because communication to the server is lost. For example, if NIS alone is running, hostname resolution is performed via the following logic in AIX:



Note: The following applies if the NIS configuration maps include the host map.

As shown, if the NIS host tables were built with the **-b** flag, the NIS Server will continue its search via domain nameserving if that option is available. The key point, however, is that under certain circumstances (for example, an adapter being down, and the NIS server being unavailable), hostname resolution should be localized for quick response time. This infers that the local **/etc/hosts** file should be consulted, and the systems that may be involved in a cluster reconfiguration must be placed within the file in addition to loopback and the local hostname. It also means that the client portion of NIS that is running on that machine must be turned off. This is the case if the cluster node is an NIS client regardless of its status as a server (master or slave). Remember, even an NIS server uses the **ypbind** to contact the **ypserv** daemon running either on the master server or a slave server to perform the lookup within the current set of NIS maps.

Similarly, the logic followed by DNS (AIX) is:



In this situation, if both DNS and NIS were not disabled, a response to a hostname request might be as long as the time required for both NIS and DNS to timeout. This would hinder HACMP's system reconfiguration, and increase the takeover time required to maintain the availability of designated resources (IP address). Disabling these services is the only reliable and immediate method of expediting the communication changes HACMP/ES is attempting to accomplish.

The method HACMP/ES uses to cleanly start and stop NIS and DNS is found within the scripts **cl_nm_nis_on** and **cl_nm_nis_off** within the **/usr/es/sbin/cluster/events/utills** directory.

For NIS, checking the process table for the **ypbind** daemon lets HACMP/ES know that NIS is running, and should be stopped. A check for the existence of the file **/usr/es/sbin/cluster/hacmp_stopped_ypbind** lets HACMP/ES know that NIS client services must be restarted following the appropriate cluster configuration events. As mentioned earlier, the commands **startsrc -s ypbind** and **stopsrc -s ypbind** are used to start and stop this node from using NIS name resolution. This is effective whether the node is a master or slave server (using NIS client services), or simply an NIS client.

HACMP/ES uses the AIX command **namerslv** to stop and start DNS on the cluster node. After checking for the existence of an **/etc/resolv.conf** file (the method for determining if domain name resolution is in effect), HACMP/ES uses the **namerslv -E** command to stop nameserving by moving the **/etc/resolv.conf** file to a specified file. After reconfiguration, HACMP/ES runs the **namerslv -B** command to restore the domain name configuration file from the specified file. Using this method, it is not necessary to stop and restart the **named** daemon.

Planning for Cluster Performance

HACMP/ES 4.4 provides easier and greater control over several tuning parameters that affect the cluster's performance. Setting these tuning parameters correctly to ensure throughput and adjusting the HACMP failure detection rate can help avoid failures caused by heavy network traffic.

Cluster nodes sometimes experience extreme performance problems, such as large I/O transfers, excessive error logging, or lack of memory. When this happens, the Cluster Manager can be starved for CPU time. It might not reset the deadman switch within the time allotted. Misbehaved applications running at a priority higher than the cluster manager can also cause this problem.

The deadman switch is the AIX kernel extension that halts a node when it enters a hung state that extends beyond a certain time limit. This enables another node in the cluster to acquire the hung node's resources in an orderly fashion, avoiding possible contention problems. If the deadman switch is not reset in time, it can cause a system panic and dump under certain cluster conditions.

Setting these tuning parameters correctly may avoid some of the performance problems noted above. You can set these parameters using HACMP/ES SMIT screens:

- High and low watermarks for I/O pacing
- **syncd** frequency rate
- HACMP Failure Detection Rate (Custom)
 - HACMP cycles to failure
 - HACMP heartbeat rate.

I/O Pacing

AIX users have occasionally seen poor interactive performance from some applications when another application on the system is doing heavy input/output. Under certain conditions I/O can take several seconds to complete. While the heavy I/O is occurring, an interactive process can be severely affected if its I/O is blocked or if it needs resources held by a blocked process.

Under these conditions, the HACMP/ES software may be unable to send keepalive packets from the affected node. The Cluster Managers on other cluster nodes interpret the lack of keepalives as node failure, and the I/O-bound node is “failed” by the other nodes. When the I/O finishes, the node resumes sending keepalives. Its packets, however, are now out of sync with the other nodes, which then kill the I/O-bound node with a RESET packet.

You can use I/O pacing to tune the system so that system resources are distributed more equitably during high disk I/O. You do this by setting high- and low-water marks. If a process tries to write to a file at the high-water mark, it must wait until enough I/O operations have finished to make the low-water mark.

By default, AIX is installed with high- and low-water marks set to **zero**, which disables I/O pacing.

Though enabling I/O pacing may have only a slight performance effect on very I/O intensive processes, it is required for an HACMP/ES cluster to behave correctly during large disk writes. If you anticipate heavy I/O on your HACMP/ES cluster, you should enable I/O pacing.

Although the most efficient high- and low-water marks vary from system to system, an initial high-water mark of **33** and a low-water mark of **24** provides a good starting point. These settings only slightly reduce write times and consistently generate correct fallover behavior from the HACMP/ES software.

See the *AIX Performance Monitoring & Tuning Guide* for more information on I/O pacing.

Syncd Frequency

The **syncd** setting determines the frequency with which the I/O disk-write buffers are flushed. Frequent flushing of these buffers reduces the chance of deadman switch time-outs.

The AIX default value for **syncd** as set in `/sbin/rc.boot` is 60. It is recommended to change this value to 10. Note that the I/O pacing parameter setting should be changed first. You should not need to adjust this parameter again unless you get frequent time-outs.

Failure Detection Parameters

Each supported cluster network in a configured HACMP/ES cluster has a corresponding cluster network module. Each network module monitors all I/O to its cluster network.

Each network module maintains a connection to other network modules in the cluster. The Cluster Managers on cluster nodes send messages to each other through these connections. Each network module is responsible for maintaining a working set of service adapters and for verifying connectivity to cluster peers. The network module also is responsible for reporting when a given link actually fails. It does this by sending and receiving periodic heartbeat messages to or from other network modules in the cluster.

Currently, HACMP/ES network modules support communication over the following types of networks:

- Serial (RS232)
- Target-mode SCSI
- Target-mode SSA
- IP
- Ethernet
- Token-Ring
- FDDI
- SP Switch
- ATM.

The Failure Detection Rate is made up of two components:

- *cycles to fail (cycle)*: the number of heartbeats that must be missed before detecting a failure
- *heartbeat rate (hbrate)*: the number of microseconds between heartbeats.

The time needed to detect a failure can be calculated using this formula:

$$(\text{heartbeat rate}) \times (\text{cycles to fail}) \times 2 \text{ seconds}$$

The default failure detection rate is usually optimal, though speeding up or slowing down failure detection is a small, but potentially significant area where you can adjust cluster fallover behavior. However, the amount and type of customization you add to event processing has a much greater impact on the total fallover time. You should test the system for some time before deciding to change the failure detection speed of any network module.

If HACMP/ES cannot get enough CPU resources to send heartbeats on IP and serial networks, other nodes in the cluster will assume the node has failed, and initiate takeover of the node's resources. In order to ensure a clean takeover, the deadman switch crashes the busy node if it is not reset within a given time period. The deadman switch uses the following formula:

$$N = ((\text{keepalives} * \text{missed_keepalives}) - 1)$$

Where *keepalives* and *missed_keepalives* are for the *slowest* network in the cluster.

The table below shows the Deadman Switch Timeout for each network. Each of these times is one second less than the time to trigger an HACMP event on that network. Remember that the deadman switch is triggered on the slowest network in your cluster.

NETWORK	SLOW	NORMAL	FAST
ATM	63	31	15
Ethernet	11	5	4
FDDI	11	5	4
Token-Ring	11	5	4
RS232	17	11	5
SP Switch	63	15	7
TMSSA	17	11	5
TMSCSI	17	11	5

Deadman Switch Timeouts in Seconds Per Network

If you decide to change the failure detection rate of a network module, keep the following considerations in mind:

- Failure detection is dependent on the *fastest* network linking two nodes.
- The failure rate of networks varies, depending on their characteristics.
- Before altering the network module, you should give careful thought to how much time you want to elapse before a real node failure is detected by the other nodes and the subsequent takeover is initiated.
- Faster heartbeat rates may lead to false failure detections, particularly on busy networks. For example, bursts of high network traffic may delay heartbeats and this may result in nodes being falsely ejected from the cluster. Faster heartbeat rates also place a greater load on networks. If your networks are very busy and you experience false failure detections, you can try slowing the failure detection speed on the network modules to avoid this problem.
- It is recommended that you first change the Failure Detection Rate from normal to slow (or fast) before trying to customize this rate.
- The Failure Detection Rate should be set equally for every network module used by the cluster. The change must be synchronized across cluster nodes. The new values will become active when you do a topology DARE.

See Chapter 18, Configuring an HACMP/ES Cluster for more details on configuring these parameters.

Completing the Network Worksheets

In the following sections you fill out the following network planning worksheets:

- TCP/IP Networks Worksheet
- TCP/IP Network Adapter Worksheet
- Serial Network Worksheet
- Serial Adapter Worksheet

Appendix A includes blank copies of these worksheets which you can copy for your own use along with examples of completed TCP/IP network worksheets.

Completing the TCP/IP Networks Worksheet

The TCP/IP Networks Worksheet helps you organize the networks for an HACMP/ES cluster. To complete the worksheet:

Enter the cluster ID in the **Cluster ID** field and the cluster name in the **Cluster Name** field. This information was determined in Chapter 3, Initial Cluster Planning.

Assigning a Name to Each Network

In the **Network Name** field, give each network a symbolic name. You use this value during the install process when you configure the cluster. This name can be up to 31 characters long and can include alphanumeric characters and underscores.

The *network name* is a symbolic value that identifies a network in an HACMP/ES environment. Cluster processes use this information to determine which adapters are connected to the same physical network. The network name is arbitrary, but must be used consistently. If several adapters share the same physical network, make sure that you use the same network name when defining these adapters.

Indicating the Type of Network

Indicate the network's type in the **Network Type** field. For example, it may be an Ethernet, a Token-Ring, and so on.

Note: For SP Switch networks, ARP must be enabled, and the network type must be **hps**.

Determining Each Network's Attribute

Indicate the network's function in the **Network Attribute** field. A TCP/IP network's attribute is either *public* or *private*.

A public network connects from two to 32 nodes and allows clients to access cluster nodes. Ethernet, Token-Ring, and FDDI are considered public networks.

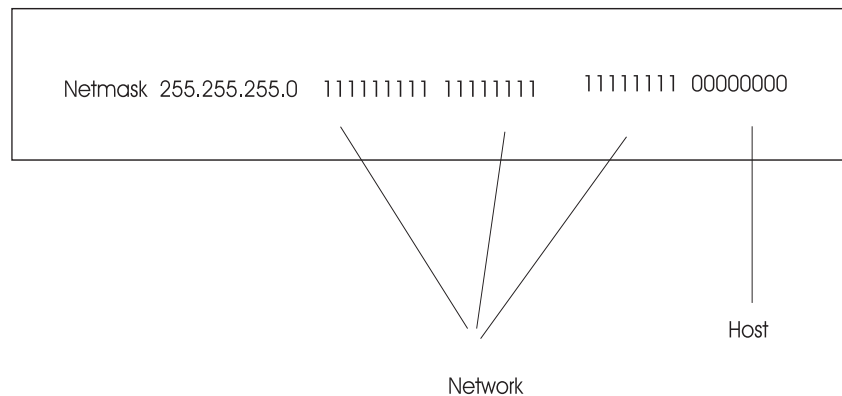
A private network provides point-to-point communication between two nodes; it normally does not allow client access. Exceptions to this rule include ATM and the SP Switch network. If an SP node is used as a client, the SP Switch network can allow client access. An ATM network does allow client connections and may contain standby adapters.

Determining the Netmask of the Network

In the **Netmask** field, provide the network mask of each network. The network mask is site dependent.

The HACMP/ES software uses the subnet feature of TCP/IP to divide a single physical network into separate logical subnets. To use subnets, you must define a network mask for your system.

An IP address consists of 32 bits. Some of the bits form the network address; the remainder form the host address. The *network mask* (or netmask) determines which bits in the IP address refer to the network and which bits refer to the host, as shown in the following example:



In the preceding figure, the netmask is shown in both dotted decimal and binary format. A binary 1 indicates that a bit is part of the network address. A binary 0 indicates that a bit is part of the host address. In the example above, the network portion of the address occupies 24 bits; the host portion occupies 8 bits. It is convenient (but not necessary) to define a subnet on an octet boundary.

Subnetting is relevant only for the local site. Remote sites view the network portion of the IP address by the network's class.

See Chapter 14 of the *IBM AIX Communication Concepts and Procedures* manual for further information about address classes. Also, ask your network administrator about the class and subnets used at your site.

Subnet Considerations for Cluster Monitoring with Tivoli

If you plan to monitor an HACMP node with your Tivoli management software, and you do not have a dedicated network for the Tivoli Management Region (TMR), you must create an IP address alias in order to ensure the proper monitoring of IP address takeover. You must place this alias in the **/etc/hosts** file.

The subnet of this alias must be *different* than the node's service and standby adapters, and the *same* as the subnet of the Tivoli Management Region (server) node.

Here is an example of what you might insert into the **/etc/hosts** file for a Tivoli-monitored cluster node named HAnode and a Tivoli server node named TMRnode. HAnode has service, standby, and alias IP addresses; TMRnode simply has a service IP address.

The netmask for this example network is 255.255.255.0

Adapter Label	Address
HAnode_svc	10.10.20.88
HAnode_stby	10.50.25.88
HAnode_alias	10.50.21.89
TMRnode	10.50.21.10

You can see in this example that the alias address and the TMR address are on the same subnet, and this subnet is *in addition to* the two already used for the cluster node's service and standby adapters.

For complete details on setting up cluster monitoring with Tivoli, see Appendix D, Installing and Configuring Cluster Monitoring with Tivoli.

List the Nodes Connected by Each Network

In the **Node Names** field, list the names of the nodes connected to each network. Refer to your cluster diagram.

Completing the TCP/IP Network Adapter Worksheet

On the TCP/IP Network Adapter Worksheet in Appendix A you define the network adapters connected to each node in the cluster. Complete the following steps for each node on a separate worksheet:

Enter the node name in the **Node Name** field. You assigned this name in Chapter 3, Initial Cluster Planning. For each node, perform the following tasks for each network adapter configured on the node.

Record the Network Adapter Interface Name

You will enter a value in the **Interface Name** field after you configure the adapter following the instructions in Chapter 18, Configuring an HACMP/ES Cluster. AIX assigns an interface name to the adapter when it is configured. The interface name is made up of two or three characters that indicate the type of adapter, followed by a number which AIX assigns in sequence for each adapter of a certain type. For example, AIX assigns an interface name such as *en0* for the first Ethernet adapter it configures, *en1* for the second Ethernet adapter it configures, and so on.

Assign a Label to Each Network Adapter

Assign each adapter a symbolic name and record it in the **Adapter IP Label** field. This name can be up to 31 characters long and can include alphanumeric characters, hyphens, and underscores.

For TCP/IP networks, the adapter label is the name in the */etc/hosts* file associated with a specific IP address. Thus, a single node can have several adapter labels and IP addresses assigned to it.

When choosing adapter labels, adopt a naming convention that helps identify the adapter's role in the cluster and be consistent. For example, an adapter label can contain the network interface name, such as *nodea_en0* and *nodea_en1*. You can also choose a naming convention that indicates each adapter's function, such as *nodea_svc*, *nodea_stdby* or *nodea_boot*.

For example, the following are two entries from an **/etc/hosts** file for the service and standby adapters:

```
100.100.50.1  nodea_svc
100.100.51.1  nodea_stdby
```

The adapter labels, however, should not be confused with the hostname. In a non-HACMP/ES environment, a node typically is identified by a hostname associated with a network interface. If a node has only one network interface, the hostname usually also uniquely identifies the node. But in an HACMP/ES cluster environment, a node typically can have more than one network interface, and this hostname-to-network interface association does not uniquely identify the node. Instead, a node is identified by a unique node name. An IP label association with node adapters uniquely identifies the host. At any given time, the hostname corresponds to only one of the node's IP labels.

Identify the Adapter's Function

Identify the adapter's function as service, standby, or boot in the **Adapter Function** field.

Assign an IP Address to Each Adapter

Assign an IP address for each adapter in the **Adapter IP Address** field.

Pay careful attention to the IP addresses you assign to standby adapters. Standby adapters must be on a separate logical subnet from the service adapters, even though they are on the same physical network. To place a standby adapter on a different subnet from a service adapter, give it an IP address that has a different network address portion. Placing the standby adapters on a different subnet from the service adapter limits the choices available to TCP/IP and allows you to determine which adapter TCP/IP will use to send a packet to a network.

If you configure multiple standby adapters on cluster nodes, they all must be configured on the same subnet to handle cluster node failures. In addition, keep in mind that with multiple standby adapters configured, a **swap_adapter** event on the standby adapter routing heartbeats on a node may cause all standbys on the node to fail, since only one heartbeat route exists per node for the standbys on that node.

If you are using IP address takeover or rotating resources, enter in the **Boot Address** field the boot address for each service address that can be taken over. The boot address and the service address must be on the same subnet. That is, the two addresses must have the same value for the network portion of the address; the host portion must be different. Use a standard formula for assigning boot addresses. For example, the boot address could be the host address plus 64. This formula yields the boot addresses shown below:

```
NODE A service address 100.100.50.135
NODE A boot address:   100.100.50.199
Network mask:         255.255.255.0
```

Record the Network Name and Attribute

Enter the name of the network to which this network adapter is connected in the **Network Name** field and the network attribute in the **Network Attribute** field. Refer to the TCP/IP Networks Worksheet for this information.

Enter the Adapter Hardware Address

If you are using hardware address swapping, enter in the **Adapter HW Address** field the alternate hardware address for each service address that has a boot address. The hardware address is a 12 or 14-digit hexadecimal value. Usually, hexadecimal numbers are prefaced with “0x” (zero x) for readability. *Do not use colons to separate the numbers in the adapter hardware address.* For more information, see the section on Defining Hardware Addresses on page 4-26.

Completing the Serial Networks Worksheet

The Serial Networks Worksheet in Appendix A helps you organize the networks for an HACMP/ES cluster. To complete the worksheet:

Enter the Cluster ID and the Cluster Name

Enter the cluster ID in the **Cluster ID** field and the cluster name in the **Cluster Name** field. You determined these values in Chapter 3, Initial Cluster Planning.

Assign a Name to the Serial Network

In the **Network Name** field, give each network a symbolic name. Names cannot exceed 31 characters. They can include alphanumeric characters and underscores. You use this value during the install process when you configure the cluster.

The *network name* is a symbolic value that identifies a network in an HACMP/ES environment. Cluster processes use this information to determine which adapters are connected to the same physical network. The network name is arbitrary, but must be used consistently. If several adapters share the same physical network, make sure that you use the same network name when defining these adapters.

Record the Network Type and Attribute

Indicate the network’s type in the **Network Type** field. For serial networks, you can specify RS232, Target Mode SCSI (tmssi), or Target Mode SSA (tmssa). The **Network Attribute** field is pre-filled with the only choice for serial networks.

List the Nodes That the Serial Network Connects

In the **Node Names** field, list the names of the nodes connected to each network. Refer to your cluster diagram.

Completing the Serial Network Adapter Worksheet

The Serial Network Adapter Worksheet in Appendix A lets you define the network adapters connected to each node in the cluster. Complete the following steps for each node on a separate worksheet:

Enter the node name in the **Node Name** field. You assigned this name in Completing the TCP/IP Networks Worksheet on page 4-21. Record the following information for each serial network configured on the node.

Enter the Slot Number and Interface Name of the Serial Adapter

Enter the number of the slot in which the serial adapter is located in the **Slot Number** field.

You will enter a value in the **Interface Name** field after you configure the adapter following the instructions in Chapter 18, Configuring an HACMP/ES Cluster. AIX assigns an interface name to the adapter when it is configured. The interface name is made up of two or three characters that indicate the type of adapter, followed by a number which AIX assigns in sequence for each adapter of a certain type. For example, AIX assigns an interface name such as en0 for the an Ethernet adapter.

Assign a Name to the Serial Adapter

Assign each adapter a symbolic name and record it in the **Adapter Label** field.

When choosing a serial adapter label, be consistent with the naming convention that helps identify the adapter's role in the cluster and be consistent. For example, if your naming convention indicates each adapter's function, such as *nodea_svc* and *nodea_stdby*, assign a name such as *nodea_tty1*.

Record the Network Name, Attribute, and Function

In the **Network Name** field, enter the name you assigned to the network in the Serial Network Worksheet.

The **Network Attribute** and **Adapter Function** fields are pre-filled with appropriate values.

Defining Hardware Addresses

Note: You cannot use hardware address swapping on the SP Ethernet or SP Switch networks.

The hardware address swapping facility works in tandem with IP address takeover. Hardware address swapping maintains the binding between an IP address and a hardware address, which eliminates the need to flush the ARP cache of clients after an IP address takeover. This facility is supported for Ethernet, Token-Ring, FDDI and ATM adapters.

Note that hardware address swapping takes about 60 seconds on a Token-Ring network, and up to 120 seconds on a FDDI network. These periods are longer than the usual time it takes for the Cluster Manager to detect a failure and take action.

Selecting an Alternate Hardware Address

This section provides hardware addressing recommendations for Ethernet, Token Ring, FDDI, and ATM adapters. Note that any alternate hardware address you define for an adapter should be similar in form to the default hardware address the manufacturer assigned to the adapter.

To determine an adapter's default hardware address, use the **netstat -i** command (when the networks are active).

Using netstat

To retrieve hardware addresses using the **netstat -i** command, enter:

```
netstat -i | grep link
```

which returns output similar to the following:

lo0	16896	link#1		186303	0	186309	0	0
en0	1500	link#2	2.60.8c.2f.bb.93	2925	0	1047	0	0
tr0	1492	link#3	10.0.5a.a8.b5.7b	104544	0	92158	0	0
tr1	1492	link#4	10.0.5a.a8.8d.79	79517	0	39130	0	0
fi0	4352	link#5	10.0.5a.b8.89.4f	40221	0	1	1	0
fi1	4352	link#6	10.0.5a.b8.8b.f4	40338	0	6	1	0
at0	9180	link#7	8.0.5a.99.83.57	54320	0	8	1	0
at2	9180	link#8	8.0.46.22.26.12	54320	0	8	1	0

Specifying an Alternate Ethernet Hardware Address

To specify an alternate hardware address for an Ethernet interface, begin by using the first five pairs of alphanumeric characters as they appear in the current hardware address. Then substitute a different value for the last pair of characters. Use characters that do not occur on any other adapter on the physical network.

For example, you could use 10 and 20 for node A and node B, respectively. If you have multiple adapters for hardware address swapping in each node, you can extend to 11 and 12 on node A, and 21 and 22 on node B.

Specifying an alternate hardware address for adapter interface en0 in the output above thus yields the following value:

Original address 02608c2fbb93

New address 02608c2fbb10

To define this alternate hardware address to the cluster environment, see Defining Adapters on page 18-2.

Specifying an Alternate Token-Ring Hardware Address

To specify an alternate hardware address for a Token-Ring interface, set the first two digits to **42**, indicating that the address is set locally.

Specifying an alternate hardware address for adapter interface tr0 in the output above thus yields the following value:

Original address 10005aa8b57b

New address 42005aa8b57b

To define this alternate hardware address to the cluster environment, see the section on Defining Adapters on page 18-2.

Specifying an Alternate FDDI Hardware Address

To specify an alternate FDDI hardware address, enter the new address into the **Adapter Hardware Address** field as follows, *without any decimal separators*:

1. Use 4, 5, 6, or 7 as the first digit (the first nibble of the first byte) of the new address.
2. Use the last 6 octets of the manufacturer's default address as the last 6 digits of the new address.

Here's a list of some sample valid addresses, shown with decimals for legibility:

```
40.00.00.b8.10.89  
40.00.01.b8.10.89  
50.00.00.b8.10.89  
60.00.00.b8.10.89  
7f.ff.ff.b8.10.89
```

Specifying an Alternate ATM Hardware Address

The following procedure applies to ATM Classic IP interface only. Hardware address swapping for ATM LAN Emulation adapters works just like hardware address swapping for the Ethernet and Token-Ring adapters that are being emulated.

Note: An ATM adapter has a hardware address which is 20 bytes in length. The first 13 bytes are assigned by the ATM switch, the next 6 bytes are burned into the ATM adapter, and the last byte represents the interface number (known as the *selector byte*). The above example only shows the burned in 6 bytes of the address. To select an alternate hardware address, you replace the 6 burned in bytes, and keep the last selector byte. The alternate ATM adapter hardware address is a total of 7 bytes.

To specify an alternate hardware address for an ATM Classic IP interface:

1. Use a value in the range of 40.00.00.00.00.00 to 7f.ff.ff.ff.ff.ff for the first 6 bytes.
2. Use the interface number as the last byte.

Here's a list of some sample alternate addresses for adapter interface at2 in the preceding output, shown with decimals for readability:

```
40.00.00.00.00.00.02  
40.00.01.00.00.00.02  
50.00.00.00.01.00.02  
60.00.00.01.00.00.02  
7f.ff.ff.ff.ff.ff.02
```

Since the interface number is hard-coded into the ATM hardware address, it must move from one ATM adapter to another during hardware address swapping. This imposes certain requirements and limitations on the HACMP/ES configuration.

HACMP/ES Configuration Requirements for ATM Hardware Address Swapping (Classical IP Only)

- If the hardware address moves to another adapter on the same machine (adapter swapping), the interface will have to be configured on that adapter as well. Likewise, when IP address takeover occurs, the interface associated with the adapter on the remote node will need to be configured on the takeover node.

- There can be *no more than 7 ATM service adapters per cluster* that support hardware address takeover.
- Each of these service interfaces *must have a unique ATM interface number*.
- On nodes that have one standby adapter per service adapter, the standby adapters on *all* cluster nodes will use the eighth possible ATM device (*at7*), so that there is no conflict with the service interface used by any of the nodes. This will guarantee that during IP address takeover with hardware address swapping the interface associated with the hardware address is not already in use on the takeover node.
- On nodes that have more than one standby adapter per service adapter, the total number of available service interfaces is reduced by that same number. For example, if two nodes have two standby adapters each, then the total number of service interfaces is reduced to 5.
- Any ATM adapters that are not being used by HACMP/ES, but are still configured on any of the cluster nodes that are performing ATM hardware address swapping, will also reduce the number of available ATM interfaces on a one-for-one basis.

Network Configuration Requirements for ATM Hardware Address Swapping (Classical IP and LAN Emulation)

- Hardware address swapping for ATM requires that all adapters that can takeover for a given service address be attached to the same ATM switch.

Avoiding Network Conflicts

Each network adapter is assigned a unique hardware address when it is manufactured, which ensures that no two adapters have the same network address. When defining a hardware address for a network adapter, ensure that the defined address does not conflict with the hardware address of another network adapter in the network. Two network adapters with a common hardware address can cause unpredictable behavior within the network.

To reduce the chance that the chosen address is not a duplicate of another network adapter, consider using the following hardware addresses:

0x08007c5d76d1	0x08007c5d6efd
0x08007c5d3c43	0x08007c5d2ea3
0x08007c5d2e44	0x08007c5d26ae
0x08007c5d4d9a	0x08007c5d6484
0x08007c5d774b	0x08007c5d65df
0x08007c5d6455	0x08007c5d6398
0x08007c5d774b	0x08007c5d65df
0x08007c5d6455	0x08007c5d6398

Afterwards, to confirm that no duplicate addresses exist on your network, bring the cluster up on your new address and ping it from another machine. If you receive two packets for each ping (one with a trailing **DUP!**), you have probably selected an address already in use. Select another and try again. Cycle the Cluster Manager when performing these operations, because the alternate address is used only after the HACMP/ES software is running and has reconfigured the adapter to use the service IP address the alternate hardware address associates with it.

Note: You cannot use the preceding addresses in a Token-Ring network or ATM Classic IP network. Token-Ring hardware addresses begin with the number 42. Alternate ATM Classic IP hardware addresses use 14 digits.

Adding the Network Topology to the Cluster Diagram

You now add the network topology to your cluster diagram.

Sketch in the networks, including all TCP/IP networks, any TCP/IP point-to-point connections, and serial networks. Identify each network by name and attribute. In the boxes in each node that represent slots, record the adapter label. If you are using IP address takeover or rotating resources, remember to include a boot address for each service adapter that can be taken over.

You can now add networking to the sample cluster diagram started in Chapter 2, Overview: Planning an HACMP/ES Cluster.

A sample network set up might include the use of five networks:

- A token ring network, named *clus1_TR*, used to connect clients to the ten cluster nodes that run the customer database “front end” application. The token ring network is a public network; it allows client access.
- An ethernet network, named *db_net*, used to connect the four cluster nodes that run the database “backend” application, which handles the update of the actual database records. The ethernet network *db_net* is a private network that is not intended for client use.
- The SP ethernet is configured as a service network, named *sp_ether*. It is defined with a public attribute but is not available for client access
- The SP Switch network is configured as a service network and is also used for high speed data transfer between the front end nodes and back-end nodes. You might name the SP Switch network *clus1_HPS* because its name must include the characters “HPS”, all in uppercase.
- To avoid corruption of critical database storage, the *db_net* nodes are also connected by a series of point-to-point serial networks. These serial networks provide additional protection against failure of the TCP/IP system.

The HACMP/ES software sends heartbeats across all networks defined to the cluster, to monitor their status.

Where You Go From Here

You have now planned the network topology for the cluster. The next step in the planning process is to lay out the shared disk configuration for your cluster, described in Chapter 5, Planning Shared Disk Devices.

Chapter 5 Planning Shared Disk Devices

This chapter discusses information you must consider before configuring shared external disks in an HACMP/ES cluster.

Prerequisites

By now, you should have completed the planning steps in the previous chapters:

- Planning applications (Chapter 3, Initial Cluster Planning)
- Planning networks (Chapter 4, Planning Cluster Network Connectivity)

You should refer to AIX documentation for the general hardware and software setup for your disks.

Overview

In an HACMP/ES cluster, shared disks are external disks that are connected to more than one cluster node. The disks are “owned” by only one node at a time. If the owner node fails, the cluster node with the next highest priority in the resource chain acquires ownership of the shared disks and restart applications to restore critical services to clients. This ensures that the data stored on the disks remains accessible to client applications. Typically, takeover occurs within 30 to 300 seconds; but this range depends on the number and types of disks being used, the number of volume groups, the filesystems (whether shared or cross-mounted), and the number of critical applications in the cluster configuration.

When planning the shared external disk for your cluster, the objective is to eliminate single points of failure in the disk storage subsystem. The following table lists the disk storage subsystem components, with recommended ways to eliminate them as single points of failure.

Cluster Object	Eliminated as Single Point of Failure by...
Disk adapter	Using redundant disk adapters
Controller	Using redundant disk controllers
Disk	Using redundant hardware and LVM disk mirroring or RAID devices, the mirroring provided by the disk subsystem

In this chapter, you will perform the following planning tasks:

- Choosing a shared disk technology. HACMP/ES supports several types of storage devices including SCSI-2 disks and Serial Storage Architecture (SSA) disk subsystems.
- Planning the installation of the shared disk storage. This includes:

- Determining the number of disks required to handle the projected storage capacity. You need multiple physical disks on which to put the mirrored logical volumes. Putting copies of a mirrored logical volume on the same physical device defeats the purpose of making copies. See Chapter 6, Planning Shared LVM Components, for more information on creating mirrored logical volumes.
- Determining the number of disk adapters each node will contain to connect to the disks or disk subsystem.

Physical disks containing logical volume copies should be on separate adapters. If all logical volume copies are connected to a single adapter, the adapter is potentially a single point of failure. If the single adapter fails, you must move the volume group to an alternate node. Separate adapters prevent any need for this move.

- Understand the cabling requirements for each type of disk technology.
- Completing planning worksheets for the disk storage
- Adding the selected disk configuration to the cluster diagram.

Choosing a Shared Disk Technology

The HACMP/ES software supports the following disk technologies as shared external disks in a highly available cluster:

- SCSI-2 SE, SCSI-2 Differential and SCSI-2 Differential Fast/Wide adapters and drives, including RAID subsystems.
- IBM SSA adapters and SSA disk subsystems.

You can combine these technologies within a cluster. Before choosing a disk technology, however, review the considerations for configuring each technology as described in this section.

SCSI Disk Planning Considerations

The benefit of the SCSI implementation is its low cost. It provides a shared disk solution that requires minimal hardware overhead.

The HACMP/ES software supports the following SCSI disk devices and arrays as shared external disk storage in cluster configurations:

- SCSI-2 Differential and SCSI-2 Differential Fast/Wide disk devices
- The IBM 7135-110 and 7135-210 RAIDiant Disk Arrays
- The IBM 7137 Disk Array
- The IBM 2105-B09 and 2105-100 Versatile Storage Servers

The following restrictions, however, apply to using shared SCSI disks in a cluster configuration:

- Different types of SCSI busses can be configured in an HACMP/ES cluster. Specifically, SCSI-2 Differential and SCSI-2 Differential Fast/Wide devices can be configured in clusters of up to four nodes, where all nodes are connected to the same SCSI bus attaching the separate device types. (You cannot mix SCSI-2 SE, SCSI-2 Differential and SCSI-2 Differential Fast/Wide devices on the same bus.)

- You can connect the IBM 7135-210 RAIDiant Disk Array to *only* High Performance SCSI-2 Differential Fast/Wide adapters, while the 7135-110 RAIDiant Array *cannot* use those High Performance Fast/Wide adapters.
- You can connect up to sixteen devices to a SCSI-2 Differential Fast/Wide bus. Each SCSI adapter and disk is considered a separate device with its own SCSI ID. SCSI-2 Differential Fast/Wide maximum bus length of 25 meters provides enough length for most cluster configurations to accommodate the full sixteen-device connections allowed by the SCSI standard.
- Do not connect other SCSI devices, such as CDROMs or tape drives, to a shared SCSI bus.
- The IBM High Performance SCSI-2 Differential Fast/Wide Adapter cannot be assigned SCSI IDs 0, 1, or 2; the adapter restricts the use of these IDs. The IBM SCSI-2 Differential Fast/Wide Adapter/A (FC 2416) cannot be assigned SCSI IDs 0 or 1.
- Physical disks containing logical volume copies should be connected to different power supplies; otherwise, loss of a single power supply can prevent access to all copies. Thus, you should plan on using multiple disk subsystem drawers or desk-side units to avoid dependence on a single power supply.
- If you plan to enable quorum on the volume group to be defined over the physical storage, include at least three separate disks in the volume group. A two-disk volume group puts you at risk for losing quorum and data access. Either build three-disk volume groups or disable quorum.

IBM 7135 RAIDiant Disk Array Planning Considerations

The benefits of using an IBM 7135 RAIDiant Disk Array in an HACMP/ES cluster are its storage capacity, speed, and reliability features. The IBM 7135 RAIDiant Disk Array contains a group of disk drives that work together to provide enormous storage capacity (up to 135 GB of nonredundant storage) and higher I/O rates than single large drives.

When using an IBM 7135 RAIDiant Disk Array, do not use LVM mirroring. You must, however, account for the data redundancy maintained by the IBM 7135 RAIDiant Disk Array when calculating total storage capacity requirements. For example, in RAID level 1, because the IBM 7135 RAIDiant Disk Array maintains two copies of the data on separate drives, only half the total storage capacity is usable. Likewise, with RAID level 5, 20 to 30 percent of the total storage capacity is used to store and maintain parity information.

RAID Levels

The IBM 7135 RAIDiant Disk Arrays support reliability features that provide data redundancy to prevent data loss if one of the disk drives in the array fails. As a RAID device, the array can provide data redundancy through RAID levels. The IBM 7135 RAIDiant Disk Arrays support RAID levels 0, 1, and 5. RAID level 3 can only be used with a raw disk.

In RAID level 0, data is striped across a bank of disks in the array to improve throughput. Because RAID level 0 does not provide data redundancy, it is not recommended for use in HACMP/ES clusters.

In RAID level 1, the IBM 7135 RAIDiant Disk Array provides data redundancy by maintaining multiple copies of the data on separate drives (mirroring).

In RAID level 5, the IBM 7135 RAIDiant Disk Array provides data redundancy by maintaining parity information that allows the data on a particular drive to be reconstructed if the drive fails.

All drives in the array are hot-pluggable. When you replace a failed drive, the IBM 7135 RAIDiant Disk Array reconstructs the data on the replacement drive automatically. Because of these reliability features, you should not define LVM mirrors in volume groups defined on an IBM 7135 RAIDiant Disk Array.

Dual Active Controllers

To eliminate adapters or array controllers as single points of failure in an HACMP/ES cluster, the IBM 7135 RAIDiant Disk Array can be configured with a second array controller that acts as a backup controller in the event of a fallover. This configuration requires that you configure each cluster node with two adapters. You connect these adapters to the two array controllers using separate SCSI busses.

In this configuration, each adapter and array-controller combination defines a unique path from the node to the data on the disk array. The IBM 7135 RAIDiant Disk Array software manages data access through these paths. Both paths are active and can be used to access data on the disk array. If a component failure disables the current path, the disk array software automatically re-routes data transfers through the other path.

Note: This dual-active path-switching capability is independent of the capabilities of the HACMP/ES software, which provides protection from a node failure. When you configure the IBM 7135 RAIDiant Disk Array with multiple controllers and configure the nodes with multiple adapters and SCSI busses, the disk array software prevents a single adapter or controller failure from causing disks to become unavailable.

The following restrictions apply to using shared IBM 7135 RAIDiant Disk Arrays in a cluster configuration:

- You can connect the IBM 7135-210 RAIDiant Disk Array to *only* High Performance SCSI-2 Differential Fast/Wide adapters, while the 7135-110 RAIDiant Array *cannot* use those High Performance Fast/Wide adapters.
- You can connect an IBM 7135 RAIDiant Disk Array to up to four cluster nodes using a SCSI bus. You also can include up to two IBM 7135 RAIDiant Disk Arrays per bus.

Note: Each array controller and adapter on the same SCSI bus requires a unique SCSI ID.

- You may need to configure the drives in the IBM 7135 RAIDiant Disk Array into logical units (LUNs) before setting up the cluster. A standard SCSI disk equates to a single LUN. AIX configures each LUN as a hard disk with a unique logical name of the form *hdiskn*, where *n* is an integer.

An IBM 7135 RAIDiant Disk Array comes preconfigured with several LUNs defined. This configuration depends on the number of drives in the array and their sizes and is assigned a default RAID level of 5. You can, if desired, use the disk array manager utility to configure the individual drives in the array into numerous possible combinations of LUNs, spreading a single LUN across several individual physical drives.

For more information about configuring LUNs on an IBM 7135 RAIDiant Disk Array, see the documentation you received with your disk array for the specific LUN composition of your unit.

Once AIX configures the LUNs into hdisks, you can define the ownership of the disks as you would any other shared disk.

IBM 7137 Disk Array Planning Considerations

The IBM 7137 disk array contains multiple SCSI-2 Differential disks. On the IBM 7137 array, these disks can be grouped together into multiple LUNs, with each LUN appearing to the host as a single SCSI device (hdisk)

IBM 2105 Versatile Storage Server

The IBM 2105 Versatile Storage Server (VSS) provides multiple concurrent attachment and sharing of disk storage for a variety of open systems servers. RISC System/6000 processors can be attached, as well as other UNIX and non-UNIX platforms.

The VSS uses IBM SSA disk technology. Existing IBM 7133 SSA disk drawers can be used in the VSS. See the section, IBM Serial Storage Architecture Disk Subsystem on page 5-5 for more information about SSA systems.

The RISC System/6000 attaches to the IBM 2105 Versatile Storage Server via a PCI SCSI-2 Fast/Wide Differential Adapter. A maximum of 64 open system servers can be attached to the VSS (16 SCSI channels with 4 adapters per channel).

There are many availability features included in the VSS. All storage is protected with RAID technology. RAID-5 techniques can be used to distribute parity across all disks in the array. *Sparing* is a function which allows you to assign a disk drive as a spare for availability. Predictive Failure Analysis techniques are utilized to predict errors *before* they affect data availability. Failover Protection enables one partition, or *storage cluster*, of the VSS to takeover for the other so that data access can continue.

The VSS includes other features such as a web-based management interface, dynamic storage allocation, and remote services support. For more information on VSS planning, general reference material, and attachment diagrams, see the URLs:

<http://www.storage.ibm.com/hardsoft/products/vss/books/vssrefinfo.htm>
<http://www.storage.ibm.com/hardsoft/products/vss/books/vsrlag.htm>

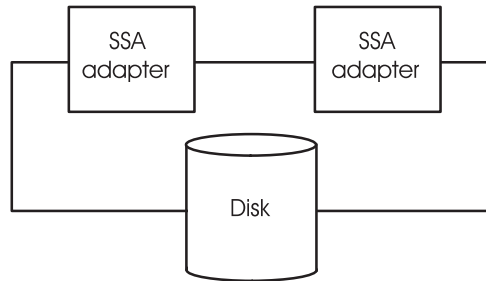
IBM Serial Storage Architecture Disk Subsystem

Serial Storage Architecture (SSA) offers many features for minimizing single points of failure and achieving high availability in an HACMP/ES environment.

You can use IBM 7133 and 7131-405 SSA disk subsystems as shared external disk storage devices to provide concurrent access in an HACMP/ES cluster configuration.

SSA is hot pluggable. Consequently, if you include SSA disks in a volume group using LVM mirroring, you can replace a failed disk drive without powering off the entire system.

The figure below shows the basic SSA loop configuration.



Basic SSA Loop Configuration

Disk Power Supply Considerations

Reliable power sources are critical for a highly available cluster. Each mirrored disk chain in the cluster should have a separate power source. As you plan the cluster, make sure that the failure of any one power source (a blown fuse, for example) does not disable more than one node or mirrored chain. The following sections discuss specific power supply considerations for supported disk types.

SCSI Device Power Considerations

If the cluster site has a multiple phase power supply, you must ensure that the cluster nodes are attached to the same power phase. Otherwise, the ground will move between the systems across the SCSI bus and cause write errors.

The bus and devices shared between two nodes are subject to the same operational power surge restrictions as standard SCSI systems. Uninterruptible power supply (UPS) devices are necessary for preventing data loss. When power is first applied to a SCSI device, the attached bus may have active data corrupted. You can avoid such errors by briefly halting data transfer operations on the bus while a device (disk or adapter) is turned on. For example, if cluster nodes are installed on two different power grids and one node has a power surge that causes it to reboot, the surviving node may lose data if a data transfer is active.

The IBM 7135 RAIDiant Disk Arrays and IBM 2105 Versatile Storage Servers are less prone to power supply problems because they come with redundant power supplies.

IBM SSA Disk Subsystem Configurations

Clusters with IBM SSA disk subsystems are less prone to power supply problems because they come with redundant power supplies.

Planning for Non-Shared Disk Storage

Keep the following considerations in mind regarding non-shared disk storage:

- The internal disks on each node in a cluster must provide sufficient space for:
 - AIX software (approximately 320 MB)
 - HACMP/ES software (approximately 35 MB for a server node)
 - Executable modules of highly available applications.
- The root volume group (**rootvg**) for each node must not reside on the shared SCSI bus.
- Use the AIX Error Notification Facility to monitor the **rootvg** on each node. Problems with the root volume group can be promoted to node failures. See Chapter 13, Tailoring AIX for HACMP/ES, for more information on using the Error Notification facility.
- Because shared disks require their own adapters, you cannot use the same adapter for both a shared and a non-shared disk. The internal disks on each node require one SCSI adapter apart from any other adapters within the cluster.
- Internal disks must be in a different volume group from the external shared disks.
- The executable modules of the highly available applications should be on the internal disks and not on the shared external disks, for the following reasons:

Licensing

Some vendors require a unique license for each processor or multi-processor that runs an application, and thus license-protect the application by incorporating processor-specific information into the application when it is installed. As a result, it is possible that even though the HACMP/ES software processes a node failure correctly, it is unable to restart the application on the fallover node because of a restriction on the number of licenses available within the cluster for that application. To avoid this problem, be sure that you have a license for each processor in the cluster that may potentially run an application.

Starting Applications

Some applications (such as databases) contain configuration files that you can tailor during installation and store with the binaries. These configuration files usually specify startup information, such as the databases to load and log files to open, after a fallover situation.

If you plan to put these configuration files on a shared file system, they will require additional tailoring. You will need to determine logically which system (node) actually is to invoke the application in the event of a fallover. Making this determination becomes particularly important in fallover configurations where conflicts in the location and access of control files can occur.

For example, in a two-node mutual takeover configuration, where both nodes are running different instances of the same application (different databases) and are standing by for one another, the takeover node must be aware of the location of specific control files and must be able to access them to perform the necessary steps to start critical applications after a fallover or else the fallover will fail, leaving critical applications unavailable to clients. If the configuration files are on a shared file system, a conflict can arise if the takeover node is not aware of the file's location.

You can avoid much of the tailoring of configuration files by placing slightly different startup files for critical applications on local file systems on either node. This allows the initial application parameters to remain static; the application will not need to recalculate the parameters each time it is invoked.

Planning a Shared SCSI-2 Disk Installation

The following list summarizes the basic hardware components required to set up an HACMP/ES cluster that includes SCSI-2 SE, SCSI-2 Differential or SCSI-2 Differential Fast/Wide devices as shared storage. Your exact cluster requirements will depend on the configuration you specify. To ensure that you account for all required components, complete a diagram for your system.

Note: When planning a shared SCSI disk installation, consult the *HACMP Installation Guide*, specifically, the restrictions listed in Chapter 5, in the section, SCSI Disks.

Also be sure to consult the hardware manuals for detailed information on cabling and attachment for the particular devices you are configuring.

Disk Adapters

The HACMP/ES software supports the following:

- IBM SCSI-2 Differential High-Performance External I/O Controller.
- IBM High Performance SCSI-2 Differential Fast/Wide Adapter/A.
- PCI SCSI-2 Differential Fast/Wide Adapter.
- PCI SCSI-2 Fast/Wide Differential Adapter.

For each IBM 7135 RAIDiant Disk Array, an HACMP/ES configuration requires that you configure each cluster node with two host adapters.

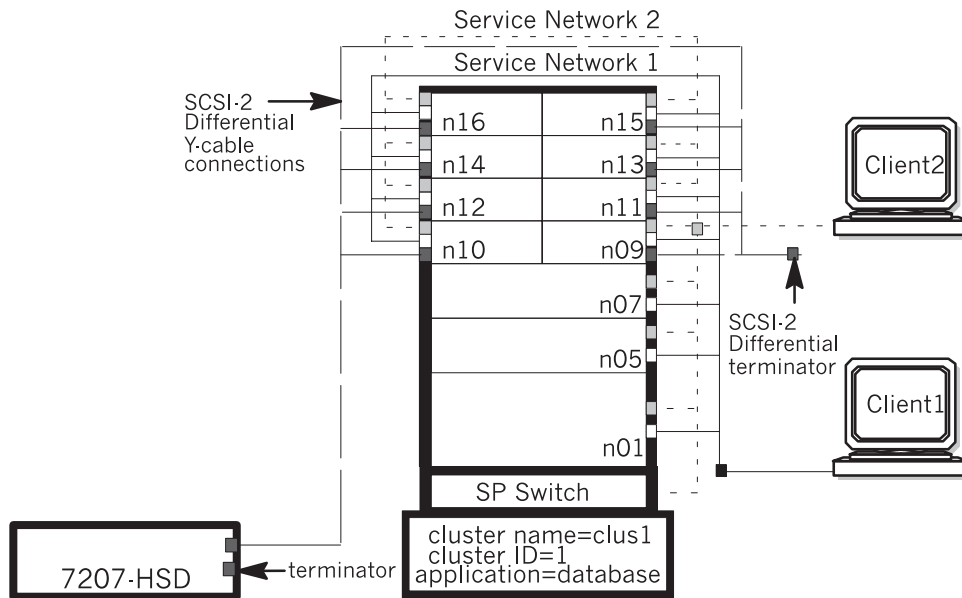
Note: Remove any SCSI terminators on the adapter card. You must use external terminators in an HACMP/ES cluster. If you terminate the shared SCSI bus on the adapter, you lose termination when the cluster node that contains the adapter fails.

Cables

The cables required to connect nodes in your cluster depend on the type of SCSI bus you are configuring. Be sure to choose cables that are compatible with your disk adapters and controllers. For information on the specific type and length SCSI-2 Differential or SCSI-2 Differential Fast/Wide cable requirements, see the hardware documentation that accompanies each device you want to include on the SCSI bus.

Sample SCSI-2 Differential Configuration

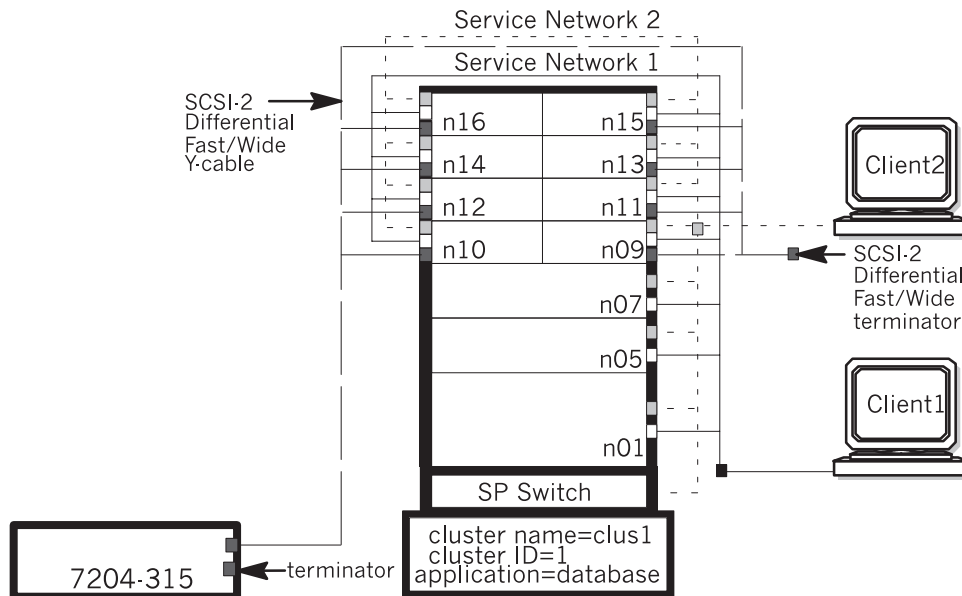
The following diagram illustrates a SCSI-2 Differential configuration in which four nodes are attached to four SCSI-2 Differential disks in a IBM 7207-HSD enclosure. Each adapter is connected to a SCSI-2 Differential Y-cable, terminated at node 9 (n09). The last SCSI-2 Differential Y-cable connection is made at node 10 (n10). Its other end connects to the disk adapter, and the configuration is terminated at the disk enclosure.



Shared SCSI-2 Differential Disk Configuration

Sample SCSI-2 Differential Fast/Wide Configuration

The following diagram illustrates a SCSI-2 Differential Fast/Wide configuration in which the nodes are attached to four SCSI-2 Differential disks in an IBM 7204-315 enclosure. Each adapter is connected to a SCSI-2 Differential Fast/Wide Y-cable, terminated at node 9 (n09). The last SCSI-2 Differential Fast/Wide Y-cable connection is made at node 10 (n10). Its other end connects to the disk adapter, and the configuration is terminated at the disk enclosure.



Shared SCSI-2 Differential Fast/Wide Disk Configuration

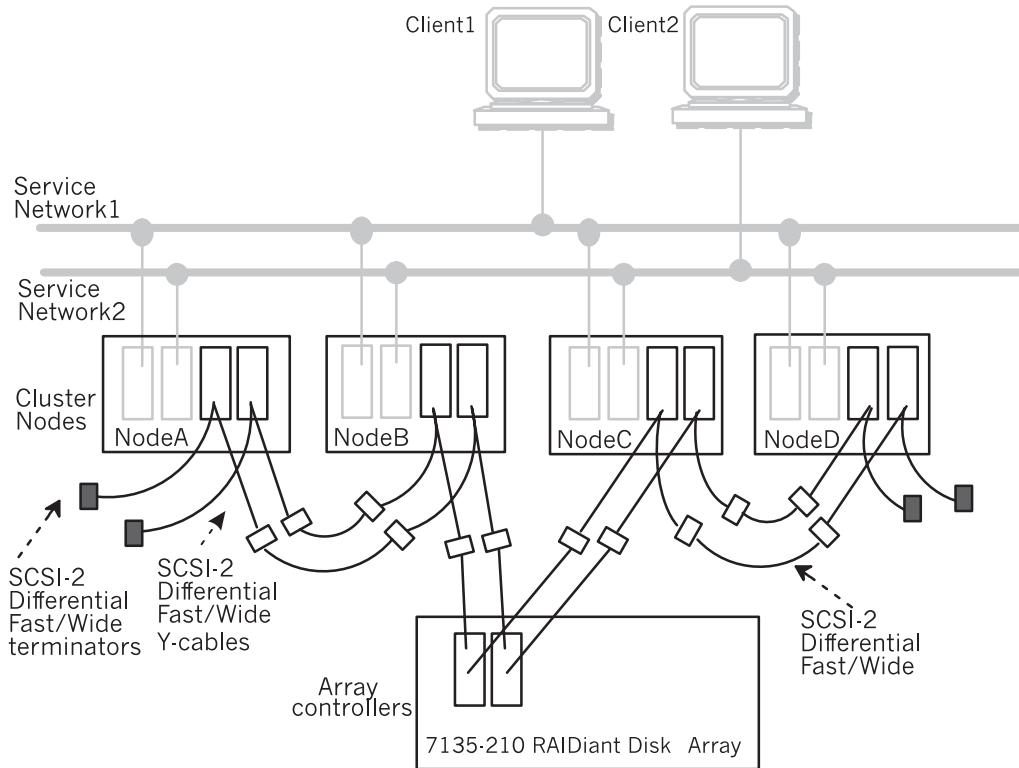
Sample IBM 7135-210 RAIDiant Disk Array Configuration

To take advantage of the path-switching capability of the IBM 7135-210 RAIDiant Disk Array software, you must configure each node with two adapters. In this way, the device driver can define multiple paths between the host and the disk array, eliminating the adapters, the controllers, and the SCSI bus as single points of failure. If a component failure disables one path, the IBM 7135-210 RAIDiant Disk Array device driver can switch to the other path automatically. This switching capability is the dual-active feature of the array.

Note: Although each controller on the IBM 7135-210 RAIDiant Disk Array contains two connectors, each controller requires only one SCSI ID.

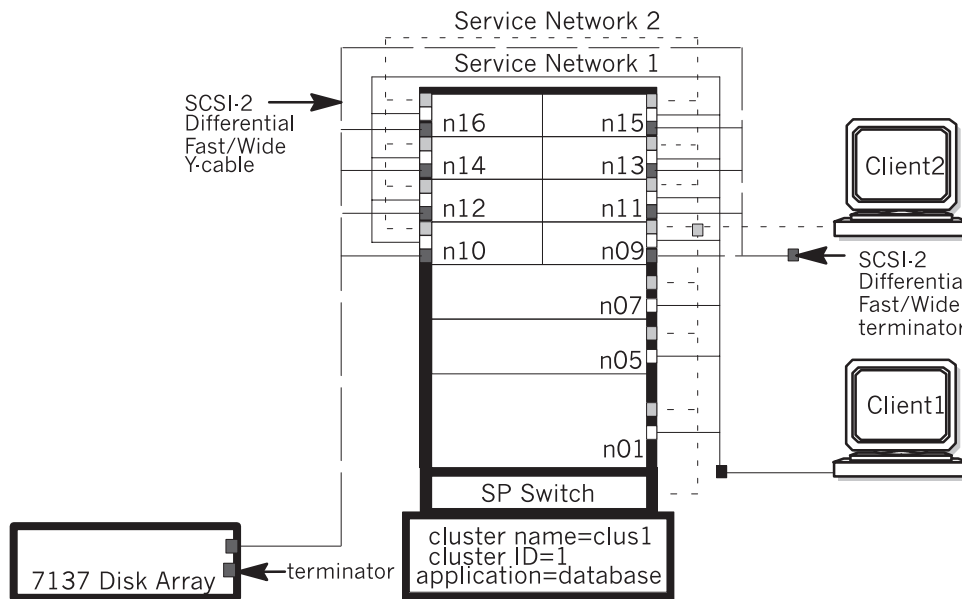
SCSI-2 Differential Fast/Wide IBM 7135-210 RAIDiant Disk Array Configuration

The following diagram illustrates an HACMP/ES cluster with an IBM 7135-210 RAIDiant Disk Array connected to two SCSI-2 Differential Fast/Wide busses. IBM 7135-210 RAIDiant Disk Array Configuration with SCSI-2 Differential Fast/Wide Busses.



SCSI-2 Differential Fast/Wide IBM 7137 Disk Array Configuration

The following diagram illustrates an HACMP/ES cluster with an IBM 7137 Disk Array connected to a SCSI-2 Differential Fast/Wide bus. Each adapter is connected to a SCSI-2 Differential Fast/Wide Y-cable, terminated at node 9 (n09). The last SCSI-2 Differential Fast/Wide Y-cable connection is made at node 10 (n10). Its other end connects to the disk adapter, and the configuration is terminated at the disk enclosure.



IBM 7137 Disk Array with a SCSI-2 Differential Fast/Wide Bus

Sample IBM 2105 Versatile Storage Server Configuration

Several diagrams are included in the information provided on the IBM world wide web pages. See the list of figures in the attachment guide for the VSS:

<http://www.storage.ibm.com/hardsoft/products/vss/books/vsrlag.htm>

Using VSS Features for High Availability

When using the VSS in an HACMP/ES environment, the following is recommended for high availability:

- Use the Sparring function to assign disks as spares and reduce the exposure to data loss. When the VSS detects that a disk is failing, it transfers the data from the failing disk to a spare device. You are required to specify at least one disk as a spare per drawer; however you can specify two spares to a drawer for increased availability.
- Configure the two host interface cards in a bay to device interface cards in the same bay.
- Configure the SCSI ports on the same interface card to the same partition of the VSS.

Planning a Shared IBM SSA Disk Subsystem Installation

The discussion in this section is specific to using SSA disks with HACMP/ES. It is meant to supplement the IBM manuals listed below.

IBM Manuals

Use the IBM manuals for:

- Rules for connecting SSA drives in loops to nodes
- Examples of configurations
- Planning charts.

The following manuals cover SSA disk subsystem hardware.

- 7133 Models 010 & 020 SSA Disk Subsystem: Installation Guide, Order Number GA33-3260
- 7133 Models 500 and 600 SSA Disk Subsystem: Installation Guide, Order Number GA33-3263-02
- 7133 SSA Disk Subsystem: Operator Guide, Order Number GA33-3259-01
- 7133 SSA Disk Subsystems: Additional Installation and Service Information, Order Number 97HO536
- 7133 SSA Disk Subsystem: Service Guide, Order Number SY33-0185-02
- 7133 Hardware Technical Reference, Order Number SA33-3261
- Planning SSA RAID Subsystems, Order Number GA33-3271

The following manuals cover SSA adapter hardware.

- IBM SSA 4-Port Adapter: Installation and Reference, Order Number SC23-2775-00
- SSA Adapters User's Guide and Maintenance Information, Order Number SA33-3272-01
- SSA 4-Port Adapter Enhanced SSA 4-Port Adapter: Technical Reference, Order Number S31H-8612-01
- Adapters, Devices, and Cable Information for Micro Channel Bus Systems, Order Number SA23-2764-02

The following manual covers general SSA reference material. It includes a discussion of high availability.

- A Practical Guide to Serial Storage Architecture for AIX, Order Number SG24-4599

Adapters

All of the IBM manuals listed above are good sources on how to connect SSA disk subsystems to nodes.

SSA 4-Port Adapter (Feature code 6214, Type 4-D)

The 6214 SSA adapter (also called a 2-way adapter) can support SSA loops containing up to two adapters per loop. In a high availability environment, then, an SSA loop is limited to one of these two configurations:

- one node containing two adapters
- two nodes, each containing one adapter.

Because of the limit imposed by the two-way adapter, if you configure both the two-way and eight-way adapters in one loop, you can use only two adapters.

These adapters must be at Microcode level 2401 or later.

Enhanced SSA 4-Port Adapter (Feature Code 6216, Type 4-G)

The 6216 SSA adapter (also called an eight-way adapter) can support SSA loops containing up to eight eight-way adapters per loop. Most multi-node configurations set up with a minimal number of single points of failure require eight-way adapters.

These adapters must be at Microcode level 2402 or later.

SSA Multi-Initiator RAID Adapters (Feature Codes 6215 and 6219)

These adapters must be at microcode level 1801 or later.

Identifying Adapters

The two-way and eight-way adapters look the same. What differs is their microcode. So the easiest way to determine which adapter you have is to install it in a machine and run either of the following commands:

```
lsdev -Cc adapter
```

or

```
lscfg -vl ssaX
```

where X is the adapter number.

These commands display identifying information about the microcode.

Bypass Cards

The 7133 Models 020 and 600 disk subsystems contain four bypass cards. Each bypass card has two external SSA connectors. Through these, you connect the bypass cards and, therefore, the disk drive module strings to each other or to a node.

The bypass cards can operate in either bypass or forced inline mode.

Bypass Mode

When you set its jumpers so a bypass card operates in bypass mode, it monitors both of its external connections. If it detects that one of its connectors is connected to a powered-on SSA adapter or device, it switches to inline mode; that is, it connects the internal SSA links to the external connector. This effectively heals the break in the SSA loop.

If the bypass card detects that neither of its connectors is connected to a powered-on SSA adapter or device, it switches to bypass state; that is, it connects the internal disk strings and disconnects them from the external connector.

Forced Inline Mode

When you set its jumpers so a bypass card operates in forced inline mode, it behaves permanently like a signal card of Models 010 and 500; that is, none of its electronic switching circuits are in use. Its internal SSA links connect to the external connector and can never make an internal bypass connection.

Using SSA Features For High Availability

Using SSA Loops

Configure so that all SSA devices are in a loop, not just connected in a string. Although SSA devices function connected in a string, a loop provides two paths of communications to each device for redundancy. The adapter chooses the shortest path to a disk.

Using SSA Fiber Optic Extenders

The SSA Fiber Optic Extenders use cables up to 2.4 Km to replace a single SSA cable. The SSA Fiber Optic Extender (Feature code 5500) is supported on all Model 7133 disk subsystems.

Using Fiber Optic extender, you can make the distance between disks greater than the LAN allows. If you do so, you can not use routers and gateways. Consequently, under these circumstances, you cannot form an HACMP/ES cluster between two LANs.

Daisy-chaining the Adapters

In each node, for each loop including that node, daisy-chain all its adapters. The SSAR router device uses another adapter when it detects that one adapter has failed. You need only one bypass switch for the whole daisy chain of adapters in the node rather than a bypass switch for each individual adapter.

Bypass Cards In the 7133, Models 020 and 600 Disk Subsystems

Bypass cards maintain high availability when a node fails, when a node is powered off, or when the adapter(s) of a node fail. Connect the pair of ports of one bypass card into the loop that goes to and from one node. That is, connect the bypass card to only one node. Remember to daisy-chain adapters if you are using more than one adapter in a node.

Avoid two possible conditions when a bypass card switches to bypass mode:

- Do not connect two independent loops through a bypass card. When the bypass card switches to bypass mode, you want it to reconnect the loop inside the 7133 disk subsystem, rather than connecting two independent loops. So both ports of the bypass card must be in the same loop.
- Dummy disks are connectors used to fill out the disk drive slots in a 7133 disk subsystem so the SSA loop can continue unbroken. Make sure that when a bypass card switches to bypass mode, it connects no more than three dummy disks consecutively in the same loop. Put the disks next to the bypass cards and dummy disks between real disks.

Configuring To Minimize Single Points of Failure

To minimize single points of failure, consider the following points:

- Use logical volume mirroring and place logical volume mirror copies on separate disks and in separate loops using separate adapters. In addition, it is a good idea to mirror between the front row and the back row of disks or between disk subsystems.
- The bypass card itself can be a single point of failure. Two ways to avoid this are:
 - With one loop: Put two bypass cards into a loop connecting to each node.
 - With two loops: Do logical volume mirroring to disks in a second loop. Set each loop to go through a separate bypass card to each node.

Set the bypass cards to forced inline mode for the following configurations:

- When connecting multiple 7133 disk subsystems.
- When the disk drives in one 7133 Model 010 or Model 600 are not all connected to the same SSA loop. In this type of configuration, forced inline mode removes the risk of a fault condition, namely, that a shift to bypass mode might cause the disk drives of different loops to be connected.

Configuring For Optimal Performance

- Multiple Nodes And SSA Domains
 - A node and the disks it accesses make up an SSA domain. For configurations containing shared disk drives and multiple nodes, you need to minimize the path length from each node to the disk drives it accesses. Measure the path length by the number of disk drives and adapters in the path. Each device has to receive and forward the packet of data.
 - With multiple adapters in a loop, put the disks near the closest adapter and make that the one that access the disks. In effect, try to keep I/O traffic within the SSA domain. Although any host can access any disk it is best to minimize I/O traffic crossing over to other domains.
 - When multiple hosts are in a loop, set up the volume groups so that a node uses the closest disks. This prevents one node's I/O from interfering with another's.
- Distribute read and write operations evenly throughout the loops.
- Distribute disks evenly among the loops.
- Download microcode when you replace hardware.

To ensure that everything works correctly, install the latest file sets, fixes, and microcode for your disk subsystem.

Testing Loops

- Test all loop scenarios thoroughly, especially in multiple-node loops. Test for loop breakage (failure of one or more adapters).
- Test bypass cards for power loss in adapters and nodes to verify that they follow configuration guidelines.

Planning for Concurrent SSA Volume Groups

If you plan to create concurrent volume groups on SSA disk subsystem, assign unique non-zero node numbers with `ssar` on each cluster node before using the concurrent volume group failed drive replacement procedure.

If you specify the use of SSA disk fencing in your concurrent resource group, HACMP/ES assigns the node numbers when you synchronize the resources. If you don't specify the use of SSA disk fencing in your concurrent resource group, assign the node numbers with

```
chdev -l ssar -a node_number=x
```

where `x` is the number to assign to that node. Then reboot the system.

SSA Disk Fencing in Concurrent Access Clusters

Preventing data integrity problems that can result from the loss of TCP/IP network communication is especially important in concurrent access configurations where multiple nodes have simultaneous access to a shared disk. Chapter 4, *Planning Cluster Network Connectivity*, describes using HACMP-specific serial networks to prevent partitioned clusters.

Concurrent access configurations using SSA disk subsystems can also use disk fencing to prevent data integrity problems that can occur in partitioned clusters.

The SSA disk subsystem includes fence registers, one per disk, capable of permitting or disabling access by each of the eight possible connections. Fencing provides a means of preventing uncoordinated disk access by one or more nodes.

The SSA hardware has a fencing command for automatically updating the fence registers. This command provides a tie-breaking function within the controller for nodes independently attempting to update the same fence register. A compare-and-swap protocol of the fence command requires that each node provide both the current and desired contents of the fence register. If competing nodes attempt to update a register at about the same time, the first succeeds, but the second fails because it does not know the revised contents.

SSA Disk Fencing Implementation

The HACMP/ES software manages the contents of the fence registers. At cluster configuration, the fence registers for each shared disk are set to allow access for the designated nodes. As cluster membership changes as nodes enter and leave the cluster, the event scripts call the `cl_ssa_fence` utility to update the contents of the fence register. If the fencing command succeeds, the script continues normal processing. If the operation fails, the script exits with failure, causing the cluster to go into reconfiguration.

Disk Fencing with SSA Disks in Concurrent Mode

You can only use SSA disk fencing under these conditions:

- Only disks contained in concurrent mode volume groups will be fenced.
- You configure all nodes of the cluster to have access to these disks and to use disk fencing.
- All resource groups with the disk fencing attribute enabled must be concurrent access resource groups.
- Concurrent access resource groups must contain all nodes in the cluster.

The purpose of SSA disk fencing is to provide a safety lockout mechanism for protecting shared SSA disk resources in the event that one or more cluster nodes become isolated from the rest of the cluster.

Concurrent mode disk fencing works as follows:

- The first node up in the cluster fences out all other nodes of the cluster from access to the disks of the concurrent access volume group(s) for which fencing is enabled, by changing the fence registers of these disks.
- When a node joins a cluster, the active nodes in the cluster allow the joining node access by changing the fence registers of all disks participating in fencing with the joining node.
- When a node leaves the cluster, regardless of how it leaves, the remaining nodes that share access to a disk with the departed node should fence out the departed node as soon as possible.
- If a node is the last to leave a cluster, whether the shutdown is forced or graceful, it clears the fence registers to allow access by all nodes. Of course, if the last node stops unexpectedly (is powered off or crashes, for example), it doesn't clear the fence registers. In this case, you must manually clear the fence registers using the proper SMIT options.

Enabling SSA Disk Fencing

The process of enabling SSA disk fencing for a concurrent resource group requires that *all volume groups containing SSA disks on cluster nodes must be varied off and the cluster must be down* when the cluster resources are synchronized. Note that this means all volume groups containing any of the SSA disks whether concurrent or non-concurrent, whether configured as part of the cluster or not, must be varied off for the disk fencing enabling process to succeed during the synchronization of cluster resources. If these conditions are not met, you will have to reboot the nodes to enable fencing.

The enabling process takes place on each cluster node: as follows

1. Assign a `node_number` to the `ssar` which matches the `node_id` of the node in the HACMP configuration. This means that any `node_numbers`, that were set prior to enabling disk fencing for purposes of replacing a drive or C-SPOC concurrent LVM functions, will be changed for disk fencing operations. The other operations will not be affected by this `node_number` change.
2. First remove, then remake all `hdisks`, `pdisks`, `ssa adapter`, and `tmssa` devices of the SSA disk subsystem seen by the node, thus picking up the `node_number` for use in the fence register of each disk.

This process is repeated each time cluster resources are synchronized while the cluster is down.

Disk Fencing and Dynamic Reconfiguration

When a node is added to the cluster through dynamic reconfiguration while cluster nodes are up, the disk fencing enabling process is performed on the added node only, during the synchronizing of topology. Any `node_numbers` that were set prior to enabling disk fencing for purposes of replacing a drive or C-SPOC concurrent LVM functions will be changed for disk fencing operations. Therefore, when initially setting SSA disk fencing in a resource group, the resources must be synchronized while the cluster is *down*. The other operations will not be affected by this `node_number` change.

Benefits Of Disk Fencing

Disk fencing provides the following benefits to concurrent access clusters:

- It enhances data security by preventing nodes that are not active members of a cluster from modifying data on a shared disk. By managing the fence registers, the HACMP/ES software can ensure that only the designated nodes within a cluster have access to shared SSA disks.
- It enhances data reliability by assuring that competing nodes do not compromise the integrity of shared data. By managing the fence registers HACMP/ES can prevent uncoordinated disk management by partitioned clusters. In a partitioned cluster, communication failures lead separate sets of cluster nodes to believe they are the only active nodes in the cluster. Each set of nodes attempts to take over the shared disk, leading to race conditions. The disk fencing tie-breaking mechanism arbitrates race conditions, ensuring that only one set of nodes gains access to the disk.

Completing the Disk Worksheets

After determining the disk storage technology you will include in your cluster, complete all of the appropriate worksheets from the following list:

- Shared SCSI-2 Differential or Differential Fast/Wide Disk Worksheet
- Shared SCSI Disk Array Worksheet
- Shared IBM Serial Storage Architecture Disk Subsystems Worksheet

Completing the Shared SCSI-2 Disk Worksheet

Complete a worksheet in Appendix A for each shared SCSI bus.

1. Enter the cluster ID and the Cluster name in the appropriate fields. This information was determined in Chapter 3, Initial Cluster Planning.
2. Check the appropriate field for the type of SCSI-2 bus.
3. Fill in the host and adapter information including the **node name**, the number of the **slot** in which the disk adapter is installed and the **logical name** of the adapter, such as scsi0. AIX assigns the logical name when the adapter is configured.
4. Determine the SCSI IDs for all the devices connected to the SCSI bus.
5. The IBM SCSI-2 Differential High Performance Fast/Wide adapter cannot be assigned SCSI IDs 0, 1, or 2.
6. Record information about the Disk drives available over the bus, including the logical device name of the disk on every node. (This name, an hdisk name, is assigned by AIX when the device is configured and may vary on each node.)

Completing the Shared SCSI-2 Disk Array Worksheet

Complete a worksheet in Appendix A for each shared SCSI disk array.

1. Enter the cluster ID and the Cluster name in the appropriate fields. This information was determined in Chapter 3, Initial Cluster Planning.

Planning Shared Disk Devices

Adding the Disk Configuration to the Cluster Diagram

2. Fill in the host and adapter information including the **node name**, the number of the **slot** in which the disk adapter is installed and the **logical name** of the adapter, such as scsi0. AIX assigns the logical name when the adapter is configured.
3. Assign SCSI IDs for all the devices connected to the SCSI bus. For disk arrays, the controller on the disk array are assigned the SCSI ID.
4. Record information about the LUNs configured on the disk array.
5. Record the logical device name AIX assigned to the array controllers when it was configured. If you have configured an IBM RAIDiant disk array, you can optionally configure the REACT software that configures a pseudo-device called a Disk Array Router.

Completing the IBM Serial Storage Architecture Disk Subsystems Worksheet

Complete a worksheet in Appendix A for each shared SSA configuration.

1. Enter the cluster ID and the Cluster name in the appropriate fields. This information was determined in Chapter 3, Initial Cluster Planning.
2. Fill in the host and adapter information including the **node name**, the SSA adapter label, and the number of the **slot** in which the disk adapter is installed. Include dual-port number of the connection. This will be needed to make the loop connection clear.
3. Determine the SCSI IDs for all the devices connected to the SCSI bus.

Adding the Disk Configuration to the Cluster Diagram

Once you have chosen a disk technology, add your disk configuration to the cluster diagram you started in Chapter 3, Initial Cluster Planning.

For the cluster diagram, draw a box representing each shared disk; then label each box with a shared disk name.

Where You Go From Here

You have now planned your shared disk configuration. The next step is to plan the shared volume groups for your cluster. This step is described in Chapter 6, Planning Shared LVM Components.

Chapter 6 Planning Shared LVM Components

This chapter describes planning shared volume groups for an HACMP/ES cluster.

Prerequisites

By now, you should have completed the planning steps in the previous chapters:

- Planning applications (Chapter 3, Initial Cluster Planning)
- Planning networks (Chapter 4, Planning Cluster Network Connectivity)
- Planning disk configuration (Chapter 5, Planning Shared Disk Devices)

This chapter discusses LVM issues as they relate to the HACMP/ES environment. It does not provide an exhaustive discussion of LVM concepts and facilities in general. Refer to the *AIX System Management Guide* for more information on the AIX LVM.

Overview

Planning shared LVM components for an HACMP/ES cluster differs depending on the method of shared disk access and the type of shared disk device. This discussion assumes the goal of no single point of failure. Data redundancy through the use of mirroring or RAID devices should always be used.

Planning for Non-Concurrent Access

Non-concurrent access configurations typically use journaled filesystems. (In some cases, a database application running in a non-concurrent environment may bypass the journaled filesystem and access the raw logical volume directly.)

Configurations that use IBM 7135-210, 7135-110 RAIDiant disk arrays or IBM 2105-B09, 2105-100 Versatile Storage Servers do not use LVM mirroring. These systems provide their own data redundancy.

Planning for Concurrent Access

Concurrent access configurations do not support journaled filesystems. Concurrent access configurations that use IBM 7131-405 and 7133 SSA serial disk subsystems should use LVM mirroring.

Concurrent access configurations that use IBM 7135-110, 7135-210 RAIDiant Disk Arrays or IBM 2105-B09, 2105-100 Versatile Storage Servers do not use LVM mirroring. Instead, these systems provide their own data redundancy.

This chapter presents information necessary for both methods of shared disk access and points out differences where applicable. After presenting various planning considerations and guidelines, this chapter provides instruction for completing the shared LVM components worksheets for the appropriate method of shared disk access.

Cluster Object	Eliminated as Single Point of Failure by...
LVM components	Using LVM mirroring or mirroring provided by RAID devices

LVM Components in the HACMP/ES Environment

The LVM controls disk resources by mapping data between physical and logical storage. *Physical storage* refers to the actual location of data on a disk. *Logical storage* controls how data is made available to the user. Logical storage can be discontinuous, expanded, replicated, and can span multiple physical disks. These features provide improved availability of data.

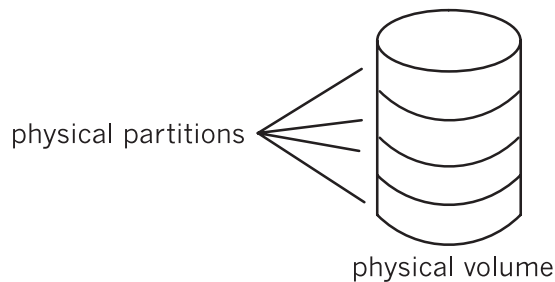
The LVM organizes data into the following components:

- Physical volumes
- Volume groups
- Logical volumes
- filesystems.

Considerations for each component as it relates to planning an HACMP/ES cluster are discussed below.

Physical Volumes

A physical volume is a single physical disk. The physical volume is partitioned to provide AIX with a way of managing how data is mapped to the volume. The following diagram shows how the physical partitions within a physical volume are conventionally diagrammed.

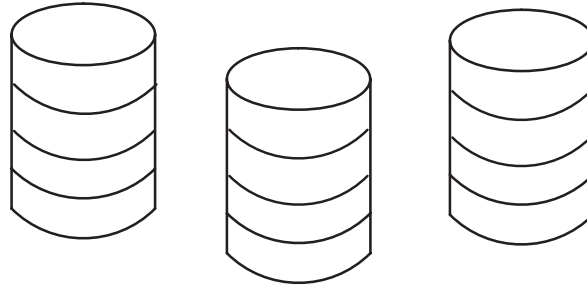


Physical Partitions on a Physical Volume

Volume Groups

A volume group is a set of physical volumes that AIX treats as a contiguous, addressable disk region. You can place from 1 to 32 physical volumes in the same volume group.

The following diagram shows a volume group of three physical volumes.



A Volume Group of Three Physical Volumes

In the HACMP/ES environment, a *shared volume group* is a volume group that resides entirely on the external disks shared by the cluster nodes. A shared volume group can be varied on by only one node at a time. Do not include an internal disk in a shared volume group, since it cannot be accessed by other nodes. If you include an internal disk in a shared volume group, the **varyonvg** command fails.

The shared volume groups in an HACMP/ES cluster should not be activated (varied on) automatically at system boot, but by cluster event scripts.

Note: Each volume group that has filesystems residing on it has a log logical volume (**jfslog**) that must also have a unique name.

Logical Volumes

A logical volume is a set of logical partitions that AIX makes available as a single storage unit—that is, the logical view of a disk. A logical partition is the logical view of a physical partition. Logical partitions may be mapped to one, two, or three physical partitions to implement mirroring.

In the HACMP/ES environment, logical volumes can be used to support a journaled filesystem or a raw device.

Filesystems

A filesystem is written to a single logical volume. Ordinarily, you organize a set of files as a filesystem for convenience and speed in managing data.

In the HACMP/ES system, a *shared filesystem* is a journaled filesystem that resides entirely in a shared logical volume.

You want to plan shared filesystems to be placed on external disks shared by cluster nodes. Data resides in filesystems on these external shared disks in order to be made highly available.

LVM Mirroring

This section does not apply to the IBM 7135-110, 7135-210 RAIDiant Disk Arrays or IBM 2105-B09, 2105-100 Versatile Storage Servers which provide their own data redundancy.

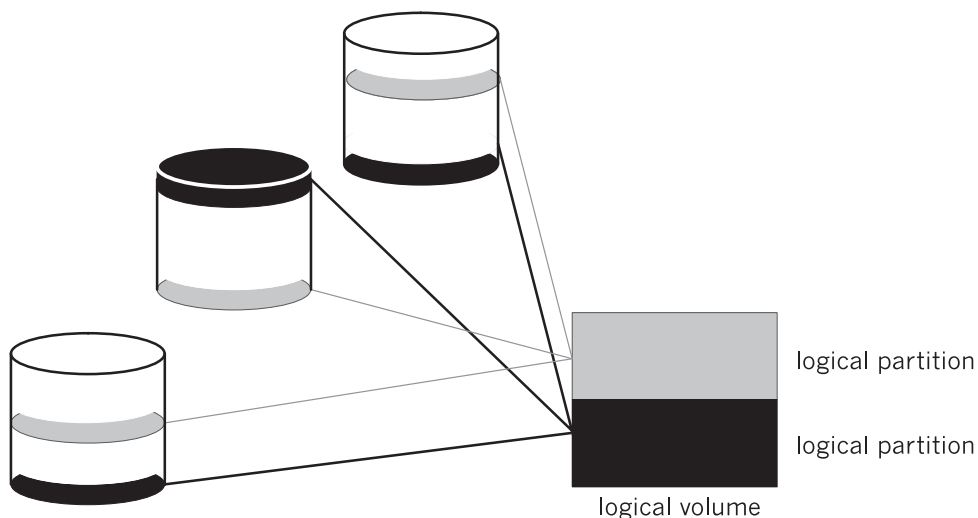
LVM mirroring provides the ability to specify more than one copy of a physical partition. Using the LVM mirroring facility increases the availability of the data in your system. When a disk fails and its physical partitions become unavailable, you still have access to the data if there is a mirror on an available disk. The LVM performs mirroring within the logical volume. Within an HACMP/ES cluster, you mirror:

- Logical volume data in a shared volume group
- Log logical volume data for each shared volume group with filesystems.

Mirroring Physical Partitions

For a logical volume, you allocate one, two, or three copies of the physical partition that contains data. This allocation lets you mirror data, which improves the availability of the logical volume. If a copy is lost due to an error, the other unaffected copies are accessed, and AIX continues processing with an accurate copy. After access is restored to the failed physical partition, AIX resynchronizes the contents (data) of the physical partition with the contents (data) of a consistent mirror copy.

The following diagram shows a logical volume composed of two logical partitions with three mirrored copies. In the diagram, each logical partition maps to three physical partitions. Each physical partition should be designated to reside on a separate physical volume within a single volume group. This provides the maximum number of alternative paths to the mirror copies and, therefore, the greatest availability.



A Logical Volume of Two Logical Partitions with Three Mirrored Copies

The mirrored copies are transparent, meaning that you cannot isolate one of these copies. For example, if a user deletes a file from a logical volume with multiple copies, the deleted file is gone from all copies of the logical volume.

Using mirrored copies improves the availability of data on your cluster. The following considerations also improve data availability:

- Allocating three copies in a logical partition provides greater protection than allocating one or two copies.
- Allocating the copies of a logical partition on different physical volumes provides greater protection than allocating the copies on the same physical volume.
- Allocating the copies of a logical partition on different adapters provides greater protection than allocating the copies on a single adapter.

Keep in mind that anything that improves availability may increase the time necessary for write operations. Nevertheless, using mirrored copies spanning multiple disks (on separate power supplies) together with multiple adapters ensures that no disk is a single point of failure for your cluster.

Mirroring Journal Logs

This section applies to non-concurrent access configurations, which support journaled filesystems. AIX uses journaling for its filesystems. In general, this means that the internal state of a filesystem at startup (in terms of the block list and free list) is the same state as at shutdown. In practical terms, this means that when AIX starts up, the extent of any file corruption can be no worse than at shutdown.

Each volume group contains a **jfslog**, which is itself a logical volume. This log typically resides on a different physical disk in the volume group than the journaled filesystem. If access to that disk is lost, however, changes to filesystems after that point are in jeopardy.

To avoid the possibility of that physical disk being a single point of failure, you can specify mirrored copies of each **jfslog**. Place these copies on separate physical volumes.

Quorum

This section does not apply to the IBM 7135-110, 7135-210 RAIDiant Disk Arrays or IBM 2105-B09, 2105-100 Versatile Storage Servers which provide their own data redundancy. In these systems, where a single volume group can contain multiple LUNs (collections of physical disks) that appear to the host as a single device (hdisk), quorum is not an issue because of the large storage capacity the LUNs provide and because of the data redundancy capabilities of the array.

Quorum is a feature of the AIX LVM that determines whether or not a volume group can be placed on-line, using the **varyonvg** command, and whether or not it can remain on-line after a failure of one or more of the physical volumes in the volume group.

Each physical volume in a volume group has a Volume Group Descriptor Area (VGDA) and a Volume Group Status Area (VGSA).

VGDA Describes the physical volumes (PVs) and logical volumes (LVs) that make up a volume group and maps logical partitions to physical partitions. The **varyonvg** command reads information from this area.

VGSA Maintains the status of all physical volumes and physical partitions in the volume group. It stores information regarding whether a physical partition is potentially inconsistent (stale) with mirror copies on other physical partitions, or is consistent or synchronized with its mirror copies. Proper functioning of LVM mirroring relies upon the availability and accuracy of the VGSA data.

Quorum at Varyon

When a volume group is brought on-line using the **varyonvg** command, VGDA and VGSA data structures are examined. If more than half of the copies are readable and identical in content, quorum is achieved and the **varyonvg** command succeeds. If exactly half the copies are available, as with two of four, quorum is not achieved and the **varyonvg** command fails.

Quorum after Varyon

If a write to a physical volume fails, the VGSA's on the other physical volumes within the volume group are updated to indicate this physical volume has failed. As long as more than half of all VGDA's and VGSA's can be written, quorum is maintained and the volume group remains varied on. If exactly half or less than half of the VGDA's and VGSA's are inaccessible, quorum is lost, the volume group is varied off, and its data becomes unavailable.

Keep in mind that a volume group can be varied on or remain varied on with one or more of the physical volumes unavailable. However, data contained on the missing physical volume will not be accessible unless the data is replicated using LVM mirroring and a mirror copy of the data is still available on another physical volume. Maintaining quorum without mirroring does not guarantee that all data contained in a volume group is available.

Quorum has nothing to do with the availability of mirrored data. It is possible to have failures that result in loss of all copies of a logical volume, yet the volume group remains varied on as a quorum of VGDA's/VGSA's are still accessible.

Disabling and Enabling Quorum

Quorum checking is enabled by default. Quorum checking can be disabled using the **chvg -Qn vgname** command, or by using the **smit chvg** fastpath.

Quorum Enabled

With quorum enabled, more than half of the physical volumes must be available and the VGDA and VGSA data structures must be identical before a volume group can be varied on with the **varyonvg** command.

With quorum enabled, a volume group will be forced off-line if one or more disk failures causes a majority of the physical volumes to be unavailable. Having three or more disks in a volume group avoids a loss of quorum in the event of a single disk failure.

Quorum Disabled

With quorum disabled, *all* the physical volumes in the volume group must be available and the VGDA data structures must be identical for the **varyonvg** command to succeed. With quorum disabled, a volume group will remain varied on until the last physical volume in the volume group becomes unavailable. The following information summarizes the effect quorum has on the availability of a volume group.

Forcing a Varyon

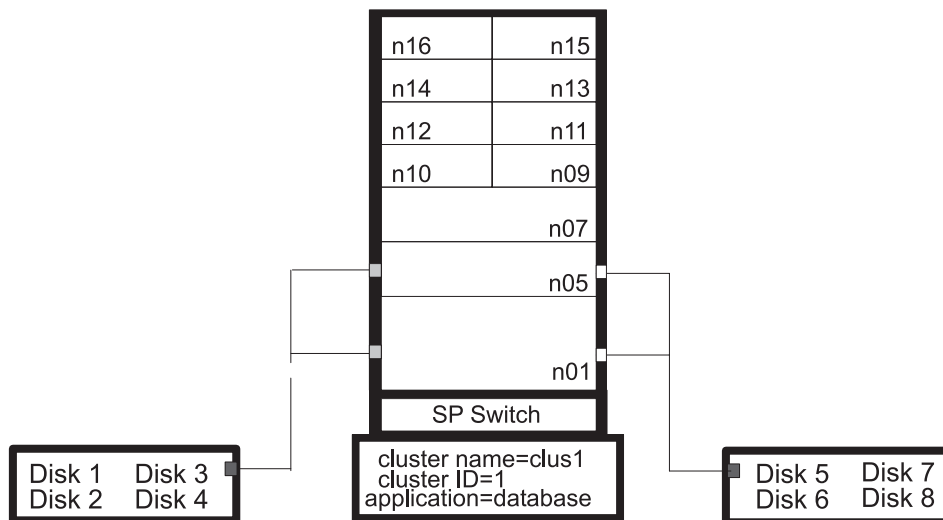
A volume group with quorum disabled and one or more physical volumes unavailable can be “forced” to vary on by using the **-f** flag with the **varyonvg** command. Forcing a varyon with missing disk resources could cause unpredictable results, including a **reducevg** of the physical volume from the volume group. Forcing a varyon should be an overt (manual) action and should only be performed with a complete understanding of the risks involved.

The HACMP/ES software assumes that a volume group is not degraded and all physical volumes are available when the **varyonvg** command is issued at startup or when a volume group resource is taken over during a failover. The cluster event scripts provided with the HACMP/ES software do not “force” varyon with the **-f** flag, which could cause unpredictable results. For this reason, modifying the cluster event scripts to use the **-f** flag is strongly discouraged.

Using Quorum in Non-Concurrent Access Configurations

While specific scenarios can be constructed where quorum protection does provide some level of protection against data corruption and loss of availability, quorum provides very little actual protection. In fact, enabling quorum may mask failures by allowing a volume group to varyon with missing resources. Also, designing logical volume configuration for no single point of failure with quorum enabled may require the purchase of additional hardware. Although these facts are true, you must keep in mind that disabling quorum can result in subsequent loss of disks—after varying on the volume group—that go undetected.

Often it is not practical to configure disk resources as shown below because of the expense. Take, for example, a cluster that requires 8 GB of disk storage (4 GB double mirrored). This requirement could be met with two disk subsystems and two disk adapters in each node. For data availability reasons, logical volumes would be mirrored across disk subsystems.



Quorum in HACMP/ES Non-Concurrent Configurations

With quorum enabled, the failure of a single adapter, cable, or disk subsystem power supply would cause exactly half the disks to be inaccessible. Quorum would be lost and the volume group varied off even though a copy of all mirrored logical volumes is still available. One solution is to turn off quorum checking for the volume group. The trade-off is that, with quorum disabled, all physical volumes must be available for the **varyonvg** command to succeed.

Using a Quorum Buster Disk

You may want to mirror the data between two 7133 drawers. Be careful to set up the configuration so that no adapter or enclosure is a single point of failure for access to the data. In order to avoid the problem mentioned above, where a power failure results in the LVM varying off the volume group (if quorum is enabled), consider using a “quorum buster” disk.

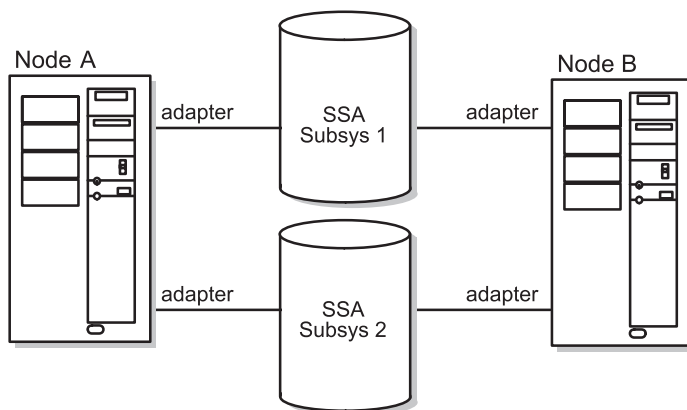
Add a single additional disk to the volume group, on a separate power and FRU boundary from either of the mirrors of the data. This disk contains no data, it simply serves as a “quorum buster” so that if one enclosure fails, or connectivity to it is lost, quorum is maintained and the data remains available on the other enclosure.

Using the quorum buster disk also ensures that, in the rare case where a clean failover does not occur, only one node has access to the quorum disk and thus you avoid the scenario where two nodes battle over ownership of a shared disk. Only one node will be able to varyon the volume group and access the data.

Using Quorum in Concurrent Access Configurations

Quorum must be enabled for an HACMP/ES concurrent access configuration. Disabling quorum could result in data corruption. Any concurrent access configuration where multiple failures could result in no common shared disk between cluster nodes has the potential for data corruption or inconsistency.

Take, for example a cluster with two sets of IBM SSA disk subsystems configured for no single point of failure. In this configuration, logical volumes are mirrored across subsystems and each disk subsystem is connected to each node with separate adapters as shown in the following figure.



An IBM SSA Concurrent Access Configuration

If multiple failures result in a communications loss between each node and one set of disks in such a way that Node A can access subsystem 1 but not subsystem 2 and Node B can access subsystem 2 but not subsystem 1. Both nodes continue to operate on the same baseline of data from the mirror copy they can access. However, each node will cease to see modifications to data on disk that the other node makes. The result is that the data diverges and becomes inconsistent between nodes.

On the other hand, if quorum protection is enabled, the communications failure results in one or both nodes varying off the volume group. Although this is a harsh action as far as the application is concerned, data consistency is not compromised.

Using NFS with HACMP/ES

In HACMP/ES version 4.4, functionality previously available in the High Availability for Network File System for AIX (HANFS for AIX) product has been added to the basic HACMP/ES architecture. The following enhancements are included:

- Reliable NFS server capability that allows a backup processor to recover current NFS activity should the primary NFS server fail, preserving the locks on NFS filesystems and duncache. *This functionality is available for 2-node clusters only.*
- Ability to specify a network for NFS mounting.
- Ability to define NFS exports and mounts at the directory level.
- Ability to specify export options for NFS-exported directories and filesystems.

In order for NFS to work as expected on an HACMP cluster, you must be aware of certain configuration issues, so you can plan accordingly:

- Creating shared volume groups
- Exporting NFS filesystems
- NFS Mounting and Fallover.

The HACMP for AIX scripts have only minimal NFS support. You may need to modify them to handle your particular configuration. The following sections contain some suggestions for planning for a variety of issues.

Reliable NFS Server Capability

An HACMP/ES two-node cluster can take advantage of AIX extensions to the standard NFS functionality that enable it to handle duplicate requests correctly and restore lock state during NFS server fallover and reintegration. This support was previously only available in the HANFS feature. More detail can be found in the `/usr/lpp/cluster/doc/release_notes`.

Creating Shared Volume Groups

When creating shared volume groups, normally you can leave the **Major Number** field blank and let the system provide a default for you. However, if you are using NFS, all nodes in your cluster must have identical major numbers, as HACMP uses the major number as part of the file handle to uniquely identify a Network Filesystem.

In the event of node failure, NFS clients attached to an HACMP/ES cluster operate exactly the way they do when a standard NFS server fails and reboots. If the major numbers are not the same, when another cluster node takes over the filesystem and re-exports the filesystem, the client application will not recover, since the filesystem exported by the node will appear to be different from the one exported by the failed node.

NFS Exporting Filesystems and Directories

The process of NFS-exporting filesystems and directories in HACMP for AIX is different from that in AIX. Remember the following when planning for NFS-exporting in HACMP:

- **Specifying Filesystems and Directories to NFS Export:** In AIX, you specify filesystems and directories to NFS-export by using the `smit mknfsexp` command (which creates the `/etc/exports` file). In HACMP for AIX, you specify filesystems and directories to NFS-export by including them in a resource group in the HACMP SMIT **NFS Filesystems/Directories to export** field.
- **Specifying Export Options for NFS Exported Filesystems and Directories:** If you want to specify special options for NFS-exporting in HACMP, you can create a `/usr/sbin/cluster/etc/exports` file. This file has the same format as the regular AIX `/etc/exports` file.

Note: Use of this alternate exports file is optional. HACMP checks the `/usr/sbin/cluster/etc/exports` file when NFS-exporting a filesystem or directory. If there is an entry for the filesystem or directory in this file, HACMP will use the options listed. If the filesystem or directory for NFS-export is not listed in the file, or, if the alternate file does not exist, the filesystem or directory will be NFS-exported with the default option of root access for all cluster nodes.

NFS Mounting and Fallover

For HACMP/ES and NFS to work properly together, you must be aware of the following mount issues.

- To NFS mount, a resource group must be configured with IPAT.
- If you want to use the Reliable NFS Server capability that preserves NFS locks and the dupcache in two-node clusters, the IPAT adapter for the resource group must be configured to use Hardware Address Takeover.

Cascading Takeover with Cross Mounted NFS Filesystems

This section describes how to set up cascading resource groups with cross mounted NFS filesystems so that NFS filesystems are handled gracefully during takeover and reintegration.

Note: Only cascading resource groups support automatic NFS mounting across servers during fallover. Rotating resource groups do not provide this support. Instead, you must use additional post events or perform NFS mounting using normal AIX routines.

Creating NFS Mount Points on Clients

A mount point is required in order to mount a filesystem via NFS. In a cascading resource group, all the nodes in the resource group will NFS mount the filesystem; thus you must create a mount point on each node in the resource group. On each of these nodes, create a mount point by executing the following command:

```
mkdir /mountpoint
```

where *mountpoint* is the name of the local mountpoint over which the remote filesystem will be mounted.

Setting Up NFS Mount Point Different from Local Mount Point

HACMP handles NFS mounting in cascading resource groups as follows: The node that currently owns the resource group will mount the filesystem over the filesystem's local mount point, and this node will NFS export the filesystem. All the nodes in the resource group (including the current owner of the group) will NFS mount the filesystem over a different mount point. Therefore the owner of the group will have the filesystem mounted twice - once as a local mount, and once as an NFS mount.

Since IPAT is used in resource groups that have NFS mounted filesystems, the nodes will not unmount and remount NFS filesystems in the event of a fallover. When the resource group falls over to a new node, the acquiring node will locally mount the filesystem and NFS-export it. (The NFS mounted filesystem will be temporarily unavailable to cluster nodes during fallover.) As soon as the new node acquires the IPAT label, access to the NFS filesystem is restored.

All applications must reference the filesystem through the NFS mount. If the applications used are dependent upon always referencing the filesystem by the same mount point name, you can change the mount point for the local filesystem mount (for example, change it to mount point `_local`, and use the previous local mount point as the new NFS mount point.

In the **Change/Show Resources/Attributes for a Resource Group** SMIT screen, the **Filesystem to NFS Mount** field must specify both mount points. Put the nfs mount point, then the local mount point, separating the two with a semicolon: for example “`nfspoint;nfslocalpoint`.” If there are more entries, separate them with a space:

```
nfspoint1;local1 nfspoint2;local2
```

If there are nested mount points, the nfs mount points should be nested in the same manner as the local mount points so that they match up properly. When cross mounting NFS filesystems, you must also set the “*filesystems mounted before IP configured*” field of the Resource Group to **true**.

Server-to-Server NFS Cross Mounting: Example

HACMP/ES allows you to configure a cluster so that servers can NFS-mount each other’s filesystems. Configuring cascading resource groups allows the Cluster Manager to decide which node should take over a failed resource, based on priority and node availability.

Ensure that the shared volume groups have the same major number on the server nodes. This allows the clients to re-establish the NFS-mount transparently after the takeover.

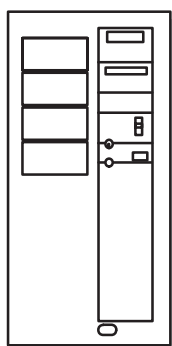
In the example cluster shown below, you have two resource groups, `NodeA_rg` and `NodeB_rg`. These resource groups are defined in SMIT as follows:

Resource Group	<i>NodeA_rg</i>
Participating node names	Node A Node B
Filesystems	<i>/afs</i> (filesystems to be locally mounted by node currently owning the resource group)
Filesystems to export	<i>/afs</i> (Filesystem to NFS-export by node currently owning resource group. Filesystem is subset of filesystem listed above.)
Filesystems to NFS mount	<i>/mountpointa;/afs</i> (Filesystems/directories to be NFS-mounted by all nodes in the resource group. First value is NFS mount point; second value is local mount point)
Resource Group	<i>NodeB_rg</i>
Participating node names	Node B Node A
Filesystems	<i>/bfs</i>
Filesystems to export	<i>/bfs</i>
Filesystems to NFS mount	<i>/mountpointb;/bfs</i>

The filesystem you want the local node (Node A) in this resource group to locally mount and export is **/afs**, on Node A. You want the remote node (Node B) in this resource group to NFS-mount **/afs**, from Node A.

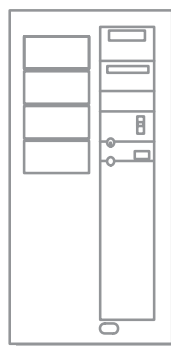
Setting up your cascading resource groups like this ensures the expected default server-to-server NFS behavior described above. On reintegration, **/afs** is passed back to Node A, locally mounted and exported. Node B mounts it via NFS again.

When the cluster as originally defined is up and running on both nodes, the filesystems are mounted as shown:



Node A

/afs locally mounted
/afs NFS exported
a_svc:/afs NFS mounted over /mountpointa
b_svc:/bfs NFS mounted over /mountpointb



Node B

/bfs locally mounted
/bfs NFS exported
b_svc:/bfs NFS mounted over /mountpointb
a_svc:/afs NFS mounted over /mountpointa

Cross-Mounted Nodes, Normal Operation

When Node A fails, Node B uses the **cl_nfskill** utility to close open files in Node A:/afs, unmounts it, mounts it locally, and re-exports it to waiting clients.

After takeover, Node B has:

- **/bfs** locally mounted
- **/bfs** NFS-exported
- **/afs** locally mounted
- **/afs** NFS-exported
- **a_svc:/afs** NFS mounted over **/mountpointa**
- **b_svc:/bfs** NFS mounted over **/mountpointb**

See the man page in **/usr/sbin/cluster/events/utlils** for information about the usage and syntax for the **cl_nfskill** command.

Caveats about Node Names and NFS

In the configuration described above the node name is used as the NFS hostname for the mount. This can fail if the node name is not a legitimate TCP/IP adapter label.

To avoid this problem do one of the following:

- Ensure that node name and the service adapter label are the same on each node in the cluster; *or*
- Alias the node name to the service adapter label in the `/etc/hosts` file.

LVM Planning Considerations

Consider the following guidelines as you plan shared LVM components.

- In general, planning for logical volumes concerns the availability of your data. However, creating logical volume copies is not a substitute for regularly scheduled backups. Backups protect against loss of data regardless of cause; logical volume copies protect against loss of data from physical access failure.
- All operating system files should reside in the root volume group (**rootvg**) and all user data should reside outside that group. This makes it more manageable to update or reinstall the operating system and to back up data.
- Volume groups that contain at least three physical volumes provide the maximum availability when implementing mirroring.
- When using copies, each physical volume containing a copy should get its power from a separate source. If one power source fails, separate power sources maintain the no single point of failure objective.
- Consider quorum issues when laying out a volume group. With quorum enabled, a two-disk volume group puts you at risk for losing quorum and data access. Either build three-disk volume groups or disable quorum.
- Plan for NFS mounted filesystems and directories.
- Keep in mind the cluster configurations that you have designed. A node whose resources are not taken over should not own critical volume groups.

Completing the Shared LVM Components Worksheets

You can now fill out a set of worksheets that help you plan the physical and logical storage for your cluster. Refer to the completed worksheets when you define the shared LVM components following the instructions in Chapter 12, Defining Shared LVM Components, and the cluster resource configuration following the instructions in Chapter 18, Configuring an HACMP/ES Cluster.

Appendix A, Planning Worksheets, has blank copies of the worksheets discussed below. Photocopy these worksheets before continuing.

Complete the following worksheets to plan the volume groups and filesystems for your cluster:

- Non-Shared Volume Group Worksheet
- Shared Volume Group/filesystem Worksheet

- NFS-Exported Filesystem/Directory Worksheet

If you plan to use the online cluster planning worksheets, refer to Appendix B, Using the Online Cluster Planning Worksheet Program, for instructions.

Completing the Non-Shared Volume Group Worksheet

For each node in the cluster, complete a Non-Shared Volume Group Worksheet for each volume group residing on a local (non-shared) disk.

1. Fill in the node name in the **Node Name** field.
2. Record the name of the volume group in the **Volume Group Name** field.
3. List the device names of the physical volumes comprising the volume group in the **Physical Volumes** field.

In the remaining sections of the worksheet, enter the following information for each logical volume in the volume group. Use additional sheets if necessary.

4. Enter the name of the logical volume in the **Logical Volume Name** field.
5. If you are using LVM mirroring, indicate the number of logical partition copies (mirrors) in the **Number Of Copies Of Logical Partition** field. You can specify one or two copies (in addition to the original logical partition, for a total of three).
6. If you are using LVM mirroring, specify whether each copy will be on a separate physical volume in the **On Separate Physical Volumes?** field.
7. Record the full-path mount point of the filesystem in the **filesystem Mount Point** field.
8. Record the size of the filesystem in 512-byte blocks in the **Size** field.

Completing the Shared Volume Group/Filesystem Worksheet

Fill out a Shared Volume Group/filesystem Worksheet for each volume group that will reside on the shared disks. You need a separate worksheet for each shared volume group, so be sure to make sufficient copies of the worksheet before you begin.

1. Fill in the name of each node in the cluster in the **Node Names** field. You determined the node names in Chapter 3, Initial Cluster Planning.
2. Assign a name to the shared volume group and record it in the **Shared Volume Group Name** field.

The name of a shared volume group must be unique. It cannot conflict with the name of an existing volume group on any node in the cluster.

3. Leave the **Major Number** field blank for now. You will enter a value in this field when you address NFS issues in the following Chapter 7, Planning Resource Groups.
4. Record the name of the log logical volume (**jfslog**) in the **Log Logical Volume Name** field.
5. Pencil-in the planned physical volumes in the **Physical Volumes** field. You will enter exact values for this field after you have installed the disks following the instructions in Chapter 11, Checking Installed Hardware.

Planning Shared LVM Components

Completing the Shared LVM Components Worksheets

Physical volumes are known in the AIX operating system by sequential **hdisk** numbers assigned when the system boots. For example, `/dev/hdisk0` identifies the first physical volume in the system, `/dev/hdisk1` identifies the second physical volume in the system, and so on.

When sharing a disk in an HACMP/ES cluster, the nodes sharing the disk each assign an **hdisk** number to that disk. These **hdisk** numbers may not match, but refer to the same physical volume. For example, each node may have a different number of internal disks, or the disks may have changed since AIX was installed.

The HACMP/ES software does not require that the **hdisk** numbers match across nodes (though your system is easier to manage if they do). In situations where the **hdisk** numbers must differ, be sure that you understand each node's view of the shared disks. Draw a diagram that indicates the **hdisk** numbers that each node assigns to the shared disks and record these numbers on the appropriate volume group worksheets in Appendix A. When in doubt, use the **hdisk**'s PVID to verify its identity on a shared bus.

In the remaining sections of the worksheet, enter the following information for each logical volume in the volume group. Use additional sheets as necessary.

6. Assign a name to the logical volume and record it in the **Logical Volume Name** field.
A shared logical volume must have a unique name within an HACMP/ES cluster. By default, AIX assigns a name to any logical volume that is created as part of a journaled filesystem (for example, `lv01`). If you rely on the system generated logical volume name, this name could cause the import to fail when you attempt to import the volume group containing the logical volume into another node's ODM structure, especially if that volume group already exists. Chapter 12, Defining Shared LVM Components, describes how to change the name of a logical volume.
7. *This step does not apply to the IBM 7135 RAIDiant disk array or IBM 2105 Versatile Storage Server.* If you are using LVM mirroring, indicate the number of logical partition copies (mirrors) in the **Number Of Copies of Logical Partition** field. You can specify that you want one or two copies (in addition to the original logical partition, for a total of three).
8. *This step does not apply to the IBM 7135 RAIDiant disk array or IBM 2105 Versatile Storage Server.* If you are using LVM mirroring, specify whether each copy will be on a separate physical volume in the **On Separate Physical Volumes?** field.
9. Record the full-path mount point of the filesystem in the **filesystem Mount Point** field.
10. Record the size of the filesystem in 512-byte blocks in the **Size** field.

Completing the NFS-Exported Filesystem Worksheet

Complete an NFS-Exported Filesystem or Directory Worksheet for filesystems or directories to be NFS-exported from a node. The information you provide will be used to update the `/usr/sbin/cluster/etc/exports` file.

1. Record the name of the resource group from which the filesystems or directories will be NFS exported in the **Resource Group** field.
2. In the **Network for NFS Mount** field record the preferred network to NFS mount the filesystems or directories.

3. In the **Filesystem Mounted Before IP Configured** field, write *true* if you want the takeover of filesystems to occur before the takeover of IP address(es). Specify *false* for the IP address(es) to be taken over first.
4. Record the full pathname of the filesystem or directory to be exported in the **Exported Directory** field.
5. (optional) Record the export options you want to assign the directories and/or filesystems to be NFS exported. Refer to the **exports** man page for a full list of export options.
6. Repeat steps 4 and 5 for each filesystem or directory to be exported.

Concurrent Access Worksheets

Complete the following worksheets to plan the volume groups for a concurrent access configuration:

- Non-Shared Volume Group Worksheet
- Shared Volume Group Worksheet

Completing the Non-Shared Volume Group Worksheet (Concurrent Access)

For each node, complete a Non-Shared Volume Group Worksheet for each volume group that resides on a local (non-shared) disk.

1. Fill in the node name in the **Node Name** field.
2. Record the name of the volume group in the **Volume Group Name** field.
3. Enter the name of the logical volume in the **Logical Volume Name** field.
4. List the device names of the physical volumes that comprise the volume group in the **Physical Volumes** field.

In the remaining sections of the worksheet, enter the following information for each logical volume in the volume group. Use additional sheets if necessary.

5. Enter the name of the logical volume in the **Logical Volume Name** field.
6. If you are using LVM mirroring, indicate the number of logical partition copies (mirrors) in the **Number Of Copies Of Logical Partition** field. You can specify one or two copies (in addition to the original logical volume, for a total of three).
7. If you are using LVM mirroring, specify whether each copy will be on a separate physical volume in the **On Separate Physical Volumes?** field.
8. Record the full-path mount point of the filesystem in the **filesystem Mount Point** field.
9. Record the size of the filesystem in 512-byte blocks in the **Size** field.

Completing the Shared Volume Group Worksheet (Concurrent Access)

Fill out a Shared Volume Group Worksheet for each volume group that will reside on the shared disks. You need a separate worksheet for each shared volume group, so be sure to make sufficient copies of the worksheet before you begin.

If you plan to create concurrent volume groups on SSA disk subsystem, assign unique non-zero node numbers with *ssar* on each cluster node before using the concurrent volume group failed drive replacement procedure.

Planning Shared LVM Components

Adding LVM Information to the Cluster Diagram

If you specify the use of SSA disk fencing in your concurrent resource group, HACMP/ES assigns the node numbers when you synchronize the resources.

If you don't specify the use of SSA disk fencing in your concurrent resource group, assign the node numbers with

```
chdev -l ssar -a node_number=x
```

where *x* is the number to assign to that node. Then reboot the system.

1. Fill in the name of each node in the cluster in the **Node Names** field.
2. Record the name of the shared volume group in the **Shared Volume Group Name** field.
3. Pencil in the planned physical volumes in the **Physical Volumes** field. You will enter exact values for this field after you have installed the disks following the instructions in Chapter 10, Overview: Installing and Configuring HACMP/ES.

In the remaining sections of the worksheet, enter the following information for each logical volume in the volume group. Use additional sheets as necessary.

4. Enter the name of the logical volume in the **Logical Volume Name** field.
5. *This step does not apply to the IBM 7135 RAIDiant disk array.* Indicate the number of logical partition copies (mirrors) in the *Number Of Copies Of Logical Partition* field. You can specify one or two copies (in addition to the original logical partition, for a total of three).
6. *This step does not apply to the IBM 7135 RAIDiant disk array.* Specify whether each copy will be on a separate physical volume in the **On Separate Physical Volumes?** field.

Adding LVM Information to the Cluster Diagram

Add the LVM information to the cluster diagram, including volume group and logical volume definitions.

Where You Go From Here

You have now planned the shared LVM components for your cluster. Use this information when you define the volume groups, logical volumes, and filesystems during the install.

In the next step of the planning process, you address issues relating to planning for your resource groups. Chapter 7, Planning Resource Groups, describes this step of the planning process.

Chapter 7 Planning Resource Groups

This chapter describes how to plan resource groups within an HACMP/ES cluster.

Prerequisites

By now, you should have completed the planning steps in the previous chapters:

- planning applications (Chapter 3, Initial Cluster Planning)
- planning networks (Chapter 4, Planning Cluster Network Connectivity)
- planning disk configuration (Chapter 5, Planning Shared Disk Devices)
- planning shared volume groups (Chapter 6, Planning Shared LVM Components)

Overview

At the beginning of the planning process in Chapter 3, Initial Cluster Planning, you made preliminary choices about the type of resource group configuration—cascading, rotating, or concurrent—and the takeover priority for each node in the resource chains. In this section you:

- Identify the individual resources that constitute the resource group.
- Define the resource chain for the resource group. A *resource chain* consists of the nodes assigned to participate in the takeover of a given resource group.

Planning Considerations

The HACMP/ES software does not restrict the number of individual resources in a resource group or the number of resource groups in a cluster. Nevertheless, as a general rule you should strive to keep your design as simple as possible. Doing so makes it easier to configure and maintain resource groups; it also makes the takeover of a resource group faster. In general, keep the following considerations in mind:

- Every cluster resource must be part of a resource group. If you want a resource to be kept separate, you define a group for that resource alone. A resource group may have one or more resources defined.
- A resource may *not* be included in more than one resource group.
- The components of a resource group must be unique down to the physical volume. If applications access the same data, put them in the same resource group. If applications access different data, put them in different resource groups.
- A rotating resource group must have a service IP label defined for it.
- A cascading resource group may or may not include a service IP label. If you want to do IP address takeover, then you must include a service label in the resource group.
- If you include the same node in more than one resource chain, make sure that it has the ability to manage all resource groups simultaneously.

- A resource group based on cascading or rotating resources cannot include any concurrent volume groups.
- Concurrent resource groups consist only of application servers and concurrent volume groups.

Special Considerations when Planning for a CWOFF Resource Group

Keep in mind the following when planning for a cascading without fallback resource group.

Sticky and Non-Sticky Migration in a CWOFF Resource Group

DARE migration is enhanced in a cascading resource group with cascading without fallback set to **true**. A cascading resource group with CWOFF set to **false** supports a DARE migration with the sticky option only. This means that the node to which this resource group migrates becomes the highest priority node until another DARE migration changes this (until a DARE to another node, DARE to stop, or a DARE to default). Resource groups with CWOFF = **true** support both sticky and non-sticky DARE migrations.

Resource Group “Clumping”

A Cascading without fallback resource group tends to remain on the node which acquires it. Some nodes may therefore host many resource groups while others have no resource groups.

You can manage the uneven distribution of resource groups, or “clumping,” in several ways:

- Use DARE Migration to redistribute resource groups after node failure or reintegration.
- Plan resource group participation in order to minimize clumping.
- Set the Inactive Takeover flag to **false** in order to manage clumping during cluster start-up.

CWOFF Resource Group Down Though Highest Priority Node is Up

In a cascading resource group with cascading without fallback set to **true**, the possibility exists that while the highest priority node is up, the resource group is down. This situation can occur if you bring the resource group down by either a graceful shutdown or a **clbare stop** command. If inactive takeover is set to **false** in such a situation, then the resource group will not be acquired by another node (because for inactive takeover = **false**, only the highest priority node will acquire the resource group). Unless you bring the resource group up manually, it will remain in an inactive state. For more information on this issue, see Common Problems and Solutions on page 29-24 of Volume 2.

Completing the *Resource Group Worksheet*

You now fill out a worksheet that helps you plan the resource groups for the cluster. Complete one for the cluster, using additional sheets if necessary. Appendix A, Planning Worksheets, has blank copies of the *Resource Group Worksheet*. Photocopy this worksheet before continuing.

If you plan to use the online cluster planning worksheets, refer to Appendix B, Using the Online Cluster Planning Worksheet Program, for instructions.

Record the cluster ID in the **Cluster ID** field and the cluster name in the **Cluster Name** field. You first assigned these values in Chapter 3, Initial Cluster Planning. The following sections describe how to fill in the following fields for each resource group.

Assign a Name to the Resource Group

Assign a name to the resource group and record it in the **Resource Group Name** field. Use no more than 31 characters. You can use alphabetic or numeric characters and underscores. Duplicate entries are not allowed.

Record Node Information

Record the node/resource relationship (that is, the type of resource for the resource group) in the **Node Relationship** field. Indicate whether the resource group is cascading or rotating. You made this choice in Chapter 3, Initial Cluster Planning.

Record the names of the nodes you want to be members of the resource chain for this resource group in the **Participating Node Names** field. List the node names in order from highest to lowest priority.

List the Resources to be Included in the Resource Group

You have identified the resources in previous chapters. In this section of the Resource Group worksheet, you record the following resources:

- Enter the IP Address in the **IP Address** field if your cluster uses IP Address Takeover.
- List the filesystems to include in this resource group in the **Filesystems** field.
- List in the **Filesystems to Export** field the filesystems in this resource group that should be NFS-exported by the node currently holding the resource. These filesystems should be a subset of the filesystems listed above.
- List which filesystems, of the ones you defined in the Filesystems field, that should be NFS-mounted by the nodes in the resource chain not currently holding the resource in the **Filesystems to NFS Mount** field.
- List in the **Volume Groups** field the shared volume groups that should be varied on when this resource group is acquired or taken over.

Note: If you plan on using raw logical volumes, you only need to specify the volume group in which the raw logical volume resides in order to include the raw logical volumes in the resource group.

- If you plan on using an application that directly accesses raw disks, list the raw disks in the **Raw Disks** field.
- If you are using **AIX Connections Services**, define the realm/service pairs in this field.
- If you are using **AIX Fast Connect Resources**, define the resources in this field.

Note: You cannot configure both AIX Connections and AIX Fast Connect in the same resource group.

- List the **Highly Available Communications Links** if you are using CS/AIX Communications Server.
- List the names of the application servers to include in the resource group in the **Application Servers** field.

Determine Initial Ownership of the Resource Group (Cascading Only)

Indicate in the **Inactive Takeover** field how you want to control the initial acquisition of a resource by a node. (This field applies only to cascading resource groups.)

- If Inactive Takeover is **false**, the first node to join the cluster acquires only those resource groups for which it has been designated the highest priority node. Each subsequent node that joins the cluster acquires all resource groups for which the node has a higher priority than any other node currently up in the cluster. Depending on the order in which the nodes are brought up, this may result in resource groups cascading to higher priority nodes as those nodes join the cluster.

The default is **false**.

- If Inactive Takeover is **true**, then the first node in the resource chain to join the cluster acquires the resource. Subsequently the resource will cascade to nodes in the chain with higher priority as they join the cluster. Note that this will cause an interruption in service as resource ownership transfers to the node with the higher priority.

Cascading without Fallback Attribute

In the **Cascading without Fallback** field, choose how you would like to define this cascading resource group attribute. Cascading without fallback interacts with Inactive Takeover in the following ways.

- If the Cascading resource group configuration **Cascading without Fallback** is set to **true** while the **Inactive Takeover Activated** option is also set to **true**, the resource group acquired by the first node in the resource chain will *not* fallback to nodes with higher ownership priority as they join the cluster.
- If the **Cascading without Fallback** option is set to **true** while **Inactive Takeover Activated** is **false**, a resource group will not migrate from the node which brought it online. The resource group will move to another node only if the owner node leaves the cluster. It will not, however, fall back to the owner node when the owner node comes back online.
- If **Cascading without Fallback** is set to **false**, a cascading resource group will follow the rules associated with Inactive Takeover as described above.

Where You Go From Here

You have now planned the resource groups for the cluster. You next tailor the cluster event processing for your cluster. Chapter 8, *Cluster Events: Tailoring and Creating*, discusses this step.

Chapter 8 Cluster Events: Tailoring and Creating

This chapter describes tailoring and creating cluster event processing for your cluster.

Note: The directory `/usr/sbin/cluster` and subdirectories have symbolic links to the `/usr/es/sbin/cluster` directory and subdirectories. Individual files in those directories are *not* linked, as they were in previous releases.

Prerequisites

By now, you should have completed these steps in planning for your cluster:

- Applications (Chapter 3, Initial Cluster Planning)
- Networks (Chapter 4, Planning Cluster Network Connectivity)
- Disk configuration (Chapter 5, Planning Shared Disk Devices)
- Shared volume groups (Chapter 6, Planning Shared LVM Components)
- Resource groups (Chapter 7, Planning Resource Groups)

Overview

HACMP/ES has two facilities for managing event processing:

- The ability to tailor predefined events
- The ability to define new events.

Tailoring Cluster Event Processing

The Cluster Manager's ability to recognize a specific series of events and subevents permits a very flexible customization scheme. The HACMP/ES event customization facility lets you tailor cluster event processing to your site. Customizing event processing allows you to provide a highly efficient path to the most critical resources in the event of a failure. This efficiency, however, depends on your configuration.

As part of the planning process, you need to decide whether to customize event processing. If the actions taken by the default scripts are sufficient for your purposes, then you do not need to do anything further to configure events during the configuration process.

If you do decide to tailor event processing to your environment, it is strongly recommended that you use the HACMP/ES event customization facility, described in this chapter. If you tailor event processing, you must register these user-defined scripts with HACMP/ES during the configuration process.

If necessary, you can modify the default event scripts or write your own. *Modifying the default event scripts or replacing them with your own is strongly discouraged.* This makes maintaining, upgrading, and troubleshooting an HACMP/ES cluster much more difficult. Again, if you write your own event customization scripts, you need to configure the HACMP/ES software to use those scripts.

The event customization facility includes the following features:

- Event notification
- Pre- and post-event processing
- Event recovery and retry

Complete customization for an event includes a notification to the system administrator (before and after event processing), and user-defined commands or scripts that run before and after event processing, as shown in the figure below.

```
Notify sysadmin of event to be processed
Pre-event script or command
HACMP/ES event script
Post-event script or command
Notify sysadmin that event processing is complete
```

A Tailored Event

See Chapter 21, Monitoring an HACMP/ES Cluster, for a discussion of event emulation, which lets you emulate HACMP/ES event scripts without actually affecting the cluster.

Event Notification

You can specify a **notify** command that sends mail to indicate that an event is about to happen (or has just occurred), and that an event script succeeded or failed. For example, a site may want to use a **network_down** notification event to inform system administrators that traffic may have to be re-routed. Afterwards, you can use a **network_up** notification event to tell system administrators that traffic can again be serviced through the restored network.

Event notification in an HACMP/ES cluster can also be done using pre- and post-event scripts.

Pre- and Post-Event Scripts

You can specify commands or multiple user-defined scripts that execute before and after the Cluster Manager calls an event script.

For example, you can specify one or more pre-event scripts that run before the **node_down** event script is processed. When the Cluster Manager recognizes that a remote node is down, it first processes these user-defined scripts. One such script might designate that a message be sent to all users to indicate that performance may be affected (while adapters are swapped, while application servers are stopped and restarted). Following the **node_down** event script, a post processing event script for **network_up** notification might be included to broadcast a message to all users that a certain system is now available at another network address.

The following scenarios are other examples of where pre- and post-event processing are useful:

- If a **node_down** event occurs, site specific actions might dictate that a pre-event script for the **start_server** subevent script be used. This script could notify users on the server about takeover for the downed application server that performance may vary, or that they should seek alternate systems for certain applications they may be interested in.
- Due to a network being down, a custom installation might be able to re-route traffic through other machines by creating new IP routes. The **network_up** and **network_up_complete** event scripts could reverse the procedure, ensuring that the proper routes exist after all networks are functioning.
- A site may want to initiate a graceful takeover as a post-event if a network has failed on the local node (but otherwise the network is functioning).

If you plan to create pre- or post-event scripts for your cluster, be aware that your scripts will be passed the same parameters used by the HACMP/ES event script you specify. The first parameter passed will be the event name; other parameters passed are those used by the HACMP/ES event. For example, for the **network_down** event, the following parameters are passed to your script: the event name, **network_down**, followed by the name of the node and network experiencing the failure.

All HACMP/ES event scripts are maintained in the **/usr/es/sbin/cluster/events** directory. The parameters passed to your script are listed in the event script headers.

Warning: Be careful not to kill any HACMP processes as part of your script! If you are using the output of the **ps** command and using a **grep** to search for a certain pattern, make sure the pattern does not match any of the HACMP or RSCT processes.

Event Recovery and Retry

You can specify a command that attempts to recover from an event script failure. If the recovery command succeeds and the retry count for the event script is greater than zero, the event script is rerun. You can also specify the number of times to attempt to execute the recovery command.

For example, a recovery command could include the retry of unmounting a file system after logging a user off, and making sure no one was currently accessing the file system.

If a condition that affects the processing of a given event on a cluster is identified, such as a timing issue, you can insert a recovery command with a retry count high enough to be sure to cover for the problem.

User-defined Events

You can define your own events for which HACMP/ES can run your specified recovery programs. This adds a new dimension to the predefined HACMP/ES pre- and post-event script customization facility.

You specify the mapping between events you define and recovery programs defining the event recovery actions in the **rules** file. This lets you control both the scope of each recovery action and the number of event steps synchronized across all nodes. See *IBM PSSP for AIX Event Management Programming Guide and Reference* for details about registering events.

This facility requires that you use the resource monitors supplied with version 2.3 of the PSSP program:

- AIX resource monitor; variable names start with IBM.PSSP.aixos.
- Program resource monitor: variable names start with IBM.PSSP.Prog.

You cannot use the Event Emulator to emulate a user-defined event.

Changing the Rules File

HACMP/ES comes with a rules file `/usr/es/sbin/cluster/events/rules.hacmprd` to handle the standard HACMP/ES events. For further information on events, see *IBM PSSP Event Management Programming Guide and Reference*.



Warning: You can expand this file to include events you define. Do not, however, change any of the existing definitions.

You must put the updated rules file and all the specified recovery programs on all nodes in the cluster before starting **clstrmgr** on any node.

The rules in the file consist of the following fields, with each field a separate line in the file. If you don't need one of the fields, include a blank line in the file for it.

1. Name
2. State (qualifier)
3. Recovery program path
4. Recovery type (reserved for future use)
5. Recovery level (reserved for future use)
6. Resource variable name (for Event Manager events)
7. Instance vector (for Event Manager events)
8. Predicate (for Event Manager events)
9. Rearm predicate (for Event Manager events)

The event name and state pair are the rule trigger. The Cluster Manager runs a recovery program only if it finds a rule with a trigger corresponding to the event name and state.

Writing Recovery Programs

A recovery program has a sequence of recovery command specifications, possibly interspersed with barrier commands.

The format of these specifications is:

```
:node_set recovery_command expected_status NULL
```

Where:

- *node_set* is a set of nodes on which the recovery program is to run
- *recovery_command* is a quote-delimited string specifying a path to the executable program. The recovery program must be in this path on all nodes in the cluster. The program must specify an exit status.

- *expected_status* is an integer status to be returned when the recovery command completes successfully. The Cluster Manager compares the actual status returned to the expected status. A mismatch indicates unsuccessful recovery. If you specify the character X in the expected status field, the Cluster Manager omits the comparison.
- *NULL* is not used now, but is included for future functions.

You specify node sets by dynamic relationships. HACMP/ES supports the following dynamic relationships:

- *all*—The recovery command executes on all nodes in the current membership.
- *event*—The node on which the event occurred.
- *other*—All nodes except the one on which the event occurred.

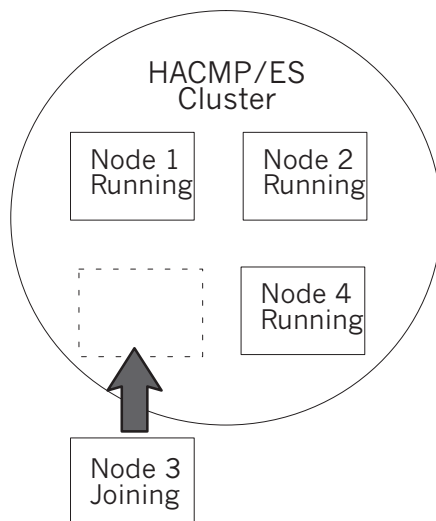
The specified dynamic relationship generates a set of recovery commands identical to the original, except that a node id replaces *node_set* in each set of commands.

The command string for user defined event commands must start with a “/”. The **clcallev** command runs commands that do not start with a “/”.

Example

Here is the recovery program for the *node_up* event:

```
#format:
#relationshipcommand to runexpected status NULL
#
other "node_up" 0 NULL
#
barrier
#
event "node_up" 0 NULL
#
barrier
#
all "node_up_complete" X NULL
#
```



node-up Recovery Program

Barrier Commands

You can put any number of barrier commands in the recovery program. All recovery commands before a barrier start in parallel. Once a node encounters a barrier command, all nodes must reach it before the recovery program continues.

The syntax of the barrier command is the string `barrier`.

Event Roll-up

If there are multiple events outstanding simultaneously, you only see the highest priority event. Node events are higher priority than network events. But user-defined events, the lowest priority, do not roll up at all, so you see all of them.

Completing the Cluster Event Worksheet

You now fill out a worksheet that helps you plan the cluster event processing for your cluster. Appendix A, Planning Worksheets, has blank copies of the worksheet referenced in the procedure below. Make photocopies of this worksheet to record all your cluster event scripts.

If you plan to use the online cluster planning worksheets, refer to Appendix B, Using the Online Cluster Planning Worksheet Program, for instructions.

Completing the Cluster Event Worksheet

Complete the following steps to plan the customized processing for a specific cluster event. Enter a value in the fields only as necessary.

For each node in the cluster, repeat these steps on a separate worksheet.

1. Record the cluster ID in the **Cluster ID** field.
2. Record the cluster name in the **Cluster Name** field.
3. Record the custom cluster event description in the **Cluster Event Description** field.
4. Record the full pathname of the cluster event method in the **Cluster Event Method** field.
5. Fill in the name of the cluster event in the **Cluster Event Name** field.
6. Fill in the full pathname of the event script in the **Event Command** field.
7. Record the full pathname of the event notification script in the **Notify Command** field.
8. Record the name of the pre-event script in the **Pre-Event Command** field.
9. Record the name of the post-event script in the **Post-Event Command** field.
10. Record the full pathname of the event retry script in the **Event Recovery Command** field.
11. Indicate the number of times to retry in the **Recovery Command Retry** field.

Repeat steps 3 through 11 for each event you plan to customize.

Where You Go From Here

You have now planned the customized and user-defined event processing for your cluster. You next address issues relating to cluster clients, described in Chapter 9, Planning for HACMP/ES Clients.

Cluster Events: Tailoring and Creating
Where You Go From Here

Chapter 9 Planning for HACMP/ES Clients

This chapter discusses planning considerations for HACMP/ES clients. This is the last step before proceeding to installation of your HACMP/ES software.

Prerequisites

By now, you should have completed these steps in planning for your cluster:

- Applications (Chapter 3, Initial Cluster Planning)
- Networks (Chapter 4, Planning Cluster Network Connectivity)
- Disk configuration (Chapter 5, Planning Shared Disk Devices)
- Shared volume groups (Chapter 6, Planning Shared LVM Components)
- Resource groups (Chapter 7, Planning Resource Groups)
- Cluster events (Chapter 8, Cluster Events: Tailoring and Creating)

Overview

Clients are end-user devices that can access the nodes in an HACMP/ES cluster. In this step of the planning process, you evaluate the cluster from the point of view of the clients. You need to consider the following:

- Different types of clients: Computers and terminal servers
- Clients with and without the Client Information Program (Clinfo)
- Network components: Routers, bridges, gateways.

Different Types of Clients: Computers and Terminal Servers

Clients may include a variety of hardware and software from different vendors. In order to maintain connectivity to the HACMP/ES cluster, you must consider the following issues.

Client Application Systems

All clients should run Clinfo if possible. If you have hardware other than RS/6000 nodes in the configuration, you may want to port Clinfo to those platforms. Clinfo source code is provided as part of the HACMP/ES release.

You need to think about what applications are running on these clients. Who are the users? Is it required or appropriate for users to receive a message when cluster events affect their system?

NFS Servers

Issues related to NFS are discussed in Chapter 28, Additional Tasks: NFS and Run-Time Parameters.

Terminal Servers

If you plan to use a terminal server on the local area network, consider the following when choosing the hardware:

- Can you update the terminal server's ARP cache? The terminal server must comply with the TCP/IP protocol, including telnet.
- Is the terminal server programmable, or does it need manual intervention when a cluster event happens?
- Can you download a control file from the cluster node to the terminal server that updates or handles cluster events' effects on clients?

If your terminal server does not meet these operating requirements, you should choose the hardware address swapping option when configuring the cluster environment.

Clients Running Clinfo

The Clinfo program calls the `/usr/es/sbin/cluster/etc/clinfo.rc` script whenever a network or node event occurs. By default, this action updates the system's ARP cache to reflect changes to network addresses. You can customize this script if further action is desired.

Reconnecting to the Cluster

Clients running the Clinfo daemon will be able to reconnect to the cluster quickly after a cluster event. If you have hardware other than RS/6000s between the cluster and the clients, you must make sure that you can update the ARP cache of those network components after a cluster event occurs.

If you configure the cluster to swap hardware addresses as well as IP addresses, you do not need to be concerned about updating the ARP cache. You simply must be aware that this option causes a longer delay for the users.

Tailoring the `clinfo.rc` Script

For clients running Clinfo, you need to decide whether to tailor the `/usr/es/sbin/cluster/etc/clinfo.rc` script to do more than update the ARP cache when a cluster event occurs. See Chapter 16, Configuring Clinfo Scripts and Files, for more information. See also Chapter 4, Sample Clinfo Client Program, in the *HACMP for AIX Programming Client Applications Guide* for a sample tailored `clinfo.rc` script.

Clients Not Running Clinfo

On clients not running Clinfo, you may have to update the local ARP cache indirectly by pinging the client from the cluster node. On the cluster nodes, add to the PING_CLIENT_LIST variable in the **clinfo.rc** script the name or address of a client host you want to notify. Now, whenever a cluster event occurs, **clinfo.rc** executes the following command for each host specified in PING_CLIENT_LIST:

```
/etc/ping hostname 1024 1
```

This assumes the client is directly connected to one of the cluster networks.

Network Components

If you have configured the network so that clients attach to networks on the other side of a router, bridge, or gateway rather than to the cluster's local networks, you must be sure that you can update the ARP cache of those network components after a cluster event occurs. If this is not possible, then make sure to use hardware address swapping when you configure the cluster environment.

Where You Go From Here

This chapter concludes the planning process. You can now begin to install the HACMP/ES software. See Part 2 of this Guide for detailed instructions on installing the software.

Planning for HACMP/ES Clients
Where You Go From Here

Part 2

Installing and Configuring HACMP/ES Clusters

This part contains instructions for installing hardware and software and then configuring the cluster.

Chapter 10, Overview: Installing and Configuring HACMP/ES

Chapter 11, Checking Installed Hardware

Chapter 12, Defining Shared LVM Components

Chapter 13, Tailoring AIX for HACMP/ES

Chapter 14, Installing the HACMP/ES Software

Chapter 15, Upgrading an HACMP/ES Cluster

Chapter 16, Configuring Clinfo Scripts and Files

Chapter 17, Installing and Configuring Clients

Chapter 18, Configuring an HACMP/ES Cluster

Chapter 10 Overview: Installing and Configuring HACMP/ES

This chapter describes the steps you take to install and configure HACMP/ES software and lists hardware supported by HACMP/ES.

Prerequisites

Read Part 1, Planning an HACMP/ES Cluster, before installing the HACMP/ES software at your site. In this part, you plan the following components of your cluster:

- Applications (Chapter 3, Initial Cluster Planning)
- Networks (Chapter 4, Planning Cluster Network Connectivity)
- Disk configuration (Chapter 5, Planning Shared Disk Devices)
- Shared volume groups (Chapter 6, Planning Shared LVM Components)
- Resource groups (Chapter 7, Planning Resource Groups)
- Cluster events (Chapter 8, Cluster Events: Tailoring and Creating)
- Clients (Chapter 9, Planning for HACMP/ES Clients)

Refer to the worksheets and diagrams completed in Part 1 as you install and configure the cluster. If you have not completed these worksheets and diagrams, return to Part 1 and do so before continuing.

Note: The directory `/usr/sbin/cluster` and subdirectories have symbolic links to the `/usr/es/sbin/cluster` directory and subdirectories. Files in those directories are *not* linked, as they were in releases prior to 4.3.1.

Note: If you are running HACMP for AIX 4.4 on your cluster, you can migrate to HACMP/ES 4.4 in a node-by-node process without bringing the cluster down. See Chapter 14, Installing the HACMP/ES Software for instructions on node-by-node migration.

Overview

The following chapters describe how to use the standard installation and configuration procedure using the SMIT interface. The standard procedure gives you the most flexibility in configuring the cluster to your specific requirements. You can use SMIT to:

- Install or upgrade the operating system to AIX 4.3.2.
- Install the HACMP/ES software
- Configure clusters, nodes, resources, and events
- Verify the cluster configuration.

Using the `xhacmpm` Utility to Configure HACMP/ES

For clusters of up to eight nodes, you can use the sample program VSM (Visual System Management), or `xhacmpm` after installing the HACMP/ES software. It is an X Windows tool for creating cluster configurations using icons that represent cluster resources. See Appendix I, VSM Graphical Configuration Application, in Volume 2 of this guide, for more information on using the `xhacmpm` facility to configure a cluster.

Steps for Installing and Configuring an HACMP/ES Server

This section identifies the steps required to set up, install, and configure an HACMP/ES server. Each of these steps is detailed in another chapter.

Check Installed Hardware

In this step you ensure that network adapters and shared external disk devices are ready to support an HACMP/ES cluster as described in Chapter 11, Checking Installed Hardware.

Define Shared LVM Components

In this step you create the shared volume groups, logical volumes, and filesystems for your cluster as described in Chapter 12, Defining Shared LVM Components.

Tailor AIX for HACMP/ES

In this step you review or edit various AIX files to ensure a proper configuration for network options and for various host files as described in Chapter 13, Tailoring AIX for HACMP/ES.

Install HACMP/ES Software

In this step you install and verify the HACMP/ES software on each cluster node as described in Chapter 14, Installing the HACMP/ES Software.

Update HACMP/ES Software

If you have been using an earlier version of the HACMP/ES software, in this step you upgrade it, as described in Chapter 15, Upgrading an HACMP/ES Cluster.

Set up the Cluster Information Program

In this step you edit the `/usr/es/sbin/cluster/etc/clhosts` file and the `/usr/es/sbin/cluster/etc/clinfo.rc` script as described in Chapter 16, Configuring Clinfo Scripts and Files.

Configure the HACMP/ES Software

In this step you define the components of your HACMP/ES cluster as described in Chapter 18, Configuring an HACMP/ES Cluster.

Steps for Installing and Configuring an HACMP/ES Client

This section identifies the steps required to set up, install, and configure an HACMP/ES client. These steps are described in Chapter 17, *Installing and Configuring Clients*.

The section also discusses how to set up HACMP/ES client code on the control workstation if the workstation is used as a client to monitor cluster status.

Install the Base System on Clients

In this step you install the base high availability software on a client.

Edit the `/usr/es/sbin/cluster/etc/clhosts` File

In this step you edit the `/usr/es/sbin/cluster/etc/clhosts` file to provide the HACMP/ES server addresses needed for clients to communicate with cluster nodes.

Edit the `/usr/es/sbin/cluster/etc/clinfo.rc` Script

In this step you review the importance of editing the `/usr/es/sbin/cluster/etc/clinfo.rc` script to ensure that the ARP cache is updated as a result of a cluster event.

Update Non-Clinfo Clients

In this step you update the ARP cache on non-Clinfo clients.

Reboot the Clients

In this step you reboot each cluster client. See Chapter 20, *Starting and Stopping Cluster Services* for more information.

Specified Operating Environment

This section describes the required and supported hardware for HACMP, as of version 4.3.1.

Hardware Requirements

HACMP 4.3.1 works with RS/6000 uniprocessors, SMP servers, and SP systems in a “no-single-point-of-failure” server configuration. HACMP 4.3.1 supports the RS/6000 models designed for server applications and meet the minimum requirements for internal memory, internal disk, and I/O slots. The following RS/6000 models and their corresponding upgrades are supported in HACMP 4.3.1:

- PCI Desktop Systems, Models 140, 150, 240, and 260
- PCI Deskside Systems, Models E20, E30, F30, F40, and F50
- PCI Rack Systems, Models H10, S70, S7A, and S80
- Entry Systems, Models 25S, 250, and 25T
- Compact Server Systems, Models C10 and C20
- Desktop Systems, Models 370, 380, 390, 397, and 39H

Overview: Installing and Configuring HACMP/ES Specified Operating Environment

- Deskside Systems, Models 570, 57F, 580, 58F, 58H, 590, 59H, 591, and 595
- Rack Systems, Models 98B, 98E, 98F, 990, 99E, 99F, 99J, 99K, R10, R20, R21, R24, R50, R5U, S70, S7A, H50, and H70
- Symmetric Multiprocessor Server Systems, Models G30, J30, R30, R3U, G40, J40, R40, R4U, J50, R4U, S70, and S7A
- SP Systems, Models 204, 205, 206, 207, 208, 209, 20A, 2A4, 2A5, 2A7, 2A8, 2A9, 2AA, 304, 305, 306, 307, 308, 309, 30A, 3A4, 3A5, 3A7, 3A8, 3A9, 3AA, 3B4, 3B5, 3B7, 3B8, 3B9, 3BA, 404, 405, 406, 407, 408, 409, 40A, 500, 50H, 550, and 55H, including the 604 High Nodes, 604E High Nodes, and the Power2 SuperChip (P2SC) nodes

Any supported RS/6000 can be joined with any other supported RS/6000 in an HACMP 4.3.1 configuration. The Models 250 and 25T can be used in the HACMP 4.3.1 server configuration, but due to slot limitations, a “single-point-of-failure” is unavoidable in shared-disk or shared-network resources.

HACMP 4.3.1 executing in a concurrent access configuration requires one of the following devices:

- IBM 7131 SSA Multi-Storage Tower Model 405 (supports up to eight nodes; no CD-ROMs or tapes can be installed)
- IBM 7133 SSA Disk Subsystem Models 020, 600, D40 and T40 (supports up to eight nodes)
- IBM 7135 RAIDiant Array Models 110 and 210 (supports up to four nodes; dual controllers recommended)
- IBM 7137 Disk Array Subsystem Models 413, 414, 415, 513, 514, or 515 (supports up to four nodes)
- IBM 2105 Versatile Storage Server (VSS) Models B09 and 100 (supports up to four nodes)

Certain non-IBM RAID systems can operate in concurrent I/O access environments. IBM will not accept Authorized Program Analysis Reports (APARs) if the non-IBM RAID offerings do not work properly with HACMP 4.3.1.

The minimum configuration and sizing of each machine is highly dependent on the user's database package and other applications.

Actual configuration requirements are highly localized according to the required function and performance needs of individual sites. In configuring a cluster, particular attention must be paid to:

- Fixed-disk capacity and mirroring (Logical Volume Manager (LVM) and database)
- Slot limitations and their effect on creating a single-point-of-failure
- Client access to the cluster
- Other LAN devices (routers, bridges) and their effect on the cluster
- Replication of I/O adapters/subsystems
- Replication of power supplies
- Other network software

Whenever a process takes over resources after a failure, consideration must be given to work partitioning. For example, if processor “A” is expected to take over for failed processor “B” and continue to perform its original duties, “A” must be configured with enough resources to perform the work of both.

HACMP 4.3.1 Device Support

At this time, the following adapters are supported in the HACMP 4.3.1 environment. Refer to individual hardware announcements for the levels of AIX that are supported.

Communications Adapters

- PCI/ISA
- 2920 IBM PCI Token-Ring Adapter
- 2931 ISA 8-Port Asynchronous Adapter
- 2932 ISA 8-Port Asynchronous Adapter
- 2933 ISA 128-Port Asynchronous Controller
- 2741 PCI FDDI-Fiber Single-Ring Upgrade
- 2742 PCI FDDI-Fiber Dual-Ring Upgrade
- 2743 PCI FDDI-Fiber Single-Ring Upgrade
- 2944 128-Port Asynchronous Controller, PCI bus
- 2943 8-Port Asynchronous EIA-232/RS-422, PCI bus Adapter
- 2963 Turboways 155 PCI UPT ATM Adapter
- 2968 PCI Ethernet 10/100 Adapter
- 2969 PCI Gigabit Ethernet Adapter
- 2979 PCI AutoLANStreamer Token-Ring Adapter
- 2985 PCI Ethernet BNC/RJ-45 Adapter
- 2986 PCI Ethernet 10/100 Adapter
- 2987 PCI Ethernet AUI/RJ-45 Adapter
- 2988 Turboways 155 PCI MMF ATM Adapter
- 4959 Token Ring PCI Adapter
- 8396 RS/6000 SP System Attachment Adapter

ATM Hardware Address Takeover is limited to adapters connected to the same switch.

MCA

- 1904 Fibre Channel Adapter
- 2402 Network Terminal Accelerator Adapter
- 2403 Network Terminal Accelerator Adapter
- 2723 FDDI-Fiber Dual-Ring Upgrade
- 2724 FDDI-Fiber Single-Ring Adapter
- 2725 FDDI-STP Single-Ring Adapter
- 2726 FDDI-STP Dual-Ring Upgrade

Overview: Installing and Configuring HACMP/ES Specified Operating Environment

- 2930 8-Port Async Adapter - EIA-232
- 2964 10/100 Mbps Ethernet Adapter - UNI
- 2972 AutoLANStreamer Token-Ring Adapter
- 2980 Ethernet High-Performance LAN Adapter
- 2989 Turboways 155 ATM Adapter
- 2992 Ethernet/FDX 10 Mbps TP/AUI MC Adapter
- 2993 Ethernet BNC MC Adapter
- 2994 10/100 Mbps Ethernet Adapter - SMP
- 4018 High-Performance Switch (HPS) Adapter-2
- 4020 Scalable POWERParallel Switch Adapter

Disk Adapters

PCI

- 6205 PCI Dual Channel Ultra2 SCSI Adapter
- 6206 PCI SCSI-2 Single-Ended Ultra-SCSI Adapter
- 6207 PCI SCSI-2 Differential Ultra-SCSI Adapter
- 6208 PCI SCSI-2 Single-Ended Fast/Wide Adapter
- 6209 PCI SCSI-2 Differential Fast/Wide Adapter
- 6215 PCI SSA Adapter
- 6225 Advanced SerialRAID Adapter

MCA

2412 Enhanced SCSI-2 Differential Fast/Wide Adapter/A 2415 SCSI-2 Fast/Wide Adapter/A
2416 SCSI-2 Differential Fast/Wide Adapter/A 2420 SCSI-2 Differential High-Performance
External I/O Controller 6212 High Performance Subsystem Adapter/A (40/80 Mbps)

- 6214 SSA 4-Port Adapter
- 6216 Enhanced SSA 4-Port Adapter
- 6219 MCA SSA Adapter

For compatibility with subsystems not listed below, refer to the individual hardware announcements.

External Storage Subsystems

- IBM 2105 Versatile Storage Server (VSS) Models B09 and 100 (supports up to four nodes)
- IBM 7131 SCSI Multi-Storage Tower Model 105 (supports up to four nodes; no CD-ROMs or tapes can be installed)
- IBM 7131 SSA Multi-Storage Tower Model 405 (supports up to eight nodes; no CD-ROMs or tapes can be installed)
- IBM 7133 SSA Disk Subsystem Models 020, 600, D40 and T40 (supports up to eight nodes)
- IBM 7135 RAIDiant Array Models 110 and 210 (supports up to four nodes; dual controllers recommended)

- IBM 7137 Disk Array Subsystem Models 413, 414, 415, 513, 514, and 515 (supports up to four nodes)
- IBM 7204 External Disk Drive Models 317, 325, 339, 402, 404, and 418 (supports up to four nodes)
- IBM 2105 Versatile Storage Server (VSS) Models B09 and 100 and Enterprise Storage Server (ESS) Models E10 and E20 (supports up to four nodes)

Router Support

The IBM RS/6000 SP Switch Router 9077-04S can be used in cluster configurations where the router is used to provide communications to client systems.

The router is not supported in the communications path between nodes in an HACMP cluster.

Rack-Mounted Storage Subsystems

IBM 7027 High Capacity Storage Drawer Model HSC (supports up to two nodes; no CD-ROMS or tapes installed)

IBM 7027 High Capacity Storage Drawer Model HSD (supports up to four nodes; no CD-ROMS or tapes installed)

Overview: Installing and Configuring HACMP/ES
Specified Operating Environment

Chapter 11 Checking Installed Hardware

This chapter describes how to verify that installed hardware is ready to support an HACMP/ES cluster.

Overview

The chapter presumes that you have already installed the devices following the instructions in the relevant AIX documentation. This chapter describes how to verify that network adapters and shared external disk devices are ready to support an HACMP/ES cluster.

Be sure to connect shared SCSI disks to the same SCSI bus as the nodes sharing the disk.

Checking Network Adapters

This section describes how to ensure that network adapters are configured properly to support the HACMP/ES software. For each node, check the settings of each adapter to make sure they match the values on the completed copies of the *TCP/IP Network Adapter Worksheet*. Special considerations for a particular adapter type are discussed below.

Ethernet, Token-Ring, and FDDI Adapters

- When using the **smit mktcpip** fastpath to define an adapter, the HOSTNAME field changes the default hostname. For instance, if you configure the first adapter as *n0_svc*, and then configure the second adapter as *n0_sby*, the default hostname at system boot is *n0_sby*.
To avoid this problem, it is recommended that you configure the adapter with the desired default hostname last. It is recommended that the hostname match the adapter label of the primary network's service adapter because some applications may depend on the hostname (though this is not required for HACMP/ES).
- Use the **smit chinnet** or **smit chghfcs** fastpath to configure each adapter for which IP address takeover might occur to boot from the boot adapter address and not from its service adapter address. Refer to your completed copies of the *TCP/IP Network Adapter Worksheet*.

Completing the TCP/IP Network Adapter Worksheets

After checking all network interfaces for a node, record the network interface names on that node's *TCP/IP Network Adapter Worksheet*. Enter the following command:

```
lsdev -Cc if
```

This displays a list of available and defined adapters for the node. At this point, all interfaces used by the HACMP/ES software should be available. List the adapters marked "Available" in the **Interface Name** field on the *TCP/IP Network Adapter Worksheet*.

Serial Networks

It is strongly recommended that a serial network connect the nodes in an HACMP/ES cluster. The serial network allows the Cluster Managers to continue to exchange keepalive packets should the TCP/IP-based subsystem, networks, or network adapters fail. Thus, the serial network prevents the nodes from becoming isolated and attempting to take over shared resources. The HACMP/ES software supports three types of serial networks: RS232 lines, the SCSI-2 Differential bus using target mode SCSI, and target mode SSA connections.

See Chapter 4, Planning Cluster Network Connectivity, for more information on serial networks.

Checking an SP Switch

The SP Switch used by an SP node serves as a network device for configuring multiple clusters, and it can also connect clients. This switch is not required for an HACMP/ES installation. When installed, the SP Switch default settings are sufficient to allow it to operate effectively in an HACMP/ES cluster environment.

You must be aware that the HPS switch (older version of the switch) differs from the SP switch (newer version of the switch): **The SP switch does not allow HACMP/ES to control the Eprimary.** You must upgrade to the SP switch before installing HACMP/ES 4.4. If you are currently running HACMP/ES Eprimary management with an HPS switch, you should run the HACMP/ES script to unmanage the Eprimary BEFORE upgrading the switch.

To check whether the Eprimary is set to be managed:

```
odmget -q'name=EPRIMARY' HACMPsp2
```

If the switch is set to MANAGE, before changing to the new switch, run the script:

```
/usr/es/sbin/cluster/events/utills/cl_HPS_Eprimary unmanage
```

Basic points to remember about the SP Switch in an HACMP/ES configuration:

- ARP must be enabled for the SP Switch network so that IP address takeover can work.
- All SP Switch addresses must be defined on a private network.
- HACMP/ES SP Switch boot and service addresses are alias addresses on the SP Switch css0 IP interface. The css0 base IP address is unused and should not be configured for IP address takeover. Standby adapters are not required for SP Switch IP address takeover. The alias service addresses appear as **ifconfig alias** addresses on the css0 interface.
- The netmask associated with the css0 base IP address is used as the netmask for all HACMP/ES SP Switch network adapters.

See Chapter 4, Planning Cluster Network Connectivity, for more information on SP switch networks.

Configuring for Asynchronous Transfer Mode (ATM)

Asynchronous Transfer Mode (ATM) is a connection-oriented, private network. A point-to-point connection between two systems makes an ATM network similar to Frame Relay, X.25, and SLIP connections rather than to non-connection methods such as Ethernet, Token Ring, and FDDI.

Because ATM is a connection-oriented technology and IP is a datagram-oriented technology, mapping IP addresses over ATM is complex. For PVC-type ATM connections, each IP station must be manually configured. SVC-type ATM networks are dynamic and are divided into logical IP subnets (LIS). An LIS is similar to a traditional LAN segment. The ATM standard requires one ARP server per LIS, where each ARP server resolves IP addresses into the corresponding ATM address without using broadcasting. Each IP station must be configured with the ATM addresses of the ARP servers that serve the IP subnet configured on that IP station. See the following section, “Configuring ATM Networks,” for details about configuring both ATM ARP servers and clients.

The current ATM support in HACMP/ES reflects the following restrictions:

HACMP/ES Configuration Restrictions

ATM networks must be defined to HACMP/ES as private networks because ATM does not support broadcasting.

ATM Configuration Restrictions

There are two basic components to configuring ATM networks:

1. Configuring the ATM ARP servers
2. Configuring the ATM ARP clients.

ATM networks require the use of another system as an ARP server to handle ATM hardware address to IP address resolution. One ARP server exists for each defined subnetwork. Thus, two ARP servers are required for each ATM network configured in an HACMP/ES cluster – one for the service subnet and one for the standby subnet. The ATM ARP client must be configured to include the following information about the ATM ARP server:

- The IP address of the ATM ARP server
- The MAC address of the ATM ARP server
- The Selector byte that reflects the TCP/IP interface number of the ATM interface being used as the ATM ARP server.

An ATM ARP server can be an HACMP/ES client, but it cannot be an HACMP/ES server because ATM ARP clients hard code the ATM link address of the ARP server (including the MAC address). Adapter swaps and IPAT, therefore, do not work on HACMP/ES nodes that are also ATM ARP servers.

ATM can support multiple interfaces per ATM adapter. Adapters used in HACMP/ES servers, however, must use only one interface per adapter, and interface numbers must match the adapter numbers. ATM switches can be used as ATM ARP servers only if the switch automatically updates its ARP cache after an HACMP/ES adapter event.

Checking Installed Hardware

Configuring for Asynchronous Transfer Mode (ATM)

Thus, ATM networks must be configured as Switched Virtual Circuits through an ATM switch, with the ATM ARP server configured on a system that is not a member of the cluster nor on the ATM switch itself (except as noted in the previous paragraph).

Configuring ATM ARP Servers for Use by HACMP/ES Nodes

Before configuring an ATM ARP server, install the ATM adapters and the switch as described in your ATM product documentation. When installation is complete, do the following:

1. Configure an ATM ARP server for the HACMP/ES service subnet.
2. Configure an ATM ARP server for the HACMP/ES standby subnet.
3. Determine the ATM server address for each ATM server.

Configuring ATM ARP Servers for HACMP/ES Service Subnetworks

To configure an ARP server for the HACMP/ES “service” subnetwork:

1. Enter:

```
smitty chinnet
```

SMIT displays a list of Available Network Interfaces.
2. Select **at0** as the ATM network interface. This interface will serve as the ARP server for the subnetwork 192.168.110 as shown in the following example of the Change/Show an ATM Interface screen.

Network Interface Name	at0
INTERNET ADDRESS (dotted decimal)	192.168.110.28
Network MASK (hex or dotted decimal)	255.255.255.0
Connection Type	svc_s
ATM Server Address	
Alternate Device	
Idle Timer	60
Current STATE	up

Note: The **Connection Type** field is set to `svc_s` to indicate that the interface is used as an ARP server.

3. Press F10 to exit SMIT.

Configuring ATM ARP Servers for HACMP/ES Standby Subnetworks

To configure an ARP server for an HACMP/ES “standby” subnetwork:

1. Repeat Steps 1 through 3 of the preceding procedure to configure an ATM ARP server for a “standby” subnetwork, selecting **at1** as the network interface and other options as shown in the following example:

Network Interface Name	at1
INTERNET ADDRESS (dotted decimal)	192.168.111.28
Network MASK (hex or dotted decimal)	255.255.255.0
Connection Type	svc_s
ATM Server Address	
Alternate Device	atm0
Idle Timer	60
Current STATE	up

Note: The interface name is (at1) for the standby adapter; the **Connection Type** designates the interface as an ARP server, svc_s. The **Alternate Device** field is set to atm0. This setting puts at1 on atm0 with at0. The “standby” subnet is 192.168.111.

Obtaining an ARP Server's Hardware Addresses:

To show an ARP server's hardware addresses, use the **arp** command as follows on the ATM ARP server:

```
arp -t atm -a svc
```

A display similar to the following appears:

SVC IP Addr	ATM address
arpserver1	(192.168.110.28) 47.0.5.80.ff.e1.0.0.0.f2.1a.21.e.8.0.5a.99.82.95.0
aprserver2	(192.168.111.28) 47.0.5.80.ff.e1.0.0.0.f2.1a.21.e.8.0.5a.99.82.95.1

Note: The ATM arp server address is the 20-byte hardware address of the ATM arp server used for the subnet of an internet address.

ARP server addresses need to be defined to ATM ARP clients, as explained in the following section.

Configuring ATM ARP Clients on HACMP/ES Cluster Nodes

To configure ATM ARP clients on cluster nodes:

1. On each cluster node, configure the service and standby ATM adapters in AIX to use the “service” and “standby” ATM ARP servers previously configured.
2. Test the configuration.
3. Define the ATM network to HACMP/ES.

Configuring the HACMP/ES Cluster Nodes as ATM ARP Clients

Use the **smitty chinet** command to configure two ATM interfaces, one on each adapter (at0 on atm0 for “service”, and at1 on atm1 for “standby”).

Configuring the “service” subnet

Indicate the following values for these interfaces:

Network Interface Name	at0
INTERNET ADDRESS (dotted decimal)	192.168.110.30
Network MASK (hex or dotted decimal)	255.255.255.0
Connection Type	svc_c
ATM Server Address	47.0.5.80.ff.el.0.0.0.f2.1a.21.e.8.0.5a.99.82.95.0
Alternate Device	
Idle Timer	60
Current STATE	up

The **Connection Type** field is set to `svc_c` to indicate that the interface is used as an ATM ARP client. Because this ATM ARP client configuration is being used for the HACMP/ES “service” subnet, the **INTERNET ADDRESS** must be a host on the 192.168.110 subnet. The ATM server address is the 20-byte address which identifies the ATM ARP server being used for the 192.168.110 subnet.

Note: If IPAT is enabled for the HACMP/ES-managed ATM network, the **INTERNET ADDRESS** represents the “boot” address. If IPAT is not enabled, the **INTERNET ADDRESS** represents the “service” address.

Configuring the “standby” subnet

Indicate the following values for these interfaces:

Network Interface Name	at1
INTERNET ADDRESS (dotted decimal)	192.168.111.30
Network MASK (hex or dotted decimal)	255.255.255.0
Connection Type	svc_c
ATM Server Address	47.0.5.80.ff.e1.0.0.0.f2.1a.21.e.8.0.5a.99.82.95.1
Alternate Device	
Idle Timer	60
Current STATE	up

The **Connection Type** field is set to `svc_c` to indicate that the interface is used as an ATM ARP client. Because this ATM ARP client configuration is being used for the HACMP/ES “standby” subnet, the **INTERNET ADDRESS** must be a host on the 192.168.111 subnet. The ATM server address is the 20-byte address which identifies the ATM ARP server being used for the 192.168.111 subnet.

Testing Communication Over the Network

To test communication over the network after configuring ARP servers and clients:

1. Run the **netstat -i** command to make sure the ATM network is recognized. You should see the device listed as **at1**.
2. Enter the following command on the first node:

```
ping IP_address_of_other_node
```

where *IP_address_of_other_node* is the address in dotted decimal that you configured as the destination address for the other node.
3. Repeat Steps 1 and 2 on the second node, entering the destination address of the first node as follows:

```
ping IP_address_of_other_node
```

Defining the ATM Network to HACMP/ES

After you have installed and tested an ATM network, you must define it to the HACMP/ES cluster topology as a “private” network. Chapter 18, *Configuring an HACMP/ES Cluster*, describes how to define an ATM network in an HACMP/ES cluster.

ATM LAN Emulation

ATM LAN emulation provides an emulation layer between protocols such as Token-Ring or Ethernet and ATM. It allows these protocol stacks to run over ATM as if it were a LAN. You can use ATM LAN emulation to bridge existing Ethernet or Token-Ring networks—particularly switched, high-speed Ethernet—across an ATM backbone network.

LAN emulation servers reside in the ATM switch. Configuring the switch varies with the hardware being used. Once you have configured your ATM switch and a working ATM network, you can configure adapters for ATM LAN emulation.

Note: You must load **bos.atm** from AIX on each machine if you have not already done so.

To configure ATM LAN emulation through SMIT, take the following steps:

1. Enter the SMIT fastpath `atmle_panel`.
SMIT displays the ATM LAN Emulation menu.
2. Select **Add an ATM LE Client**.
3. Choose one of the adapter types (Ethernet or Token-Ring). A popup appears with the adapter selected (Ethernet in this example). Press Enter.
4. SMIT displays the **Add an Ethernet ATM LE Client** screen. Make entries as follows:

Local LE Client's LAN MAC Address (dotted hex)	Assign a hardware address like the burned in address on actual network cards. Address must be unique on the network to which it is connected.
Automatic Configuration via LECS	No is the default. Toggle if you want yes .
If no, enter the LES ATM address (dotted hex)	Enter the 20-byte ATM address of the LAN Emulation server.
If yes, enter the LECS ATM address (dotted hex)	If the switch is configured for LAN Emulation Configuration Server either on the well-known address, or on the address configured on the switch, enter that address here.
Local ATM Device Name	Press F4 for a list of available adapters.
Emulated LAN Type	Ethernet/IEEE 802.3 (for this example)
Maximum Frame Size (bytes)	
Emulated LAN name	(optional) Enter a name for this virtual network.

5. Once you make these entries, press Enter. Repeat these steps for other ATM LE clients.

6. The ATM LE Clients should be visible to AIX as network cards when you execute the `lsdev -Cc adapter` command.
7. Each virtual adapter has a corresponding interface that must be configured, just like a real adapter of the same type, and it should behave as such.

Defining the ATM LAN Emulation Network to HACMP/ES

After you have installed and tested an ATM LE network, you must define it to the HACMP/ES cluster topology as a public network. Chapter 18, Configuring an HACMP/ES Cluster, describes how to define networks and adapters in an HACMP/ES cluster.

You will define these virtual adapters to HACMP/ES just as if they were real adapters, except you cannot use Hardware Address swapping.

Checking Shared External Disk Devices

This section describes how to verify that shared external disk devices are configured properly for the HACMP/ES software. Separate procedures are shown for SCSI-2 Differential disk devices, IBM SCSI-2 Differential disk arrays, IBM 9333 serial disk subsystems, and IBM Serial Storage Architecture (SSA) disk subsystems.

Verifying Shared SCSI-2 Differential Disks

Complete the following steps to verify that a SCSI-2 Differential disk is installed correctly. These steps are valid for both SCSI-2 Differential and SCSI-2 Differential Fast/Wide disks. Differences in procedures are noted as necessary. As you verify the installation, record the shared disk configuration on the *Shared SCSI-2 Differential Disk Worksheet*. Use a separate worksheet for each set of shared SCSI-2 disks. You refer to the completed worksheets when you configure the cluster.

1. Note the type of SCSI bus installation, SCSI-2 Differential or SCSI-2 Differential Fast/Wide, in the **Type of SCSI Bus** section of the worksheet.
2. Fill in the node name of each node that connected to the shared SCSI bus in the **Node Name** field.
3. Enter the logical name of each adapter in the **Logical Name** field.

To determine the logical name, use the command:

```
lscfg | grep scsi
```

The first column lists the logical name of the SCSI adapters.

+ scsi0	00-07	SCSI I/O Controller
+ scsi1	00-08	SCSI I/O Controller

logical name

4. Record the Microchannel I/O slot that each SCSI adapter uses in the **Slot Number** field.

Checking Installed Hardware

Checking Shared External Disk Devices

The second column of the existing display lists the SCSI-2 Differential adapter's location code in the format AA-BB. The last digit of that value (the last B) is the Microchannel slot number.

+ scsi0	00-07	SCSI I/O Controller
+ scsi1	00-08	SCSI I/O Controller

slot

- Record the SCSI ID of each SCSI adapter on each node in the **Adapter** field. To determine the SCSI IDs of the disk adapters, use the **lsattr** command, as in the following example to find the ID of the adapter *scsi1*:

For SCSI-2 Differential adapters: `lsattr -E -l scsi1 | grep id`

For SCSI-2 Differential Fast/Wide adapters: `lsattr -E -l ascsi1 | grep external_id`

Do not use wildcard characters or full pathnames on the command line for the device name designation.

In the resulting display, the first column lists the attribute names. The integer to the right of the **id** (or **external_id**) attribute is the adapter SCSI ID.

id	7	Adapter card SCSI ID
----	---	----------------------

SCSI ID

Note: The IBM High Performance SCSI-2 Differential Fast/Wide Adapter cannot be assigned SCSI IDs 0, 1, or 2. The IBM SCSI-2 Differential Fast/Wide Adapter/A cannot be assigned SCSI IDs 0 or 1.

- Record the SCSI IDs of the physical disks in the **Shared Drive** fields. Use the command:

```
lsdev -Cc disk -H
```

The third column of the display generated by the **lsdev -Cc disk -H** command is a numeric location with each row in the format AA-BB-CC-DD. The first digit (the first D) of the DD field is the SCSI ID.

name	status	location	description
hdisk0	Available	00-07-00-00	2.0 GB SCSI Disk Drive
hdisk1	Available	00-07-00-10	2.0 GB SCSI Disk Drive
hdisk2	Available	00-07-00-20	2.0 GB SCSI Disk Drive

SCSI ID

Note: For SCSI-2 Differential Fast/Wide disks, a comma follows the SCSI ID field of the location code since the IDs can require two digits, such as 00-07-00-12,0.

7. At this point, verify that each SCSI device connected to the shared SCSI bus has a unique ID. A common configuration is to set the SCSI ID of the adapters on the nodes to be higher than the SCSI IDs of the shared devices. (Devices with higher IDs take precedence in SCSI bus contention.) For example, the adapter on one node can have SCSI ID 6 and the adapter on the other node can be SCSI ID 7, and the external disk SCSI IDs should be an integer from 0 through 5.
8. Determine the logical names of the physical disks.
The first column of the generated by the **lsdev -Cc disk -H** command lists the logical names of the SCSI disks.

name	status	location	description
hdisk0	Available	00-07-00-00	2.0 GB SCSI Disk Drive
hdisk1	Available	00-07-00-10	2.0 GB SCSI Disk Drive
hdisk2	Available	00-07-00-20	2.0 GB SCSI Disk Drive

logical name size

Record the name of each external SCSI disk in the **Logical Device Name** field, and the size in the **Size** field. *Be aware that the nodes can assign different names to the same physical disk. You should note these situations on the worksheet.*

9. Verify that all disks have a status of available. The second column of the display generated by the **lsdev -Cc disk -H** command shows the status.

Checking Installed Hardware

Checking Shared External Disk Devices

name	status	location	description
hdisk0	Available	00-07-00-00	2.0 GB SCSI Disk Drive
hdisk1	Available	00-07-00-10	2.0 GB SCSI Disk Drive
hdisk2	Available	00-07-00-20	2.0 GB SCSI Disk Drive

↑
status

If a disk has a status of defined, instead of available, check the cable connections and then use the **mkdev** command to make the disk available.

At this point, you have verified that the SCSI-2 Differential disk is configured properly for the HACMP/ES environment.

Verifying IBM SCSI-2 Differential Disk Arrays

Complete the following steps to verify that an IBM 7135 RAIDiant Disk Array or an IBM 7137 Disk Array is installed correctly. As you verify the installation, record the shared disk configuration on the *Shared SCSI-2 Differential Disk Array Worksheet*. Use a separate worksheet for each disk array. You will refer to the completed worksheets when you define the cluster topology.

1. Fill in the name of each node connected to this shared SCSI-2 Differential bus in the **Node Name** field.
2. Record the logical device name of each adapter in the **Adapter Logical Name** field.

To determine the logical device name, use the **lscfg** command, as in the following example:

```
lscfg | grep scsi
```

(The following examples display the information for an IBM 7135-110 RAIDiant Disk Array using SCSI-2 Differential Fast/Wide adapters.)

+ ascsi0	00-03	WIDE SCSI I/O Controller Adapter
+ vscsi1	00-03-00	SCSI I/O Controller Protocol Device
+ vscil	00-03-01	SCSI I/O Controller Protocol Device

↑
logical name

The first column of the display generated by the **lscfg** command lists the logical name of the SCSI adapters.

3. Record the Microchannel I/O slot of each SCSI-2 Differential adapter used in this shared SCSI bus in the **Slot Number** field of the configuration worksheet.

+ ascsi0	00-03	WIDE SCSI I/O Controller Adapter
+ vscsi1	00-03-00	SCSI I/O Controller Protocol Device
+ vscil	00-03-01	SCSI I/O Controller Protocol Device

slot

In the existing display, the second column lists the location code of the adapter in the format AA-BB. The last digit of that value (the last B) is the Microchannel slot number.

- Record the SCSI ID of each SCSI adapter on each node in the **Adapter** field. To determine the SCSI IDs of the disk adapters, use the **lsattr** command, specifying the logical name of the adapter as an argument. In the following example, the SCSI ID of the adapter named *ascsi0* is obtained:

```
lsattr -E -l ascsi0 | grep external_id
```

Do not use wildcard characters or full pathnames on the command line for the device name designation.

In the resulting display, the first column lists the attribute names. The integer to the right of the **id (external_id)** attribute is the adapter SCSI ID.

external_id	7	Adapter Card SCSI ID
-------------	---	----------------------

SCSI ID

Note: The IBM High Performance SCSI-2 Differential Fast/Wide Adapter cannot be assigned SCSI IDs 0, 1, or 2. The IBM SCSI-2 Differential Fast/Wide Adapter/A cannot be assigned SCSI IDs 0 or 1.

- Record the SCSI IDs of the array controllers in the **Array Controller** fields. To determine that AIX has the correct SCSI IDs for the array controllers, obtain a listing of the array controllers using the **lscfg** command, as in the following example:

```
lscfg | grep dac
```

In the display generated by the command, the first column lists the logical names of the array controllers. The second column contains the location code of the array controller in the format AA-BB-CC-DD. The seventh digit of the location code (the first D) is the SCSI ID of the array controller.

Checking Installed Hardware

Checking Shared External Disk Devices

+ dac0	00-02-00-20	7135 Disk Array Controller
+ dac1	00-04-00-20	7135 Disk Array Controller

SCSI ID

Note: Since these controllers are on separate SCSI buses, they can have the same SCSI ID.

The SCSI ID in the display should match the numbers you set for the SCSI ID on each array controller.

- At this point, determine that each device connected to a shared SCSI bus has a unique SCSI ID. A common configuration is to let one of the nodes keep the default SCSI ID 7 and assign the adapter on the other cluster node SCSI ID 6. The array controller SCSI IDs should later be set to an integer starting at 0 and going up. Make sure no array controller has the same SCSI ID as any adapter.
- Verify that AIX created the physical volumes (hdisks) that you expected.

To determine the logical names of the LUNs on the disk array, use the **lsdev** command, as in the following example:

```
lsdev -Cc disk -H
```

The example illustrates how this command lists the hard disks created for a four-LUN 7135-110 RAIDiant Disk Array. The display includes the location code of the hard disk in the form AA-BB-CC-DD. The last digit of the location code included in the display represents the LUN number. The first column lists the logical names of the LUNs on the 7135-110 RAIDiant Disk Array.

name	status	location	description
hdisk0	Available	00-02-00-30	7135 Disk Array Device
hdisk1	Available	00-02-00-31	7135 Disk Array Device
hdisk2	Available	00-02-00-32	7135 Disk Array Device
hdisk3	Available	00-02-00-33	7135 Disk Array Device

Logical name

Record the logical name of each LUN in the **Logical Device Name** field. *Be aware that the nodes can assign different names to the same physical disk. You should note these situations on the worksheet.*

- Verify that all disks have a status of available. The second column of the existing display shows the status.

name	status	location	description
hdisk0	Available	00-02-00-30	7135 Disk Array Device
hdisk1	Available	00-02-00-31	7135 Disk Array Device
hdisk2	Available	00-02-00-32	7135 Disk Array Device
hdisk3	Available	00-02-00-33	7135 Disk Array Device

status

If a disk has a status of defined, instead of available, check the cable connections and then use the **mkdev** command to make the disk available.

At this point, you have verified that the disk array is configured properly for the HACMP/ES software.

Configuring Target Mode SCSI Connections

This section describes how to configure a target mode SCSI connection between nodes sharing disks connected to a SCSI-2 Differential bus. Before you can configure a target mode SCSI connection, all nodes that share the disks must be connected to the SCSI bus, and all nodes and disks must be powered on.

Note: Neither PCI SCSI-2 differential busses nor SE busses support target mode SCSI.

Checking the Status of SCSI Adapters and Disks

To define a target mode SCSI connection, each SCSI adapter on nodes that share disks on the SCSI bus must have a unique ID and must be “Defined,” known to the system but not yet available. Additionally, all disks assigned to an adapter must also be “Defined” but not yet available.

Note: The uniqueness of adapter SCSI IDs ensures that tm SCSI devices created on a given node reflect the SCSI IDs of adapters on other nodes connected to the same bus.

To check the status of SCSI adapters you intend to use, enter:

```
lsdev -C | grep scsi
```

If an adapter is “Defined,” see Enabling Target Mode SCSI Devices in AIX on page 11-16 to configure the target mode connection.

To check the status of SCSI disks on the SCSI bus, enter:

```
lsdev -Cc disk
```

If either an adapter or disk is “Available,” follow the steps in the procedure below to return both the adapter (and its disks) to a defined state so that they can be configured for target mode SCSI and made available.

Returning Adapters and Disks to a Defined State

For a SCSI adapter, use the following command to make “Defined” each available disk associated with an adapter:

```
rmdev -l hdiskx
```

where *hdiskx* is the hdisk to be made “Defined.”

For example:

```
rmdev -l hdisk3
```

Next, run the following command to return the SCSI adapter to a “Defined” state:

```
rmdev -l scsix
```

where *scsix* is the adapter to be made “Defined.”

If using an array controller, you use the same command to return a router and a controller to a “Defined” state. However, make sure to perform these steps after changing the disk and before changing the adapter. The following lists these steps in this order:

```
rmdev -l hdiskx  
rmdev -l darx  
rmdev -l dacx  
rmdev -l scsix
```

When all controllers and disks are “Defined,” see “Enabling Target Mode SCSI Devices in AIX” to enable the Target Mode connection.

Note: Target mode SCSI is automatically configured if you are using the SCSI-2 Differential Fast/Wide Adapter. Skip ahead to the section on Follow-Up Task on page 11-18.

Enabling Target Mode SCSI Devices in AIX

To define a target mode SCSI device:

1. Enable the target mode interface for the SCSI adapter.
2. Configure (make available) the devices.

Complete both steps on one node, then on the second node.

Enabling Target Mode Interface

To enable the target mode interface:

1. Enter:

```
smit devices
```

SMIT displays a list of devices.
2. Select **SCSI Adapter** and press Enter.
3. Select **Change/Show Characteristics of a SCSI Adapter** and press Enter. SMIT prompts you to identify the SCSI adapter.

4. Set the **Enable TARGET MODE interface** field to **yes** to enable the target mode interface on the device (the default value is no).
At this point, a target mode SCSI device is generated that points to the other cluster nodes that share the SCSI bus. Note, however, that the SCSI ID of the adapter on the node from which you enabled the interface will not be listed.
5. Press Enter to commit the value.
6. Press F10 to exit SMIT.

Configuring the Target Mode SCSI Device

After enabling the target mode interface, you must run **cfgmgr** to create the initiator and target devices and make them available. To configure the devices and make them available:

1. Enter:

```
smit devices
```

SMIT displays a list of devices.
2. Select **Install/Configure Devices Added After IPL** and press Enter.
3. Press F10 to exit SMIT after the **cfgmgr** command completes.
4. Run the following command to ensure that the devices are paired correctly:

```
lsdev -Cc tmsci
```

Repeat the above procedure (enabling and configuring the target mode SCSI device) for other nodes connected to the SCSI-2 bus.

Target Mode Files

Configuring the target mode connection creates two special files in the /dev directory of each node, the /dev/tmcsinn.im and /dev/tmcsinn.tm files. The file with the .im extension is the initiator, which transmits data. The file with the .tm extension is the target, which receives data.

Testing the Target Mode Connection

For the target mode connection to work, initiator and target devices must be paired correctly. To ensure that devices are paired and that the connection is working after enabling the target mode connection on both nodes:

Enter the following command on a node connected to the bus.

```
cat < /dev/tmcsinn.tm
```

where *nn* must be the logical name representing the target node. (This command hangs and waits for the next command.) On the target node, enter the following command:

```
cat filename > /dev/tmcsinn.im
```

where *nn* must be the logical name of the sending node and *filename* is a file.

The contents of the specified file are displayed on the node on which you entered the first command.

Note: Target mode SCSI devices are not always properly configured during the AIX boot process. Ensure that all tmscsi initiator devices are available on all cluster nodes before bringing up the cluster. Use the “`lsdev -Cc tmscsi`” command to ensure that all devices are available. See the *HACMP/ES Troubleshooting Guide* for more information regarding problems with target mode SCSI devices.

Note: If the SCSI bus is disconnected while running as a target mode SCSI network, you must shut down HACMP/ES before reattaching the SCSI bus to that node. *Never attach to a running system.*

Follow-Up Task

After you have installed and tested the target mode SCSI bus, you must define the target mode connection as a serial network to the HACMP/ES cluster environment. Chapter 18, *Configuring an HACMP/ES Cluster*, contains the instructions.

Configuring Target Mode SSA Connections

This section describes how to configure a target mode SSA connection between nodes sharing disks connected to SSA on Multi-Initiator RAID adapters (FC 6215 and FC 6219). The adapters must be at Microcode Level 1801 or later.

You can define a serial network to HACMP/ES that connects all nodes on an SSA loop.

Changing Node Numbers on Systems in SSA Loop

By default, node numbers on all systems are zero. In order to configure the target mode devices, you must first assign a unique non-zero node number to all systems on the SSA loop.

1. To change the node number use the following command.

```
chdev -l ssar -a node_number=#
```
2. To show the system's node number use the following command.

```
lsattr -El ssar
```

Configuring Target Mode SSA Devices

After enabling the target mode interface, you must run **cfgmgr** to create the initiator and target devices and make them available. To configure the devices and make them available:

1. Enter:

```
smit devices
```

SMIT displays a list of devices.
2. Select **Install/Configure Devices Added After IPL** and press Enter
3. Press F10 to exit SMIT after the **cfgmgr** command completes.
4. Run the following command to ensure that the devices are paired correctly:

```
lsdev -Cc tmssa
```

Repeat the above procedure (enabling and configuring the target mode SSA device) for other nodes connected to the SSA adapters.

Target Mode Files

Configuring the target mode connection creates two special files in the `/dev` directory of each node, the `/dev/tmssa#.im` and `/dev/tmssa#.tm` files. The file with the `.im` extension is the initiator, which transmits data. The file with the `.tm` extension is the target, which receives data

Testing the Target Mode Connection

For the target mode connection to work, initiator and target devices must be paired correctly. To ensure that devices are paired and that the connection is working after enabling the target mode connection on both nodes:

1. Enter the following command on a node connected to the SSA disks.

```
cat < /dev/tmssa#.tm
```

where # must be the number of the target node. (This command hangs and waits for the next command.)

2. On the target node, enter the following command:

```
cat filename > /dev/tmssa#.im
```

where # must be the number of the sending node and *filename* is a file.

The contents of the specified file are displayed on the node on which you entered the first command.

3. You can also check that the tmssa devices are available on each system using the following command:

```
lsdev -C | grep tmssa
```

Defining the Target Mode SSA Serial Network to HACMP/ES

Take the following steps to configure the Target Mode SSA serial network in the HACMP/ES cluster.

1. Select Configure Adapters from the Cluster Topology menu.
2. Select Add an Adapter and press Enter.

Enter the fields as follows.

Adapter Label	Unique label for adapter. For example, <i>adp_tmssa_1</i>
Network Type	Pick <i>tmssa</i> from the pop up pick list.
Network Name	Arbitrary name used consistently for all adapters on this network. For example, <i>tmssa_1</i>
Network Attribute	Set this field to Serial
Adapter Function	Set this field to service .
Adapter Identifier	Enter the device name <i>/dev/tmssa#</i>

Adapter Hardware Address Leave this field blank.

Node Name Enter the name of the node the adapter is connected to.

3. Press Enter. The system adds these values to the HACMP/ES ODM and returns you to the Configure Adapters menu.
4. Repeat the above procedure to define the other adapters connected on the SSA loop.

Verifying Shared IBM SSA Disk Subsystems

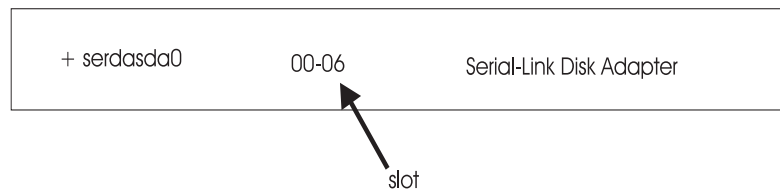
Complete the following steps to verify a shared IBM SSA disk subsystem. As you verify the installation, record the shared disk configuration on the *Shared IBM SSA Disk Subsystem Worksheet*. Use a separate worksheet for each set of shared IBM SSA disk subsystems. You refer to the completed worksheets when you configure the cluster.

1. Fill in the node name of each node connected to the shared IBM SSA disk subsystem in the **Node Name** field.
2. Record the logical device name of each adapter in the **Adapter Logical Name** field.

To get the logical device name, enter the following command at each node:

```
lscfg | grep ssa
```

The first column of the resulting display lists the logical device names of the SSA adapters.

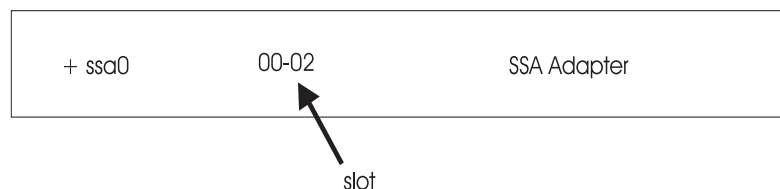


```
+ serdasda0      00-06      Serial-Link Disk Adapter
```

An arrow points from the text "slot" below to the "06" in the "00-06" column of the terminal output.

3. For each node, record the slot that each adapter uses in the **Slot Number** field. The slot number value is an integer value from 1 through 16.

The second column of the existing display lists a value of the form AA-BB. The last digit of that value (the last B) is the Microchannel slot number.



```
+ ssa0            00-02            SSA Adapter
```

An arrow points from the text "slot" below to the "02" in the "00-02" column of the terminal output.

4. Determine the logical device name and size of each physical volume and record the values on the worksheet. On each node use the command:

```
lsdev -Cc disk | grep -i ssa
```

The first column of the resulting display lists the logical names of the disks.

Checking Installed Hardware

Verifying Shared IBM SSA Disk Subsystems

name	status	location	description
hdisk1	Available	00-02-L	SSA Logical Disk Drive
hdisk2	Available	00-02-L	SSA Logical Disk Drive
hdisk3	Available	00-02-L	SSA Logical Disk Drive

logical name

Enter the name in the **Logical Device Name** field.

Record the size of each external disk in the **Size** field.

5. Verify that all disks have a status of "Available." The second column of the existing display indicates the disk status.

name	status	location	description
hdisk1	Available	00-02-L	SSA Logical Disk Drive
hdisk2	Available	00-02-L	SSA Logical Disk Drive
hdisk3	Available	00-02-L	SSA Logical Disk Drive

status

If a disk has a status of defined, instead of available, check the cable connections and then use the **mkdev** command to make the disk available.

At this point, you have verified that the IBM SSA disk is configured properly for the HACMP/ES software.

Checking Installed Hardware
Verifying Shared IBM SSA Disk Subsystems

Chapter 12 Defining Shared LVM Components

This chapter describes how to define the LVM components shared by the nodes in an HACMP/ES cluster.

Overview

Creating the volume groups, logical volumes, and filesystems shared by the nodes in an HACMP/ES cluster requires that you perform steps on all nodes in the cluster. In general, you define the components on one node (referred to in the text as the source node) and then import the volume group on the other nodes in the cluster (referred to as destination nodes). This ensures that the ODM definitions of the shared components are the same on all nodes.

HACMP/ES non-concurrent access environments typically use journaled filesystems to manage data, while concurrent access environments use raw logical volumes. This chapter provides instructions for both types of environments.

TaskGuide for Creating Shared Volume Groups

The TaskGuide is a graphical interface that simplifies the task of creating a shared volume group within an HACMP/ES cluster configuration. The TaskGuide presents a series of panels that guide the user through the steps of specifying initial and sharing nodes, disks, concurrent or non-concurrent access, volume group name, physical partition size, and cluster settings. The TaskGuide can reduce errors, as it does not allow a user to proceed with steps that conflict with the cluster's configuration. Online help panels give additional information to aid in each step.

The TaskGuide for creating a shared volume group was introduced in HACMP 4.3.0. In version 4.4, the TaskGuide has two enhancements: it automatically creates a JFS log, as you would do manually when creating a shared volume group without the TaskGuide. In addition, it now displays the physical location of available disks.

Note that you may still want to rename and mirror the default JFS log after creating the shared volume group, as discussed on page 12-4.

TaskGuide Requirements

Before starting the TaskGuide, make sure:

- You have a configured HACMP/ES cluster in place.
- You are on a graphics capable terminal.
- You have set the display to your machine using your IP address or an alias, for example:

```
export DISPLAY=<your IP address>:0.0
```

Starting the TaskGuide

If you have the TaskGuide filesets installed and your display set properly, you can start the TaskGuide from the command line by typing

```
/usr/sbin/cluster/tguides/bin/cl_ccvg
```

or you can use the SMIT interface as follows:

1. Type `smit hacmp`
2. From the SMIT main menu, choose **Cluster System Management > Cluster Logical Volume Manager > Taskguide for Creating a Shared Volume Group**

After a pause, the TaskGuide “Welcome” panel appears.

3. Proceed through the panels that guide you through the steps to create or share a volume group, and enter the information appropriate to your cluster.

In the last panel, you have the option to cancel or to back up and change what you have entered. If you are satisfied with your entries, click **Apply** to create the shared volume group.

Defining Shared LVM Components for Non-Concurrent Access

HACMP/ES non-concurrent access environments typically use journaled filesystems to manage data, though some database applications may bypass the journaled filesystem and access the logical volume directly.

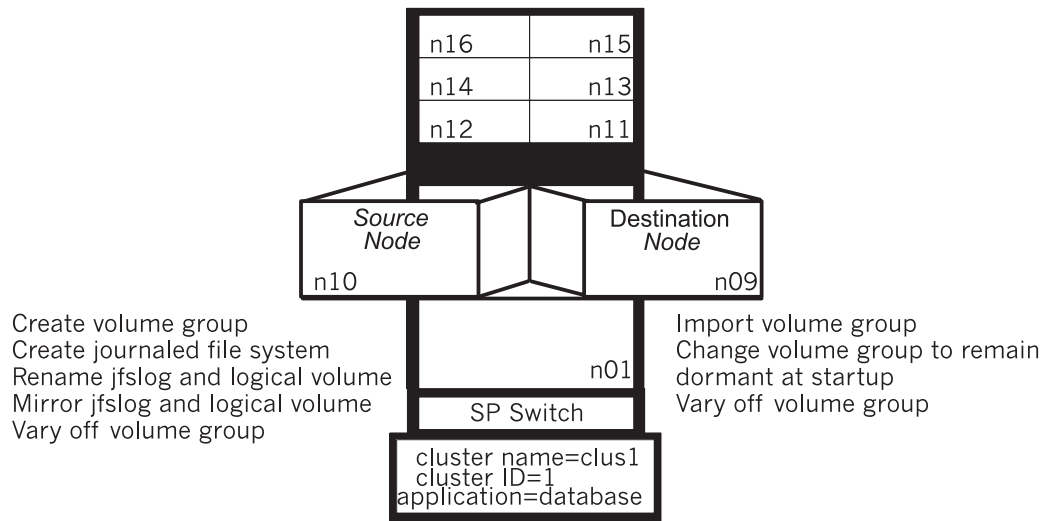
The key consideration, however, is whether the environment uses mirrors. Shared logical volumes residing on SCSI-2 Differential and SCSI-2 Differential Fast/Wide disk devices, IBM 7137 Disk Arrays, and IBM 7133 SSA devices should be mirrored in AIX to eliminate the disk as a single point of failure. Shared volume groups residing on an IBM 7135 or IBM 2105 Versatile Storage Server RAID devices should not be AIX mirrored; these systems provide their own data redundancy.

Note: The discussion of the IBM 7135 RAIDiant Disk Array assumes you are using RAID level 1, 3, or 5. RAID level 0 does not provide data redundancy and therefore is not recommended for use in an HACMP/ES configuration.

The following figures list the tasks you complete to define the shared LVM components with and without mirrors. Each task is described throughout the pages following the figures. Refer to your completed copies of the shared volume group worksheets as you define the shared LVM components.

Defining Shared LVM Components with AIX Mirroring

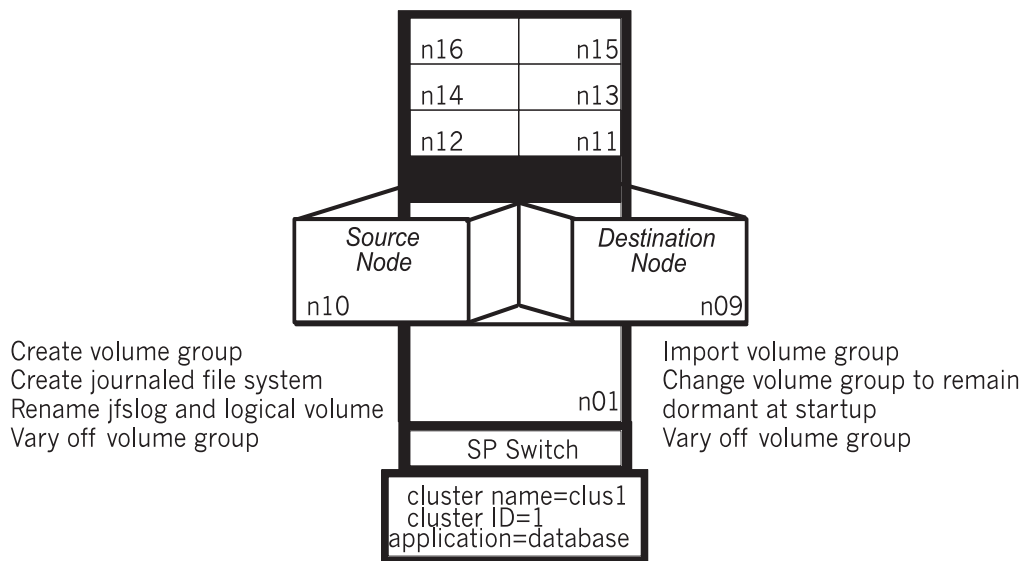
The following figure shows the steps to take to define shared LVM components with AIX mirroring.



Defining Shared LVM Components with AIX Mirroring

Defining Shared LVM Components without AIX Mirroring

The following figure shows the steps to take to define shared LVM components without AIX mirroring.



Defining Shared LVM Components Without AIX Mirroring

Creating a Shared Volume Group on Source Node

Use the **smit mkvg** fastpath to create a shared volume group. Use the default field values unless your site has other requirements, or unless you are specifically instructed otherwise here.

VOLUME GROUP name The name of the shared volume group should be unique within the cluster.

Activate volume group AUTOMATICALLY at system restart? Set to **no** so that the volume group can be activated as appropriate by the cluster event scripts.

ACTIVATE volume group after it is created? Set to **yes**.

Volume Group MAJOR NUMBER Make sure to use the same major number on all nodes. Use the **lvlstmajor** command on each node to determine a free major number common to all nodes.

Creating a Shared Filesystem on Source Node

Use the **smit crjfs** fast path to create the shared filesystem on the source node. When you create a journaled filesystem, AIX creates the corresponding logical volume. Therefore, you do not need to define a logical volume. You do, however, need to later rename both the logical volume and the log logical volume for the filesystem and volume group.

Mount AUTOMATICALLY at system restart? Make sure this field is set to **no**.

Start Disk Accounting Make sure this field is set to **no**.

Renaming jfslogs and Logical Volumes on Source Node

AIX assigns a logical volume name to each logical volume it creates. Examples of logical volume names are */dev/lv00* and */dev/lv01*. Within an HACMP/ES cluster, the name of any shared logical volume *must* be unique. Also, the journaled filesystem log (**jfslog**) is a logical volume that requires a unique name in the cluster.

To make sure that logical volumes have unique names, rename the logical volume associated with the filesystem and the corresponding **jfslog** logical volume. Use a naming scheme that indicates the logical volume is associated with a certain filesystem. For example, *lvsharefs* could name a logical volume for the */sharefs* filesystem.

1. Use the **lsvg -l volume_group_name** command to determine the name of the logical volume and the log logical volume (**jfslog**) associated with the shared volume groups. In the resulting display, look for the logical volume name that has type **jfs**. This is the logical volume. Then look for the logical volume name that has type **jfslog**. This is the log logical volume.

2. Use the **smit chlv** fastpath to rename the logical volume and the log logical volume.
After renaming the **jfslog** or a logical volume, check the **/etc/filesystems** file to make sure the **dev** and **log** attributes reflect the change. Check the **log** attribute for each filesystem in the volume group and make sure that it has the new **jfslog** name. Check the **dev** attribute for the logical volume you renamed and make sure that it has the new logical volume name.

Adding Copies to Logical Volume on Source Node

To add logical volume copies on a source node:

1. Use the **smit mklvcopy** fastpath to add copies to a logical volume. Add copies to both the **jfslog** log logical volume and the logical volumes in the shared filesystems. To avoid space problems, first mirror the **jfslog** log logical volume and then the shared logical volumes.
The copies should reside on separate disks that are controlled by different disk adapters and are located in separate drawers or units, if possible.

Note: These steps do not apply to IBM 7135-110 and 7135-210 RAIDiant Disk Arrays, which provide their own mirroring of logical volumes. Continue with Testing a Filesystem on page 12-5.

2. Verify the number of logical volume copies. Enter:

```
lsvg -l volume_group_name
```

In the resulting display, locate the line for the logical volume for which you just added copies. Notice that the number in the physical partitions column is x times the number in the logical partitions column, where x is the number of copies.

3. To verify the placement of logical volume copies, enter:

```
lspv -l hdiskx
```

where *hdiskx* is the name of each disk to which you assigned copies. That is, you enter this command for each disk. In the resulting display, locate the line for the logical volume for which you just added copies. For copies placed on separate disks, the numbers in the logical partitions column and the physical partitions column should be equal. Otherwise, the copies were placed on the same disk and the mirrored copies will not protect against disk failure.

Testing a Filesystem

To run a consistency check on each filesystem's information:

1. Enter:

```
fsck /filesystem_name
```

2. Verify that you can mount the filesystem by entering:

```
mount /filesystem_name
```

3. Verify that you can unmount the filesystem by entering:

```
umount /filesystem_name
```

Varying Off a Volume Group on the Source Node

After completing the previous tasks, use the **varyoffvg** command to deactivate the shared volume group. You vary off the volume group so that it can be properly imported onto the destination node and activated as appropriate by the cluster event scripts. Enter the following command:

```
varyoffvg volume_group_name
```

Importing a Volume Group onto Destination Nodes

Importing the volume group onto the destination nodes synchronizes the ODM definition of the volume group on each node on which it is imported. Use the **smit importvg** fastpath to import the volume group.

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 source node.
PHYSICAL VOLUME name	Enter the name of a physical volume that resides in the volume group. Note that a disk may have a different physical name on different nodes. Make sure that you use the disk name as it is defined on the destination node.
ACTIVATE volume group after it is imported?	Set the field to yes .
Volume Group MAJOR NUMBER	Use the same major number on all nodes. Use the lvstmajor command on each node to determine a free major number common to all nodes.

Changing a Volume Group's Startup Status

By default, a volume group that has just been imported is configured to automatically become active at system restart. In an HACMP/ES environment, a volume group should be varied on as appropriate by the cluster event scripts. Therefore, after importing a volume group, use the Change a Volume Group screen to reconfigure the volume group so that it is not activated automatically at system restart.

Use the **smit chvg** fastpath to change the characteristics of a volume group.

Activate volume group Automatically at system restart? Set this field to **no**.

A QUORUM of disks required to keep the volume group on-line? This field is site-dependent. See Chapter 5, Planning Shared Disk Devices, for a discussion of quorum.

Varying Off Volume Group on Destination Nodes

Use the **varyoffvg** command to deactivate the shared volume group so that it can be imported onto another destination node or activated as appropriate by the cluster event scripts. Enter:

```
varyoffvg volume_group_name
```

Defining Shared LVM Components for Concurrent Access

Concurrent access does not support filesystems. Instead, you must use logical volumes.

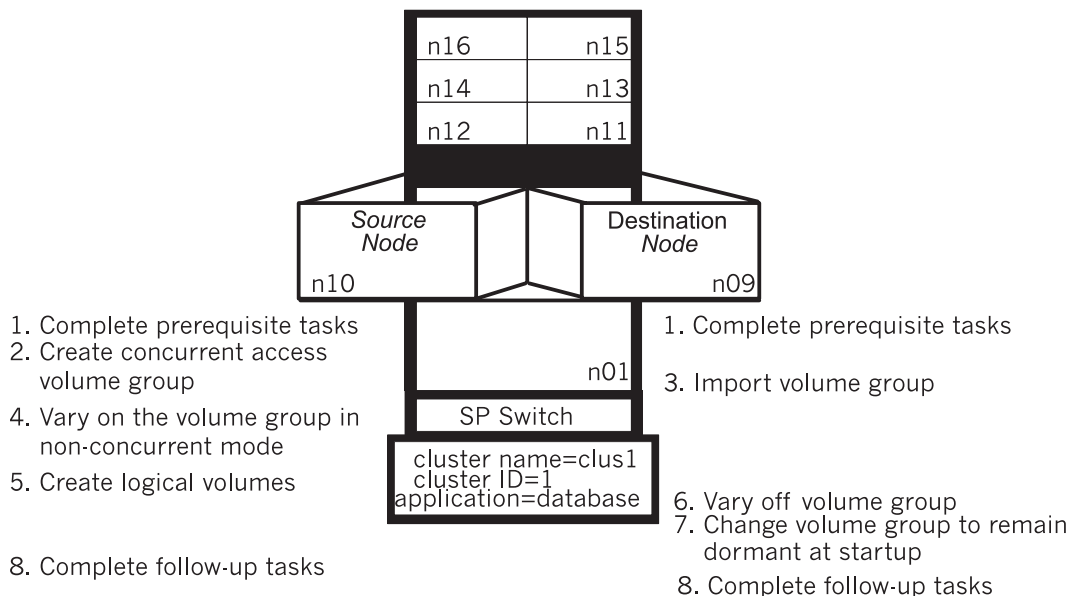
This section describes the procedure for defining shared LVM components for a concurrent access environment. Concurrent access is supported on the following devices:

- IBM 7135-110 and 210 RAIDiant Disk Array
- IBM 7137 Disk Arrays
- IBM 9333 serial disk subsystems
- IBM 7133 or 7131-405 SSA Disk Subsystems
- IBM 2105-B09 and 100 Versatile Storage Servers

Refer to your completed copies of the shared volume group worksheets as you define the shared LVM components.

Creating a Concurrent Access Volume Group on a Source Node

The figure below summarizes the steps you must complete on the source and destination nodes in an HACMP/ES cluster to create a concurrent capable volume group that HACMP/ES can vary on in concurrent access mode.



Creating a Concurrent Access Volume Group

The above steps are described in detail throughout this section.

Step 1. Complete Prerequisite Tasks

The physical volumes (hdisks) should be installed, configured, and available. You can verify the disks' status using the **lsdev -Cc disk** command.

Step 2. Create a Concurrent Access Volume Group on Source Node

The procedure used to create a concurrent access volume group varies depending on which type of device you are using: serial disk subsystem or RAID disk subsystem.

Creating Concurrent Access Volume Groups on SSA Devices

If you are creating (or plan to create) concurrent volume groups on SSA devices, be sure to assign unique non-zero node numbers through the `ssar` on each cluster node. If you plan to specify SSA disk fencing in your concurrent resource group, the node numbers are assigned when you synchronize resources. If you do not specify SSA disk fencing, assign node numbers using the following command: `chdev -l ssar -a node_number=x`, where `x` is the number to assign to that node. You must reboot the system to effect the change.

Creating a Concurrent Access Volume Group on Serial Disk Subsystems

To use a concurrent access volume group, defined on a serial disk subsystem such as an IBM 9333 or IBM 7133 disk subsystem, you must create it as a *concurrent capable* volume group. A concurrent capable volume group can be activated (varied on) in either non-concurrent mode or concurrent access mode. To define logical volumes on a concurrent capable volume group, it must be varied on in non-concurrent mode.

Use the `mkvg` command, specifying the `-c` flag, to create the concurrent capable volume group, on `myvg` as in the following example:

```
mkvg -n -s 4 -c -y myvg hdisk1 hdisk2
```

You can also use SMIT to build the `mkvg` command by using the following procedure:

1. To create a concurrent capable volume group, enter:

```
smit mkvg
```

SMIT displays the **Add a Volume Group** screen. Enter the specific field values as follows:

VOLUME GROUP name Specify name of volume group.

Physical partition SIZE in megabytes Accept the default.

PHYSICAL VOLUME NAMES Specify the names of the physical volumes you want included in the volume group.

Activate volume group AUTOMATICALLY at system restart? Set this field to **no** so that the volume group can be activated as appropriate by the cluster event scripts.

ACTIVATE volume group after it is created? Set this field to **no**.

Volume Group MAJOR NUMBER Accept the default.

Create VG concurrent capable? Set this field to **yes** so that the volume group can be activated in concurrent access mode by the HACMP/ES event scripts.

Auto-varyon concurrent mode? Set this field to **no** so that the volume group can be activated as appropriate by the cluster event scripts.

2. Press Enter. SMIT responds:
are you sure?
3. Press Enter.
4. Press F10 to exit SMIT after the command completes.

Creating a Concurrent Access Volume Group on RAID Disk Subsystems

To create a concurrent access volume group on a RAID disk subsystem, such as an IBM 7135 disk subsystem or IBM 2105 Versatile Storage Server, you follow the same procedure as you would to create a non-concurrent access volume group. A concurrent access volume group can be activated (varied on) in either non-concurrent mode or concurrent access mode. To define logical volumes on a concurrent access volume group, it must be varied on in non-concurrent mode.

Use the **smit mkvg** fastpath to create a shared volume group. Use the default field values unless your site has other requirements, or unless you are specifically instructed otherwise.

VOLUME GROUP name The name of the shared volume group should be unique within the cluster.

Activate volume group AUTOMATICALLY at system restart? Set to **no** so that the volume group can be activated as appropriate by the cluster event scripts.

ACTIVATE volume group after it is created? Set to **yes**.

Volume Group MAJOR NUMBER If you are not using NFS, use the default (which is the next available number in the valid range). If you are using NFS, you must make sure to use the same major number on all nodes. Use the `lvlstmajor` command on each node to determine a free major number common to all nodes.

Create VG concurrent capable? Set this field to **no**.

Step 3. Import Volume Group Information on Destination Nodes

On each destination node, import the volume group, using the **importvg** command, as in the following example:

```
importvg -y vg_name physical_volume_name
```

Specify the name of any disk in the volume group as an argument to the **importvg** command. By default, AIX automatically varies on non-concurrent capable volume groups when they are imported. AIX does *not* automatically vary on concurrent capable volume groups when they are imported. You can also build the **importvg** command through SMIT using the procedure below.

Defining Shared LVM Components

Defining Shared LVM Components for Concurrent Access

To import a concurrent capable volume group:

1. Enter:

```
smit importvg
```

SMIT displays the **Import a Volume Group** SMIT screen.

2. Enter specific field values as follows. For other fields, use the defaults or the appropriate entries for your operation:

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 source node.

PHYSICAL VOLUME name Enter the name of one of the physical volumes that resides in the volume group. *Note that a disk may have a different hdisk number on different nodes. Make sure that you use the disk name as it is defined on the destination node.* Use the **lspv** command to map the hdisk number to the PVID. The PVID uniquely identifies a disk.

ACTIVATE volume group after it is imported? Set the field to **no**.

Volume Group MAJOR NUMBER Accept the default.

Make this VG concurrent capable Accept the default.

Make default varyon of VG concurrent Accept the default.

3. Press Enter to commit the information. Press F10 to exit SMIT and return to the command line.

If your cluster uses SCSI external disks (including RAID devices) and the import of the volume group fails, check that no reserve exists on any disk in the volume group by executing the following command, only after installing the HACMP/ES software as described in Chapter 14, *Installing the HACMP/ES Software*:

```
/usr/es/sbin/cluster/events/utils/cl_scdiskreset /dev/hdiskn ...
```

For example, if the volume group consists of *hdisk1* and *hdisk2*, enter:

```
/usr/es/sbin/cluster/events/utils/cl_scdiskreset /dev/hdisk1 /dev/hdisk2
```

Step 4. Varyon the Concurrent Capable Volume Group in Non-concurrent Mode

Use the **varyonvg** command to activate a volume group in non-concurrent mode. To create logical volumes, the volume group must be varied on in non-concurrent access mode. For example, to vary on the concurrent capable volume group **myvg** in non-concurrent access mode, enter the following command:

```
varyonvg myvg
```

You can also use SMIT to build the **varyonvg** command by using the following procedure.

1. To vary on a concurrent capable volume group in non-concurrent mode, enter:

```
smit varyonvg
```

SMIT displays the **Add a Volume Group** SMIT screen. Enter the specific field values as follows.

VOLUME GROUP name	Specify name of volume group.
RESYNCHRONIZE stale physical partitions?	Set this field to no .
Activate volume group in SYSTEM MANAGEMENT mode?	Accept the default.
FORCE activation of the volume group?	Accept the default.
Varyon volume group in concurrent mode	Accept the default. To create logical volumes on the volume group, it must be varied on in non-concurrent mode.

2. Press Enter.
SMIT responds: **are you sure?**
3. Press Enter.
4. Press F10 to exit SMIT after the command completes.

Step 5. Create Logical Volumes on Concurrent Capable Volume Group on Source Node

Create logical volumes on the volume group, specifying logical volume mirrors to provide data redundancy. If the volume group is varied on in concurrent access mode, you will not be able to create logical volumes. A concurrent capable volume group must be varied on in non-concurrent access mode to create logical volumes on it.

For more information about creating logical volumes, see the *AIX System Management Guide: Operating System and Devices*.

Use the SMIT fastpath **smit mklv**. You must specify the size of the logical volume as the number of logical partitions. Accept default values for all other fields except the following:

Logical volume name	Specify name of logical volume. Name must be the same on all cluster nodes.
PHYSICAL VOLUME names?	Specify the physical volumes you want the logical volume to include.
Activate volume group in SYSTEM MANAGEMENT mode?	Accept the default.

Defining Shared LVM Components

Defining Shared LVM Components for Concurrent Access

Number of COPIES of each logical partition	Specify 1, 2, or 3 mirror copies. If You defined the volume group on an IBM 7135-110 or IBM 7135-210 disk array, do not create mirror copies. Instead, use the data redundancy provided by RAID levels 1, 3, or 5.
Mirror Write Consistency	Specify the value to no .

Step 6. Vary Off Volume Group on Source Node

After creating the logical volume, vary off the volume group using the **varyoffvg** command so that it can be varied on by the HACMP/ES scripts. Enter:

```
varyoffvg volume_group_name
```

Step 7. Change Volume Group to Remain Dormant at Startup on Destination Nodes

By default, AIX configures an imported volume group to automatically become active at system restart. In the HACMP/ES system, a volume group should be varied on as appropriate by the HACMP/ES scripts. Therefore, after importing a volume group, you must reconfigure the volume group so that it remains dormant at startup.

To change the startup state of a volume group, enter:

```
chvg -a n volume_group_name
```

You can also build the **chvg** command through SMIT using the following procedure:

1. Use the **smit chvg** fastpath to change the characteristics of a volume group.
Enter the specific field values as follows. For other fields use the defaults or the appropriate entries for your operation:
Set the **Activate volume group Automatically at system restart?** field to **no**.
2. Press Enter to commit this change.
3. Press F10 to exit SMIT and return to the command line.

Step 8. Complete Follow-up Tasks

Verify that the HACMP/ES scripts can activate the concurrent capable volume group as a concurrent cluster resource.

Chapter 13 Tailoring AIX for HACMP/ES

This chapter discusses several general tasks necessary for ensuring that your HACMP/ES environment works as planned.

Note: The directory `/usr/sbin/cluster` and subdirectories have symbolic links to the `/usr/es/sbin/cluster` directory and subdirectories. Files in those directories are *not* linked, as they were in releases prior to 4.3.1.

Tailoring AIX

Consider the following issues to ensure that AIX works as expected in an HACMP/ES cluster:

- I/O pacing
- User and group IDs
- Network option settings
- `/etc/hosts` file and name server edits
- `/.rhosts` file edits. (if not using Kerberos)
- Editing the `/etc/rc.net` file on NFS Clients
- SP Switch Address Resolution Protocol (ARP)
- Automounter Daemon
- Error notification facility
- Automatic error notification

Consult the *IBM Parallel System Support Programs for AIX: Administration Guide* for more information on managing AIX on the SP.

I/O Pacing

By default, AIX 4.3.3 is installed with high- and low-water marks set to **zero**, which disables I/O pacing. While enabling I/O pacing may have a slight performance effect on I/O intensive processes, it is *required* for an HACMP/ES cluster to behave correctly during large disk writes. If you anticipate heavy I/O on your HACMP/ES cluster, you should enable I/O pacing.

See Planning for Cluster Performance on page 4-17 for information about tuning networks for optimal throughput, including I/O pacing, using HACMP/ES SMIT screens. See Chapter 18, Configuring an HACMP/ES Cluster for more details on configuring these parameters.

You can also use the **smit chgsys** fastpath to set high and low water marks on the Change/Show Characteristics of Operating System screen. Although the most efficient high and low water marks vary from system to system, an initial high water mark of **33** and a low water mark of **24** provides a good starting point. These settings only slightly reduce write times and consistently generate correct fallover behavior from the HACMP/ES software.

See the *AIX Performance Tuning Guide* for more information on I/O pacing.

Checking User and Group IDs

If a node fails, users should be able to log on to the surviving nodes without experiencing problems caused by mismatches in the user or group IDs. To avoid mismatches, make sure that user and group information is propagated to nodes as necessary. User and group IDs should be the same on all nodes.

Network Options

HACMP/ES requires that the **nonlocsrcroute**, **bcasting**, **ipsrouteseend**, **ipsrouterecv**, and **ipsrouteforward** network options be set to 1; this happens during **clstrmgrES** initialization.

Changing routerevalidate Network Option

Changing hardware and IP addresses within HACMP/ES changes and deletes routes. Due to the fact that AIX caches routes, it is required that you set the “routerevalidate” network option as follows:

```
routerevalidate=1
```

This setting ensures the maintenance of communication between cluster nodes.

- To change the default value, add the following line to the end of the **/etc/rc.net** file:

```
no -o routerevalidate=1
```

Editing the /etc/hosts File and nameserver Configuration

Make sure all nodes can resolve all cluster addresses. If you are using NIS or DNS, review the section Using HACMP/ES with NIS and DNS on page 4-14.

Edit the **/etc/hosts** file (and the **/etc/resolv.conf** file, if using the **nameserver** configuration) on each node in the cluster to make sure the IP addresses of all clustered interfaces are listed. Make sure that all service, standby, and boot addresses are listed.

For each boot address, make an entry similar to the following:

```
100.100.50.200 crab_boot
```

Also, make sure that the **/etc/hosts** file on each node has the following entry:

```
127.0.0.1 loopback localhost
```

cron and NIS Considerations

If your HACMP/ES cluster nodes use NIS services which include the mapping of the **/etc/passwd** file and IPAT is enabled, users that are known only in the NIS-managed version of the **/etc/passwd** file will not be able to create crontabs. This is because **cron** is started via the **/etc/inittab** file with run level 2 (for example, when the system is booted), but **ypbind** is started in the course of starting HACMP/ES via the **rcnfs** entry in **/etc/inittab**. When IPAT is enabled in HACMP/ES, the run level of the **rcnfs** entry is changed to **-a** and run via the **telinit -a** command by HACMP/ES.

In order to let those NIS-managed users create crontabs, you can do one of the following:

1. Change the runlevel of the **cron** entry in **/etc/inittab** to **-a** and make sure it is positioned after the **rcnfs** entry in **/etc/inittab**. This solution is recommended if it is acceptable to start **cron** after HACMP/ES has started.
2. Add an entry to the **/etc/inittab** file like the following script with runlevel **-a**. Make sure it is positioned after the **rcnfs** entry in **/etc/inittab**. The important thing is to kill the **cron** process, which will respawn and know about all of the NIS-managed users. Whether or not you log the fact that **cron** has been refreshed is optional.

Sample Script

```
#!/bin/sh
# This script checks for a ypbind and a cron process. If both
# exist and cron was started before ypbind, cron is killed so
# it will respawn and know about any new users that are found
# in the passwd file managed as an NIS map.
echo "Entering $0 at `date`" >> /tmp/refr_cron.out
cronPid=`ps -ef |grep "/etc/cron" |grep -v grep |awk \
' { print $2 } '`
ypbindPid=`ps -ef | grep "/usr/etc/ypbind" | grep -v grep | \
if [ ! -z "${ypbindPid}" ]
then
    if [ ! -z "${cronPid}" ]
    then
        echo "ypbind pid is ${ypbindPid}" >> /tmp/refr_cron.out
        echo "cron pid is ${cronPid}" >> /tmp/refr_cron.out
        echo "Killing cron(pid ${cronPid}) to refresh user \
            list" >> /tmp/refr_cron.out
        kill -9 ${cronPid}
        if [ $? -ne 0 ]
        then
            echo "$PROGNAME: Unable to refresh cron." \
                >>/tmp/refr_cron.out
            exit 1
        fi
    fi
fi
echo "Exiting $0 at `date`" >> /tmp/refr_cron.out
exit 0
```

Editing the **/.rhosts** File

Note: If you configure your system to use Kerberos authentication, you do not need to use the **.rhosts** file. See Chapter 18, Configuring an HACMP/ES Cluster, for more information about setting up a Kerberos configuration.

Make sure that each node's service adapters and boot addresses are listed in the **/.rhosts** file on all nodes in the cluster. Doing so allows the **/usr/es/sbin/cluster/utilities/clruncmd** command and the **/usr/es/sbin/cluster/godm** daemon to run. (The **/usr/es/sbin/cluster/godm** daemon is used when nodes are configured from a central location.)

For security reasons, you can add entries to the **/.rhosts** file only as necessary, and delete them when they are no longer needed. The **/usr/es/sbin/cluster/clstrmgrES** daemon does not depend on the **/.rhosts** file. The cluster synchronization and verification functions, however, use **rcmd** and **rsh** and thus require these **/.rhosts** entries.

Editing the `/etc/rc.net` File on NFS Clients

By default, AIX 4.3.3 NFS clients do not respond to a ping on the broadcast address. Therefore, on any clients that use NFS, add the following line to the `/etc/rc.net` file:

```
/usr/sbin/no -o bcastping=1
```

SP Switch Address Resolution Protocol (ARP)

If your SP nodes are already installed and the switch network is up on all nodes, you can verify if ARP is enabled. On the control workstation, enter the following command:

```
dsh -av "/usr/lpp/ssp/css/ifconfig css0"
```

If NOARP appears on output from any of the nodes, you will have to enable ARP to use IP takeover on the SP Switch.

Note: You must enable ARP on all SP nodes connected to the SP Switch. Use the following method. Enter all commands from the control workstation. Ensure all nodes are up. You must type quotes listed in the commands.

Warning: Complete all steps carefully. The following sequence requires that CuAt be backed up. If CuAt is not backed up on the nodes and user error corrupts CuAt and reboots, the SP nodes may be corrupted and have to be re-installed. If you corrupt CuAt on the nodes, be sure that you copy your backup (CuAt.save) back to `/etc/objrepos/CuAt` prior to any reboot. Be careful! If you feel this is too risky, customize the nodes to turn ARP on (see the *IBM Parallel System Support Programs for AIX: Administration Guide* for this procedure).

1. `dsh -av "cp /etc/objrepos/CuAt /etc/objrepos/CuAt.save"`
2. `dsh -av "odmget -q'name=css and attribute=arp_enabled' CuAt | sed s/no/yes/ > /tmp/arpon.data"`
3. `dsh -av "odmchange -o CuAt -q'name=css and attribute=arp_enabled' /tmp/arpon.data"`
4. Save the new information onto the boot device using the `savebase` command.
5. Verify that the previous commands worked by issuing the following command:
`dsh -av "odmget -a'name=css and name=arp_enabled' CuAt | grep value"`
You should see an entry from every node reporting "value=yes".
6. `dsh -av rm /tmp/arpon.data`
7. Shut down and reboot the nodes.

AIX Automounter Daemon

For SP installations that require the AIX automounter daemon on HACMP/ES nodes, a modification is needed to ensure that Automounter starts properly (with NFS available and running) on node boot. This is due to the way HACMP/ES manages the **inittab** file and run levels upon startup.

To enable the automounter on nodes that have HACMP/ES installed, add the following line as the last line of the file **/usr/es/sbin/cluster/etc/harc.net**:

```
startsrc -s nfsd
```

AIX Error Notification Facility

Although the HACMP/ES software does not monitor the status of disk resources, it does provide a SMIT interface to the AIX Error Notification facility. The AIX Error Notification facility allows you to detect an event not specifically monitored by the HACMP/ES software—a disk adapter failure, for example—and to program a response to the event.

Permanent hardware errors on disk drives, controllers, or adapters may impact the fault resiliency of data. By monitoring these errors through error notification methods, you can assess the impact of a failure on the cluster's ability to provide high availability. A simple implementation of error notification would be to send a mail message to the system administrator to investigate the problem further. A more complex implementation could include logic to analyze the failure and decide whether to continue processing, stop processing, or escalate the failure to a node failure and have the takeover node make the volume group resources available to clients.

It is strongly recommended that you implement an error notification method for all errors that affect the disk subsystem. Doing so ensures that degraded fault resiliency does not remain undetected.

Also see the section HACMP Automatic Error Notification on page 13-6 for information on using this utility for assigning error notification methods in one step to a number of selected disk devices.

Global Network Failure Detection and Action

Several options exist for detecting failure and invoking user defined scripts to verify the failure and recover.

The switch power off will be seen as a HPS_FAULT9_ER recorded on each node, followed by HPS_FAULT6_ER (fault service daemon terminated). By modifying the AIX error notification strategies, it is possible to call a user script to detect the global switch failure and perform some recovery action. The user script would have to do the following:

- Detect global network failure (switch power failure or fault service daemon terminated on all nodes). If the failure is local, issue the **clstop -grsy** command to stop HACMP/ES on that node and fall over resources and workloads to a takeover node.
- Take recovery action, such as moving workload to another network, or reconfiguring a backup network.

- In order to recover from a major switch failure (power off, for example), you must issue **Eclock** and **Estart** commands to bring the switch back on-line. The **Eclock** command runs **rc.switch**, which deletes the aliases HACMP/ES needs for SP Switch IP address takeover. It is recommended to create an event script for either the `network_down` or the `network_down_complete` event to add back the aliases for `css0`.

Sample SP Switch Notify Method

In the following example of the Add a Notify Method screen, you specify an error notification method for an SP Switch.

Notification Object Name	HPS_ER9
Persist across system restart?	yes
Process ID for use by Notify Method	
Select Error Class	All
Select Error Type	PERM
Match Alertable Errors?	None
Select Error Label	HPS_FAULT9_ER
Resource Name	all
Resource Class	all
Resource Type	all
Notify Method	“usr/sbin/cluster/utilities/clstop -grsy”

HACMP Automatic Error Notification

Using a SMIT screen option, you can configure error notification automatically for the cluster resources listed below, list currently defined automatic error notify entries for the same cluster resources, or remove previously configured automatic error notify methods. Before you configure Automatic Error Notification, you must have a valid HACMP/ES configuration.

Warning: Automatic error notification should be configured only when the cluster is not running.

Choosing to add error notify methods automatically runs the **cl_errnotify** utility which turns on error notification on all nodes in the cluster for the following devices:

- All disks in the rootvg volume group
- All disks in HACMP/ES volume groups, concurrent volume groups, and filesystems. (To avoid single points of failure, the JFS log must be included in an HACMP/ES volume group.)
- All disks defined as HACMP/ES resources
- The SP switch adapter.

Automatic error notification applies to selected hard, non-recoverable error types: disk, disk adapter, and SP switch adapter errors. No media errors, recovered errors, or temporary errors are supported by this utility.

Executing automatic error notification assigns one of two error notification methods for all the error types noted:

- **cl_failover** is assigned if a disk or an adapter (including SP switch adapter) is determined to be a single point of failure and its failure should cause the cluster resources to fall over. In case of a failure of any of these devices, this method logs the error to **hacmp.out** and shuts down the cluster software on the node. It first tries to do a graceful shutdown with takeover; if this fails, it calls **cl_exit** to shut down the node.
- **cl_logerror** is assigned for all other error types. In case of a failure of any of these devices, this method logs the error to **hacmp.out**.

You can also use the utility to list currently defined auto error notification entries in your HACMP/ES cluster configuration and to delete all automatic error notify methods.

Configuring Automatic Error Notification

To configure automatic error notification, take the following steps:

1. Be sure that the cluster is not running.
2. Open the SMIT main HACMP/ES menu by typing `smit hacmp`.
3. From the main menu, choose **RAS Support > Error Notification > Configure Automatic Error Notification**. The SMIT interface appears as follows:
4. Select the **Add Error Notify Methods for Cluster Resources** option from the following list:

List Error Notify Methods for Cluster Resources	Lists all currently defined auto error notify entries for certain cluster resources: HACMP/ES defined volume groups, concurrent volume groups, filesystems, and disks; rootvg; SP switch adapter (if present). The list is output to the screen.
--	--

Add Error Notify Methods for Cluster Resources	Error notification methods are automatically configured on all relevant cluster nodes.
---	--

Delete Error Notify Methods for Cluster Resources	Error notification methods previously configured with the Add Error Notify Methods for Cluster Resources option are deleted on all relevant cluster nodes.
--	---

5. (optional) Since error notification is automatically configured for all the listed devices on all nodes, you must make any modifications to individual devices or nodes manually, after running this utility. To do so, choose the **Error Notification** option from the **RAS Support** SMIT screen. See the earlier section in this chapter.

Note: If you make any changes to cluster topology or resource configuration, you may need to reconfigure automatic error notification. When you run **clverify** after making any change to the cluster configuration, you will be reminded to reconfigure error notification if necessary.

Listing Error Notify Methods

To see the automatic error notify methods that currently exist for your cluster configuration, take the following steps:

1. From the main HACMP menu, choose **RAS Support > Error Notification > Configure Automatic Error Notification**.
2. Select the **List Error Notify Methods for Cluster Resources** option. The utility lists all currently defined automatic error notification entries with these HACMP/ES components: HACMP/ES defined volume groups, concurrent volume groups, filesystems, and disks; rootvg; SP switch adapter (if present). The list is output to a screen similar to that shown below, in which the cluster nodes are named *sioux* and *quahog*:

```
COMMAND STATUS
Command: OK          stdout: yes          stderr: no
Before command completion, additional instructions may appear below.
sioux:
sioux: HACMP Resource          Error Notify Method
sioux:
sioux: hdisk0                  /usr/sbin/cluster/diag/cl_failover
sioux: hdisk1                  /usr/sbin/cluster/diag/cl_failover
sioux: scsi0                   /usr/sbin/cluster/diag/cl_failover
quahog:
quahog: HACMP Resource          Error Notify Method
quahog:
quahog: hdisk0                  /usr/sbin/cluster/diag/cl_failover
quahog: scsi0                   /usr/sbin/cluster/diag/cl_failover
```

Deleting Error Notify Methods

To delete automatic error notification entries previously assigned using this utility, take the following steps:

1. From the main menu, choose **RAS Support > Error Notification > Configure Automatic Error Notification**.
2. Select the **Delete Error Notify Methods for Cluster Resources** option. Error notification methods previously configured with the **Add Error Notify Methods for Cluster Resources** option are deleted on all relevant cluster nodes.

Emulation of Error Log Entries

After you have added one or more error notification methods to the AIX Error Notification facility, you can test your methods by emulating an error. By inserting an error log entry into the AIX error device file (**/dev/error**), you will cause the AIX error daemon to run the appropriate pre-specified notify method. This will allow you to determine whether your pre-defined action is carried through.

To emulate an error log entry:

1. Open the SMIT main HACMP/ES menu by typing `smit hacmp`
2. From the main menu, choose **RAS Support > Error Notification > Emulate Error Log Entry**.

The **Select Error Label** box appears, showing a picklist of the notification objects for which notify methods have been defined.

3. Select a notification object and press return to begin the emulation.

As soon as you press the return key, the emulation process begins: the emulator inserts the specified error into the AIX error log, and the AIX error daemon runs the notification method for the specified object.

When the emulation is complete, you can view the error log by typing the **errpt** command to be sure the emulation took place. The error log entry has either the resource name EMULATOR, or a name as specified by the user in the **Resource Name** field during the process of creating an error notify object.

You will now be able to determine whether the specified notify method was carried out.

Note: Remember that the actual notify method will be run. Whatever message, action, or executable you defined will occur. Depending on what it is, you may need to take some action, for instance, to restore your cluster to its original state.

Only the root user is allowed to run an error log emulation.

Chapter 14 Installing the HACMP/ES Software

This chapter describes how to install the HACMP/ES for AIX, Version 4.4 Licensed Program Product (LPP) on cluster nodes and clients in an HACMP/ES cluster environment. The chapter contains instructions for:

- Doing a new installation of HACMP/ES 4.4
- Converting HACMP version 4.3.1 or a previous version to HACMP/ES 4.4
- Migrating node-by-node from HACMP version 4.4 to HACMP/ES 4.4.

If you are installing HACMP/ES for AIX on a client only, see Chapter 17, *Installing and Configuring Clients*.

Prerequisites

The HACMP/ES software has the following prerequisites:

- Each cluster node must have AIX, Version 4.3.3 installed.
- Version 3 release 2 of the PSSP (AIX Parallel System Support Programs) or greater must be installed on the SP control workstation and SP nodes.
- The following AIX optional bos components are mandatory for HACMP/ES:
 - bos.adt.lib
 - bos.adt.libm
 - bos.adt.syscalls
 - bos.net.tcp.client
 - bos.net.tcp.server
 - bos.rte.SRC
 - bos.rte.libc
 - bos.rte.libcfg
 - bos.rte.libcur
 - bos.rte.libpthreads
 - bos.rte.odm
 - (If installing Concurrent Resource Manager) bos.rte.lvm.usr4.3.2 or higher.
- Each cluster node requires its own HACMP/ES software license.
- The root user must perform the installation.
- The `/usr` directory must have **52 MB** of free space for a full install.
 - If you are not planning to install optional software you can plan for less space: HAView =14MB, TaskGuide= 400KB, hativoli=400 KB, VSM= 5.2MB. You should also choose to install the message catalogs for the language you will be using, rather than all message catalogs (Japanese message catalogs use 1.6MB).
- The `/` (root) directory must have **500 KB** of free space (beyond any need to extend the `/usr` directory).

- HAView requires that you install the following NetView filesets:
 - on a server:**
 - nv6000.base.obj 4.1.2.0 (or higher) and nv6000.database.obj 4.1.2.0 (or higher)
or
 - nv6000.base.obj 5.0.0.0 or higher
 - on a client:**
 - nv6000.base.obj 4.1.2.0 (or higher)
or
 - nv6000.client.obj 4.1.2.0 (or higher)
 - Cluster monitoring with Tivoli requires that you install the following Tivoli software:
 - Tivoli Framework 3.6 (on TMR and cluster nodes)
 - Tivoli Application Extension Facility (AEF) 3.6 (on TMR only)
 - Tivoli TME 10 Distributed Monitoring 3.5 (on TMR and cluster nodes)
 - Tivoli TME 10 Distributed Monitoring 3.5.1 (on TMR and cluster nodes)
- and the following filesets:
- cluster.hativoli.client
 - cluster.hativoli.server
 - cluster.msg.en_US.hativoli

Overview

Contents of the Installation Media

The HACMP/ES for AIX software installation media contains:

- The RS/6000 Cluster Technology (RSCT) subsystem images that you must install on all cluster nodes. This feature provides Group Services, Topology Services, and Event Management services.
- The High Availability Subsystem images, some of which you must install on all cluster nodes and clients. This feature provides the services for cluster membership, system management, configuration integrity and control, fallover, and recovery. It also includes cluster status and monitoring facilities for programmers and system administrators.

The HACMP/ES Concurrent Resource Manager software installation media contains:

- The Concurrent Resource Manager (CRM) images. This optional feature adds concurrent shared-access management for supported RAID and SSA disk subsystems. Concurrent access is provided at the raw logical volume level. Applications that use the CRM must be able to control access to the shared data. The CRM includes the High Availability Subsystem which provides distributed locking facilities to support access to shared data.

Note: In HACMP/ES Version 4.4 environments, concurrent access is available using only an IBM 7135-110 or 210 Disk Array, an IBM 7137 Disk Array, IBM 2105 Versatile Storage Server (Models B09 and 100), or IBM 7133 SSA disk subsystem. RAID devices from other manufacturers may not support concurrent access.

Conversion and Migration from HACMP for AIX to HACMP/ES 4.4

If you presently use the HACMP for AIX software, you can convert your cluster to the HACMP/ES for AIX, Version 4.4 software using a cluster snapshot. See the section Converting from HACMP for AIX to HACMP/ES 4.4 on page 14-12 for information.

If you presently use the HACMP for AIX version 4.4 software, You can choose to perform a node-by-node migration of your HACMP for AIX 4.4 cluster to HACMP/ES 4.4 without taking the entire cluster down. See the section Node-by-Node Migration from HACMP for AIX Version 4.4 to HACMP/ES Version 4.4 on page 14-14.

Setting the LANG Variable

US English and Japanese message catalogs are available. Set the LANG variable to one of these so that SMIT help is available in the desired language.

Changes to Symbolic Links in HACMP/ES 4.4

HACMP/ES software packaging in all previous releases contained symbolic links to individual files. Starting with HACMP/ES Version 4.3.1 and continuing in all subsequent versions, these individual file links are omitted.

Symbolic links to the following *directories* are maintained:

- **`/usr/sbin/cluster` -> `/usr/es/sbin/cluster`**
- **`/usr/sbin/cluster/conversion` -> `/usr/es/sbin/cluster/conversion`**
- **`/usr/sbin/cluster/cspoc` -> `/usr/es/sbin/cluster/cspoc`**
- **`/usr/sbin/cluster/demos` -> `/usr/es/sbin/cluster/demos`**
- **`/usr/sbin/cluster/diag` -> `/usr/es/sbin/cluster/diag`**
- **`/usr/sbin/cluster/etc` -> `/usr/es/sbin/cluster/etc`**
- **`/usr/sbin/cluster/events` -> `/usr/es/sbin/cluster/events`**
- **`/usr/sbin/cluster/local` -> `/usr/es/sbin/cluster/local`**
- **`/usr/sbin/cluster/nims` -> `/usr/es/sbin/cluster/nims`**
- **`/usr/sbin/cluster/sbin` -> `/usr/es/sbin/cluster/sbin`**
- **`/usr/sbin/cluster/snapshots` -> `/usr/es/sbin/cluster/snapshots`**
- **`/usr/sbin/cluster/tguides` -> `/usr/es/sbin/cluster/tguides`**
- **`/usr/sbin/cluster/utilities` -> `/usr/es/sbin/cluster/utilities`**

HACMP/ES Installation Choices

Use the `smit install_selectable_all` fastpath to install the HACMP/ES software on all nodes and clients. Separate procedures for servers and clients are described in the following sections.

You must install the HACMP/ES software on each cluster node (server) and on any client machines that run the Clinfo daemon. You can install the software from an installation server, directly from the installation medium, or from a hard disk to which the software has been copied.

Installation Server

To install the HACMP/ES software in a cluster environment, you can create an HACMP/ES installation server (containing all HACMP/ES software installable images) on one node and then load the images onto the remaining cluster nodes. Creating an installation server lets you load the HACMP/ES software onto other nodes faster from the server than from other media. For instructions on creating an installation server, see the *AIX Installation Guide* or the *AIX Network Installation Management Guide and Reference*.

The organization of cluster images on the base High Availability Subsystem media allows you to make individual or multiple image selections through SMIT when installing the HACMP/ES software.

The RSCT version 1.2.00 images are prerequisites for HACMP/ES.

The installable HACMP/ES for AIX, Version 4.4 images and their associated modules on the High Availability Subsystem media include the following:

- `/usr/sys/inst.images/rsct.basic`

<code>rsct.basic.hacmp</code>	RS/6000 Cluster Technology basic functions (HACMP domains)
<code>rsct.basic.rte</code>	RS/6000 Cluster Technology basic functions (all domains)
<code>rsct.basic.sp</code>	RS/6000 Cluster Technology basic functions (SP domains)
- `/usr/sys/inst.images/rsct.clients`

<code>rsct.clients.hacmp</code>	RS/6000 Cluster Technology client function (HACMP domains)
<code>rsct.clients.rte</code>	RS/6000 Cluster Technology client function (all domains)
<code>rsct.clients.sp</code>	RS/6000 Cluster Technology client function (SP domains)
- `/usr/sys/inst.images/rsct.core.hacmp`

- /usr/sys/inst.images/cluster.es
 - cluster.es.client.lib ES Client Libraries
 - cluster.es.client.rte ES Client Runtime
 - cluster.es.client.utils ES Client Utilities
 - cluster.es.server.diag ES Server Diags
 - cluster.es.server.events ES Server Events
 - cluster.es.server.rte ES Server Runtime
 - cluster.es.server.utils ES Server Utilities
- /usr/sys/inst.images/cluster.es.cspoc
 - cluster.es.cspoc.cmds ES CSPOC Commands
 - cluster.es.cspoc.dsh ES CSPOC dsh and perl
 - cluster.es.cspoc.rte ES CSPOC Runtime Commands
- /usr/sys/inst.images/cluster.adt.es
 - cluster.adt.es.client.samples.clinfo ES Client CLINFO Samples
 - cluster.adt.es.client.samples.clstat ES Client clstat Samples
 - cluster.adt.es.client.includes ES Client Include Files
 - cluster.adt.es.client.samples.libcl ES Client LIBCL Samples
 - cluster.adt.es.java.demo.monitor ES Web Based Monitor Demo
 - cluster.adt.es.client.demos ES Client Demos
 - cluster.adt.es.client.samples.demos ES Client Demos Samples
 - cluster.adt.es.server.demos ES Server Demos
 - cluster.adt.es.server.samples.demos ES Server Sample Demos
 - cluster.adt.es.server.samples.images ES Server Sample
- Images/usr/sys/inst.images/cluster.man.en_US.es
 - cluster.man.en_US.es.data ES Man Pages
- Images/usr/sys/inst.images/cluster.man.en_US.es.data
 - cluster.man.en_US.cspoc.es.data ES CSPOC Man pages
 - cluster.man.en_US.client.es.data ES Client Man pages
 - cluster.man.en_US.server.es.data ES Server Man pages

Installing the HACMP/ES Software

HACMP/ES Installation Choices

- /usr/sys/inst.images/cluster.msg.en_US.es
 - cluster.msg.en_US.es.server ES Recovery Driver Messages
 - cluster.msg.en_US.es.client ES Client Messages
 - cluster.msg.en_US.cspoc HACMP CSPOC Messages
- /usr/sys/inst.images/cluster.vsm.es
 - cluster.vsm.es ES VSM Configuration Utility
- /usr/sys/inst.images/cluster.haview
 - cluster.haview.client HACMP HAView Client
 - cluster.haview.server HACMP HAView Server
- /usr/sys/inst.images/cluster.man.en_US.haview.data
 - cluster.man.en_US.haview.es.data ES HAView Manpages
- /usr/sys/inst.images/cluster.msg.en_US.haview
 - cluster.msg.en_US.es.haview HACMP HAView Messages
- /usr/sys/inst.images/cluster.es.taskguides
 - cluster.es.taskguides.shrvolgrp ES Shr Vol Grp Task Guides

In addition, if you plan to monitor the cluster with Tivoli, install these hativoli filesets:

- cluster.hativoli.client
- cluster.hativoli.server
- cluster.msg.en_US.hativoli

The installable images on the CRM (concurrent access) installation media are:

- /usr/sys/inst.images/cluster.es.clvm ES for AIX Concurrent Access
- /usr/sys/inst.images/cluster.es.hc ES HC Daemon

The **rsct.basic.hacmp**, **rsct.client.hacmp**, **rsct.core.hacmp**, and **cluster*** images contain the HACMP/ES run-time executables and are required. The CRM is optional software and installed separately.

HAView Installation Notes

HAView requires TME 10 NetView for AIX. Install NetView before installing HAView.

The HAView fileset includes a server image and a client image. If NetView is installed using a client/server configuration, the HAView server image should be installed on the NetView server, and the client image on the NetView client. Otherwise, you can install both the HAView client and server images on the NetView server.

Note: It is recommended that you install the HAView components on nodes outside the cluster, even though you can install it on any node that has a NetView server installed. If HAView is configured on a cluster node, and that node fails, you will lose your view of the cluster configuration, and all HAView monitoring functions will be lost until another HAView node can be brought online.

For more information on using HAView to monitor a cluster, see Chapter 21, Monitoring an HACMP/ES Cluster.

Tivoli Installation Notes

If you plan to monitor your cluster through a Tivoli interface, you must have Tivoli software installed, **hativoli** filesets installed, and then perform a number of steps to make the cluster nodes known to Tivoli. The Tivoli Management Region (TMR) should be a node outside the cluster.

Complete instructions are located in Appendix D, Installing and Configuring Cluster Monitoring with Tivoli.

Installation Media

If you install the HACMP/ES software from an installation media, you must install the software directly onto each cluster node.

To install the HACMP/ES software:

1. Insert the installation medium and enter:

```
smit install_selectable_all
```
2. SMIT displays the first **Install and Update from ALL Available Software** screen. Enter the device name of the installation medium or Install Directory in the **INPUT device / directory for software** field and press Enter.

If you are unsure about the input device name or about Install Directory, press F4 to list available devices. Then select the proper drive or directory and press Enter. The correct value is entered into the **INPUT device/directory** field as the valid input device.

3. Enter field values as follows on the next screen:

SOFTWARE to install

Enter **all** to install all server and client images, or press F4 for a software listing. The **rsct.basic.rte** and **rsct.client.rte** software are prerequisites for the **cluster.es** software. On SP nodes, the PSSP Version 3 Release 2 software already includes these images. On an RS/6000, these are installed with HACMP/ES. If you press F4, a popup window appears, listing all installable software. Use the arrow keys to locate all software modules associated with the following rsct and cluster images: **rsct.basic.rte**, **rsct.client.rte**, **cluster.es**, **cluster.es.cspoc**, **cluster.adt.es**, **cluster.man.en_US.es.data**, **cluster.msg.en_US.es**, **cluster.vsm.es**, **cluster.es.haview**, **cluster.man.en_US.haview.es.data**, **cluster.msg.en_US.es.haview**, **cluster.es.taskguides**. Next press F7 to select either an image or a module. Then press Enter after making all selections. Your selections appear in this field.

OVERWRITE same or newer versions?

Leave this field set to **no**. Set it to **yes** if you are reinstalling the HACMP/ES for AIX, Version 4.3 software.

AUTOMATICALLY Install requisite software

Set this field to **no** if the prerequisite software for Version 4.4 is installed or if the **OVERWRITE same or newer versions?** field is set to **yes**; otherwise, set this field to **yes** to install required server and client software.

4. Enter values for the other fields as appropriate for your site.
5. When you are satisfied with the entries, press Enter. SMIT responds:
are you sure?
6. Press Enter again.

You are then instructed to read the HACMP/ES 4.4 **release_notes** file in the **/usr/es/lpp/cluster/doc** directory for further instructions.

Hard Disk Installation

To install the HACMP/ES for AIX, version 4.4 software from your hard disk, you must first copy the software from the installation medium to the hard disk.

Copying HACMP/ES Software to Hard Disk

Complete the following steps to copy the HACMP/ES software to your hard disk.

1. Place the HACMP/ES tape into the tape drive and enter the **smit bffcreate** fastpath to display the Copy Software to Hard Disk for Future Installation screen.

2. Enter the name of the tape drive in the **INPUT device / directory for software** field and press Enter.

If you are unsure of the input device name, press F4 to list available devices. Select the proper drive and press Enter. That value is entered into the **INPUT device/directory** field as the valid input device.

3. Press Enter to display the **Copy Software to Hard Disk for Future Installation** screen.
4. Enter field values as follows:

SOFTWARE name Enter **cluster*** or **all** to copy all server and client images, or press F4 for a software listing.

DIRECTORY for storing software Change the value to the storage directory accessed by all nodes using HACMP/ES.

5. Enter values for the other fields as appropriate for your site.
6. When you are satisfied with the entries, press Enter. SMIT responds: **are you sure?**
7. Press Enter again.

Once the HACMP/ES software has been copied to your system, install the software by following the instructions given in the following section. After you copy the software, read the HACMP/ES release notes in the **/usr/es/lpp/cluster/doc** directory.

Installing HACMP/ES from Hard Disk (SP only)

You must install the HACMP/ES software on all nodes of the SP system that will be used to form an HACMP/ES cluster, as well as HACMP/ES client code on any node or control workstation that will be used to monitor the cluster.

Use the SP **dsh** command to speed installation of the LPP filesets (images).

Complete the following steps:

1. Create a file called **/HACMPHOSTS** that contains host names of nodes in the SP frame that will have the HACMP/ES for AIX software installed.
2. Export the Working Collective (WCOLL) environment variable using the following command:

```
export WCOLL=/HACMPHOSTS
```

3. Ensure that all hosts listed in the **/HACMPHOSTS** file are up (host responds) using the following command:

```
/usr/lpp/spp/bin/SDRGetObjects host_responds
```

where **SDRGetObjects** refers to the SP database. Host responds should indicate a 1 for all nodes which respond.

4. Mount the file system (from the control workstation) onto all nodes; enter:

```
dsh /etc/mount CNAME:STORAGE_DIRECTORY /mnt
```

where **CNAME** is the hostname of the control workstation and **STORAGE_DIRECTORY** is the directory where the software is stored, as listed in Step 4 of the previous section.

5. Issue the following command to install the HACMP/ES for AIX software on the nodes; enter:

```
dsh "/etc/installlp -Xagd /mnt <LPP_NAME>"
```

where *LPP_NAME* is the name of the package/fileset you need to install. This must be done for each fileset you need to install.

Note: You can install all packages and filesets by entering **cluster*** as the *LPP_NAME* value.

6. To verify that HACMP/ES was successfully installed on each node, enter:

```
dsh "/etc/installlp -s | grep cluster"
```

Additionally, for SP nodes installed through the SP install process, you can use that process or the customize process to install HACMP/ES and PTFs. See the *IBM Parallel System Support Programs for AIX: Installation and Migration Guide*.

HACMP/ES files are installed in the **/usr/es/sbin/cluster** and **/usr/es/lpp/cluster** directories.

Installing the Concurrent Resource Manager

To install the concurrent access feature on cluster nodes, complete the procedure in this section.

Note: In HACMP/ES Version 4.4 environments, concurrent access is available using only an IBM 7135-110 or 210 Disk Array, an IBM 7137 Disk Array, an IBM 2105 Versatile Storage Server (Models B09 and 100), or an IBM 7133 SSA disk subsystem. RAID devices from other manufacturers may not support concurrent access.

Use the **smit install_selectable_all** fastpath to load the concurrent access install image on a node. See the section HACMP/ES Installation Choices on page 14-4 in this chapter for a list of software images to install. Depending on the AIX level installed on your system, not all images are required.

To install the concurrent access software on a server:

1. Insert the installation media and enter:

```
smit install_selectable_all
```
2. Enter the device name of the installation media or Install Directory in the **INPUT device / directory for software** field and press Enter.
If you are unsure about the input device name or about Install Directory, press F4 to list available devices. Then select the proper media or directory and press Enter. The correct value is entered into the **INPUT device/directory** field as the valid input device.
3. Press Enter. SMIT refreshes the screen.
4. Enter field values as follows:
SOFTWARE to install Change the value in this field to include **cluster.es.clvm** and **cluster.es.hc**. Note that the run-time executables for the HACMP/ES software and associated images are automatically installed when you select these images.

OVERWRITE same or newer versions? Leave this field set to **no**. Set it to **yes** only if you are reinstalling or reverting to Version 4.4 from a newer version of the HACMP/ES software.

AUTOMATICALLY Install requisite software Set this field to **no** if the prerequisite software for Version 4.4 is installed or if the **OVERWRITE same or newer versions?** field is set to **yes**; otherwise, set this field to **yes** to install required software.

5. Enter values for other fields appropriate for your site.
6. Press Enter when you are satisfied with the entries.

SMIT responds: **are you sure?**

7. Press Enter again.

You are then instructed to read the HACMP 4.4 **release_notes** file in the **/usr/lpp/cluster/doc** directory for further information.

Problems During the Installation

If you experience problems during the installation, the installation program automatically performs a cleanup process. If, for some reason, the cleanup is not performed after an unsuccessful installation:

1. Enter the **smit install** software to display the Installation and Maintenance menu.
2. Select **Install and Update Software**.
3. Select **Clean Up After a Interrupted Installation**.
4. Review the SMIT output (or examine the **/smit.log** file) for the interruption's cause.
5. Fix any problems and repeat the installation process.

Processor ID Licensing Issues

The Concurrent Resource Manager is licensed to the processor ID of a cluster node. Many of the **clvm** or concurrent access commands validate the processor ID against the license file. A mismatch will cause the command to fail, with an error message indicating the lack of a license.

Restoring a system image from a **mksysb** tape created on a different node or replacing the planar board on a node will cause this problem. In such cases, you must recreate the license file by removing and reinstalling the **cluster.clvm** component of the current release from the original installation images.

Verifying Cluster Software

After the install, use the AIX command **lppchk**, and check the installed directories to see that expected files are present.

Rebooting Cluster Nodes and Clients

The final step in installing the HACMP/ES software is to reboot each cluster node and client in your HACMP/ES environment.

Converting from HACMP for AIX to HACMP/ES 4.4

Consider the following choices:

- *Converting using a snapshot.* For clusters running a version of HACMP for AIX prior to version 4.4: Save the HACMP for AIX configuration in a snapshot, remove HACMP for AIX, install HACMP/ES 4.4, and apply the snapshot to the HACMP/ES cluster
 - *If your HACMP for AIX software is version 4.1 or greater,* you can save your HACMP for AIX cluster configuration in a snapshot before removing HACMP for AIX, and then reapply the snapshot to the HACMP/ES 4.4 software.
 - *If your HACMP for AIX software is pre-version 4.1,* you can upgrade your cluster to HACMP Version 4.1 or greater, then follow the procedure outlined above.
- *Converting without a snapshot.* Save the planning worksheets and configuration files from your current configuration as reference if you want to configure the HACMP/ES cluster the same as it was in HACMP. Simply remove HACMP, install HACMP/ES, and reconfigure the cluster following the saved information.

Step 1: Check on Committed Software

To see if a version of the HACMP/6000 or HACMP for AIX software exists and if your cluster configuration is committed, enter the following command:

```
lslpp -h "cluster*"
```

If the HACMP software exists, but is earlier than version 4.1, you must upgrade to this level before continuing with a conversion using the snapshot utility.

Step 2: Save HACMP for AIX Cluster Configuration in a Snapshot

To save your HACMP for AIX (Version 4.1 or greater) cluster configuration, create a snapshot in HACMP for AIX. If you have customized event scripts, save them also. See your *HACMP for AIX Administration Guide* for information on this process.

Warning: Do not save your cluster configuration or customized event scripts under the directory path **/usr/sbin/cluster**, **/usr/es/sbin/cluster** or **/usr/lpp/cluster**. These directories are deleted and recreated during the installation of the HACMP/ES software, and all configuration and data is lost.

Step 3: Remove the HACMP for AIX Software

To remove the HACMP software and your cluster configuration on cluster nodes and clients:

1. Enter the following command:

```
smit install_remove
```

The **Install/Remove** screen appears.

2. Enter field values as follows:

SOFTWARE name Enter cluster* to remove all server and client software, or press F4 for a popup window listing all installed software. Use the arrow keys to locate all software you want to remove; then press F7 to select it.

Press Enter after making all selections. Your selections appear in this field.

REMOVE dependent software? Enter no.

EXTEND file systems if space needed? Enter yes.

DETAILED output? Enter no.

Step 4: Install HACMP/ES 4.4

Follow instructions starting with section HACMP/ES Installation Choices on page 14-4. *Before verifying the software and rebooting*, continue with Step 5 to install the saved snapshot.

Step 5: Install the Saved Snapshot

After installing HACMP/ES 4.4 on cluster nodes, you now install the snapshot you saved. This must be done before rebooting the cluster nodes and clients.

Use the following command, where *version* is your HACMP for AIX version number and *snapshotfile* is the name of your snapshot file. The **-C** flag converts an HACMP for AIX snapshot to an HACMP/ES snapshot format:

```
clconvert_snapshot -C -v version -s snapshotfile
```

For example, if you are converting from HACMP for AIX version 4.2 and call the file *my42snapshot*:

```
clconvert_snapshot -C -v 4.2 -s my42snapshot
```

Apply the snapshot. See Chapter 26, Saving and Restoring Cluster Configurations, for more information.

Step 6: Reinstall Saved Customized Event Scripts

Step 7: Verify the Configuration

See section Verifying Cluster Software on page 14-12.

Step 8: Reboot Cluster Nodes and Clients

Node-by-Node Migration from HACMP for AIX Version 4.4 to HACMP/ES Version 4.4

Node-by-node migration lets you migrate from a running HACMP 4.4 cluster to a running HACMP/ES 4.4 cluster *without bringing the entire cluster offline at once*, thereby keeping all cluster resources highly available during the migration process.

Prerequisites for Node-by-Node Migration

In order to perform node-by-node migration:

- All nodes in the cluster must have HACMP version 4.4 installed and committed.
- You cannot have HAGEo for AIX installed on the cluster. Node-by-node migration functions only for HACMP 4.4 to HACMP/ES 4.4 migrations.
- All nodes in the cluster must be up and running the HACMP 4.4 software.
- The cluster must be in a stable state.
- You must have enough disk space to hold both HACMP and HACMP/ES software during the migration process (approximately 100MB in the `/usr` directory and 950 KB in the `/` (root) directory). When the migration is complete, the space requirements are reduced to the normal amount necessary for HACMP/ES 4.4 alone.
 - If you are not planning to install optional software you can plan for less space: HAView=14MB, TaskGuide=400KB, hativoli=400KB, VSM=5.2MB. You should also choose to install the message catalogs for the language you will be using, rather than all message catalogs (Japanese message catalogs use 1.6MB).
- Nodes must have enough memory to run both HACMP and HACMP/ES daemons simultaneously. This is a minimum of 64MB of RAM. 128MB of RAM is recommended.
- If any nodes in the cluster are currently set to start cluster services automatically on reboot, change this setting before beginning the migration process.

Setting System to Stop Cluster Services on System Restart

Using C-SPOC:

- Enter `smit cl_admin`
- Select **HACMP for AIX Cluster Services > Stop Cluster Services**
- For the field “Stop now, on system restart, or both?” Enter **restart**. This removes the HACMP cluster services entry in the `/etc/inittab` file, so that cluster services do *not* restart after a reboot.
- Press Enter. The setting is changed for all cluster nodes.

If you do not use C-SPOC, you must change the setting on each cluster node individually, by entering the command `smit clstop` and then changing the setting as described above.

- If you have an HACMP cluster with more than three nodes, with serial networks configured so that multiple networks connect to one node (one standby node for multiple nodes type configuration), you must change this to a daisy chain or point-to-point configuration before migrating to HACMP/ES. No HACMP/ES node can connect to more than two other nodes.

Note: As in any migration, do not attempt to make any changes to the cluster topology or configuration once you have started the migration process.

How to Perform a Node-by-Node Migration

Take the following steps for a node-by-node migration from HACMP 4.4 to HACMP/ES 4.4:

1. Save the current configuration in a snapshot as a precautionary measure. Place it in a safe directory (one that is not touched by installation procedures. Do *not* use `/usr/sbin/cluster`, for example).
2. Stop cluster services on one of the nodes running HACMP 4.4 using the “graceful with takeover” method.

The command line method:

```
clstop -gr
```

The SMIT method:

On the HACMP SMIT **Stop Cluster Services** screen, select “graceful with takeover.”

Check that the cluster services are stopped on the node and that its cluster resources have been transferred to takeover nodes before proceeding.

3. Install HACMP/ES 4.4 on the node. See instructions in the section HACMP/ES Installation Choices on page 14-4.

The HACMP/ES installation utility first checks the current version of HACMP. If it detects that your HACMP software is an earlier version than 4.4, you will see an error message and the installation will be aborted. If the node’s HACMP cluster services are not stopped, the operation will also abort. If you have version 4.4 installed and properly stopped, you will see a message indicating that a migration has been requested and is about to proceed.

The **smit.log** records the migration messages during the install process.

4. After installing the HACMP/ES software, **reboot the node**.
5. Using the HACMP SMIT **Start Cluster Services** screen, restart the HACMP software.

When you restart Cluster Services:

- The HACMP/ES software is also started
- HACMP cluster services run on the node; it rejoins the cluster
- The node reacquires the cascading resources for which it is the primary node (depending on how you have Inactive Takeover set).

Installing the HACMP/ES Software

Node-by-Node Migration from HACMP for AIX Version 4.4 to HACMP/ES Version 4.4

Both HACMP and HACMP/ES are now running on the node, but only HACMP controls the cluster events and resources. If you list daemons in SRC, you will see the following daemons listed on this hybrid node:

HACMP	HACMP/ES	RSCT
clstrmgr	clstrmgrES	grpsvcs
clsmuxpd	cllockdES (optional)	topsvcs
clinfo (optional)		emsvcs
		grpplsm
		emaixos

6. Repeat steps 2 through 5 for all other nodes in the cluster.

Warning: Starting cluster services on the last node is a point of no return. Once you have restarted HACMP (and thus HACMP/ES) on the last node, and the migration has commenced, *you cannot reverse the migration*. If you wish to return to the HACMP configuration after this point, you will have to reinstall the HACMP software and apply the saved snapshot. *Up to this point*, you can backout of the installation of HACMP/ES and return to your running HACMP cluster. In case you need to do this, see the section Backout Procedure on page 14-19. See also the section Notes About the Migration Process on page 14-17 for more detailed information on how the software handles the node-by-node migration process, and the precautions taken to facilitate the backout (recovery) procedure.

During this install-migration process, when you restart each node, the node is running both products, with the HACMP **clstrmgr** in control of handling cluster events and the **clstrmgrES** in “passive” mode.

After you start cluster services on the last node, the migration to HACMP/ES proceeds automatically. Total control of the cluster is transferred automatically from HACMP to HACMP/ES.

Messages documenting the migration process are logged to the **/hacmp.out** file as well as to the HACMP **cm.log** and to the HACMP/ES **clstrmgr.debug** log files.

Note: If you redirected the **hacmp.out** log file to a different directory than the default **/tmp** directory, messages will be logged to your chosen directory.

When the migration is complete, all cluster nodes are up and running HACMP/ES, and HACMP is de-installed.

config_too_long Message Appears

When the migration process has completed and the HACMP filesets are being deinstalled, you may see the message “config_too_long.”

This message appears when the cluster manager detects that an event has been processing for more than the six minutes allowed by default. config_too_long messages will continue to be appended to the **hacmp.out** log every 30 seconds until the event completes. If you observe these messages, you should periodically check that the event is indeed still running and has not failed.

You can avoid these messages by increasing the time to wait before calling config_too_long, using the following command:

```
chssys -s clstrmgr -a "-u milliseconds_to_wait"
```

For example:

```
chssys -s clstrmgr -a "-u 60000"
```

sets the time to 600 seconds, or 10 minutes, instead of the default six minutes.

Note: If you do change the time, you should change it back to the default time when the migration is complete.

Notes About the Migration Process

When you have installed HACMP/ES on all cluster nodes (all nodes are now in a “hybrid” state), starting Cluster Services on the last cluster node automatically triggers the transfer of control from HACMP to HACMP/ES as follows:

- Installing HACMP/ES installs a recovery file called **firstboot** in a holding directory on the cluster node, and creates a migration file (**.mig**) to be used as a flag during the migration process.
 - The HACMP/ES recovery driver sends a message to the HACMP clstrmgr telling it to run the “waiting” and “waiting_complete” events.
 - HACMP/ES Group Services verify cluster stability and membership
 - The **firstboot** file on each cluster node is moved to an active directory (**/etc**)
 - The migration flag (**.mig** file) created during installation is transferred from the HACMP/ES 4.4 directory to the HACMP 4.4 directory on all nodes
- When the **firstboot** file is moved to the active directory and the **.mig** file transfer is complete on all nodes, transfer of control to HACMP/ES continues with the HACMP/ES “migrate” event.
- HACMP/ES recovery driver issues the “migrate” event
 - HACMP/ES stops the HACMP daemons using the “forced” option.
 - HACMP/ES clinfoES and clsmuxpdES daemons are all activated, reusing the ports previously used by the HACMP versions of those daemons

Installing the HACMP/ES Software

Node-by-Node Migration from HACMP for AIX Version 4.4 to HACMP/ES Version 4.4

- HACMP/ES recovery driver runs the `migrate_complete` event
 - HACMP is deinstalled; configuration files common to both products are left untouched.
 - Base directories are relinked
 - `/etc/firstboot` files are removed
 - the migration flag (`.mig` file) in the HACMP `/usr/sbin/cluster` directory is removed.
- Migration is now complete.

You should verify and test the cluster's proper fallover and recovery functionality

Cluster Snapshots Saved During Migration

Pre-existing HACMP snapshots are saved in the `/usr/sbin/cluster/snapshots` directory.

Handling Node Failure During the Migration Process

If a node fails during the migration process *after* its `firstboot` file has moved to an active directory, it will complete the migration process during node reboot. However, the failed node may have an out-of-synch HACMP ODM when it reintegrates into the cluster. You must fix this manually by synchronizing the topology and resources of the cluster prior to reintegrating the failed node back into the cluster.

Take these steps to reintegrate and synchronize the failed node:

1. Wait until you are sure all other nodes have completed the migration process. (Check that the `migrate_complete` event has run on all other nodes.)
2. If it has not already rebooted, reboot the failed node.
On rebooting, the node will complete the migration process, but its HACMP ODM may be out-of-synch.
3. On an another active node that has successfully migrated, enter
`smitty hacmp`
4. Select **Cluster Configuration > Cluster Topology > Synchronize Cluster Topology**
5. Press F3 to return to the main HACMP menu.
6. Select **Cluster Configuration > Cluster Resources > Synchronize Cluster Resources**
7. Press F10 to exit SMIT.

This process updates the out-of-synch ODM on the node that failed during the original migration process.

Backout Procedure

If for some reason you decide not to complete the migration process, you can deinstall the HACMP/ES software on the nodes where you have installed it at any point in the process *before* starting HACMP on the last node.

1. On each node in turn (one at a time), stop cluster services.

The command line method:

```
clstop -gr
```

The SMIT method:

On the HACMP SMIT **Stop Cluster Services** screen, select “graceful with takeover.”

Check that the cluster services are stopped on the node and that its cluster resources have been transferred to takeover nodes before proceeding.

2. When you are sure the resources on the node have been properly transferred to a takeover node, remove the HACMP/ES software.

Enter the following command:

```
smit install_remove
```

The **Install/Remove** screen appears.

3. Enter field values as follows:

SOFTWARE name Press F4 for a popup window listing all installed software. Use the arrow keys to locate all the ES software; then press F7 to select it.

Be sure *NOT* to remove the manpages or the C-SPOC messages software; these are shared with the HACMP software:

Press Enter after making all selections. Your selections appear in this field.

REMOVE dependent software? Enter **yes**.

EXTEND file systems if space needed? Enter **yes**.

DETAILED output? Enter **no**.

4. Start HACMP on this node. When you are certain the resources have transferred properly (if necessary) back to this node, repeat these steps on the next node.
5. Continue this process until HACMP/ES has been removed from all nodes in the cluster.

Installing the HACMP/ES Software

Node-by-Node Migration from HACMP for AIX Version 4.4 to HACMP/ES Version 4.4

Chapter 15 Upgrading an HACMP/ES Cluster

This chapter provides instructions for upgrading an HACMP/ES cluster configuration.

Preparing for an Upgrade

This section identifies the tasks you must perform to prepare for an upgrade to HACMP/ES Version 4.4:

- Read the chapters in Part 1, Planning HACMP/ES Clusters, before starting an upgrade to the Version 4.4 software. These chapters cover the worksheets and diagrams necessary to plan an HACMP/ES installation and configuration.

Note: If you have not completed these worksheets and diagrams, do so before continuing. You should transfer all information about your existing installation, plus any changes you plan to make after the upgrade, to these new worksheets.

- Ensure that each cluster node has its own HACMP/ES license.
- Ensure that adapter SCSI IDs for the shared disk are not the same on each node. During an operating system upgrade, disk adapters are reset to the default value of seven. This setting can cause a SCSI ID conflict on the bus that prevents proper access to the shared disk.
- The base HACMP/ES software uses 44MB of disk space. Ensure that the `/usr` filesystem has the necessary free disk space for the upgrade. If you are installing only the HACMP/ES for AIX software for clients, a minimum of 3MB is recommended.
 - If planning to install these optional images add the following amounts of space: HAView =14MB, TaskGuide= 400KB, hativoli=400 KB, VSM= 5.2MB. You should also choose to install the message catalogs for the language you will be using, rather than all message catalogs (Japanese message catalogs use 1.6MB).
- Perform the installation process as the root user.

Before upgrading to HACMP/ES, Version 4.4:

1. Archive any localized script and configuration files to prevent losing them during an upgrade.
2. Commit the installation (if it is applied but not committed) so that the HACMP/ES software can be installed over the existing version. To see if your configuration is already committed, enter:

```
lslpp -h cluster.*
```

If the word “COMMIT” is displayed under the Action header, continue to the next step. If not, run the **smit install_commit** utility before installing the Version 4.4 software. SMIT displays the **Commit Applied Software Updates (Remove Saved Files)** screen.

Enter field values as follows:

SOFTWARE name Change this value to **cluster***

COMMIT old version if above version used it? Set this field to **yes**.

EXTEND filesystem if space needed? Set this field to **yes**.

3. Make a **mksysb** backup of each node's configuration.

Note: If for some reason you do restore a **mksysb** back onto your system, you will need to reset the SCSI IDs on the system.

4. Save any customized event information.

Upgrading your HACMP/ES Cluster

This section identifies steps required to upgrade an HACMP/ES cluster from version 4.2.1, 4.2.2, 4.3.0, or 4.3.1 to Version 4.4.

Note: When upgrading an HACMP/ES cluster, you cannot leave the cluster at mixed versions for long periods of time and still guarantee high availability.

Note: In version 4.4, the directory **/usr/sbin/cluster** and subdirectories have symbolic links to the **/usr/es/sbin/cluster** directory and subdirectories. Files in those directories are *not* linked, as they were in releases prior to 4.3.1.

Complete the following steps to upgrade your cluster:

Step 1: Upgrade the operating system to AIX version 4.3.3

Complete the migration installation as described in your *AIX Installation Guide*.

Step 2: Upgrade to PSSP Version 3 Release 2 (applies to SP nodes only)

Complete the installation as described in your *PSSP Installation Guide*.

Step 3: Upgrade Your Existing HACMP/ES Software Version

In this step, you upgrade your existing HACMP/ES software from version 4.2.1, 4.2.2, 4.3.0, or 4.3.1 to version 4.4, following the instructions in this chapter.

Step 4: (optional) Adding to or Changing the Cluster

In this step, you make further changes or additions to the system as defined on the planning worksheets. For example, you may want to include newly added nodes in the resource chain for one or more resource groups.

You can also modify previous version cluster snapshot files so they can be applied to a Version 4.4 cluster.

Step 5: Verifying Cluster Configuration

In this step, you use the `/usr/es/sbin/cluster/diag/clverify` utility to verify that the cluster is configured properly. Chapter 25, Verifying a Cluster Configuration, describes how to use this utility.

Step 6: Rebooting Cluster Nodes and Clients

The upgrade is complete when you have performed the steps identified above. You must then reboot the system to make the cluster topology active.

Step 7: Upgrading Applications Using Clinfo API

Check the *HACMP for AIX Programming Client Applications* guide for updated information on the Clinfo C and C++ API routines. Make any necessary changes, then recompile and link your applications.

Installing HACMP/ES Software

You must install the HACMP/ES Version 4.4 software on all cluster nodes and clients. To replace your current version's ODM object classes with the Version 4.4 ODM object classes, you must perform an upgrade when installing the Version 4.4 base High Availability Subsystem software. Chapter 14, Installing the HACMP/ES Software, provides separate procedures for installing the HACMP/ES software.

Upgrading Existing HACMP/ES Software to Version 4.4

If your site is currently running an earlier version of the HACMP/ES software in its cluster environment, perform an **install-overwrite** to upgrade your cluster configuration to Version 4.4.

1. Upgrade to AIX 4.3.3 and PSSP Version 3 release 2 on all SP nodes and on the control workstation.
2. Commit your current HACMP/ES software on all nodes.
3. Stop HACMP/ES on one node (gracefully with takeover) using the **clstop** command. For this example, stop HACMP/ES on Node A. Node B will takeover Node A's resources and make them available to clients.

See Chapter 20, Starting and Stopping Cluster Services, for more information on using the **clstop** command.

4. After the resources have moved successfully from Node A to the takeover node, install the new HACMP/ES software. See Chapter 14, Installing the HACMP/ES Software, starting at the section HACMP/ES Installation Choices, for instructions on installing the HACMP/ES Version 4.4 software.

The **cl_convert** utility runs as part of the **Install with Overwrite** procedure and automatically updates the HACMP ODM object classes to the 4.4 version. However, if installation fails, you must run **cl_convert** from the command line. For more information on **cl_convert**, see the section entitled *HACMP for AIX Common Commands on page A-5* of Appendix A in the *HACMP for AIX Administration Guide* or the **cl_convert** man page.

Upgrading an HACMP/ES Cluster

Upgrading Existing HACMP/ES Software to Version 4.4

Note: If IP address swapping is being used on this node, that is, a boot address is defined for this node, check to ensure that the HACMP changes to `/etc/inittab` and `/etc/rc.net` exist (as specified in Chapter 19, Maintaining an HACMP/ES Cluster) before rebooting the node.

5. Reboot Node A.
6. Start the HACMP/ES software on Node A using the `smit clstart` fastpath and verify that Node A successfully joins the cluster.
7. Repeat Steps 2 through 6 on remaining cluster nodes, one at a time.
8. Synchronize the node configuration and the cluster topology from Node A to all nodes.

Important: When upgrading a cluster to Version 4.4, never synchronize the cluster definition from an upgraded Version 4.4 node when a node that has not been upgraded remains in a mixed-version cluster. The `cl_convert` utility assigns node IDs that are consistent across all nodes in the cluster. These new IDs could conflict with those already assigned from the Version 4.4 cluster, even though the ODMs on the nodes may look correct.
9. Check that the `tty` device is configured as a serial network using the `smit mktty` fastpath.
10. If you are using SCSI shared disks, check that the SCSI ID of the shared disk adapter is unique and is not equal to 7.

A SCSI ID conflict can occur if SCSI ID 7 is in use by the shared adapter when the HACMP/ES cluster is restarted.
11. Check that all external disks are available on Node A and that the `lspv` command shows PVIDs for each disk.

If PVIDs are not displayed for the disks, you may need to remove the disk and reconfigure them.
12. Verify the cluster topology on all nodes using the `clverify` utility.
13. Restore the HACMPevent ODM object class to save any pre- and post-events you have configured for your cluster. See the section Making Additional Changes to the Cluster on page 15-7.
14. (optional) Upgrade the Concurrent Resource Manager software.
15. Go to the section Making Additional Changes to the Cluster on page 15-7.
16. Complete a test phase on the cluster before putting it into production.

Supported Upgrades of HACMP/ ES

If you wish to convert to HACMP/ES version 4.4 from versions earlier than those listed below, you must first do an installation upgrade to one of the following supported versions. Since a conversion from the versions below to 4.4 are supported upgrades, you will then be able to convert to HACMP/ES 4.4. For example, to convert from HACMP/ES 4.2.1 to HACMP/ES 4.4 you must first do an installation upgrade to HACMP/ES 4.2.2. (Refer to the *Enhanced Scalability Installation and Administration Guide*, Version 4.2.2 for specific information on this upgrade.) You will then be able to migrate to HACMP/ES 4.4

HACMP conversion utilities provide easy conversion between the versions and products listed below:

- HACMP 4.2.2 to HACMP 4.4
- HACMP 4.3.1 to HACMP 4.4
- HACMP/ES 4.2.2 to HACMP/ES 4.4
- HACMP/ES 4.3.1 to HACMP/ES 4.4
- HACMP 4.4 to HACMP/ES 4.4
- HANFS 4.3.1 to HACMP 4.4

cl_convert and clconvert_snapshot

The HACMP conversion utilities are **cl_convert** and **clconvert_snapshot**.

Upgrading HACMP/ES software to the newest version involves converting the ODM from a previous release to that of the current release. When you install HACMP/ES, **cl_convert** is run automatically. However, if installation fails, you must run **cl_convert** from the command line.

Note: See the *HACMP for AIX Administration Guide* Appendix A, HACMP for AIX Commands, for command line syntax. In a failed conversion, be sure to run **cl_convert** using the **-F** flag.

The **clconvert_snapshot** utility is not run automatically during installation, and must always be run from the command line. Run **clconvert_snapshot** to upgrade cluster snapshots.

The **cl_convert** utility logs conversion progress to the **/tmp/clconvert.log** file so that you can gauge conversion success. This log file is regenerated each time **cl_convert** or **clconvert_snapshot** is executed.

Note: Root user privilege is required to run a conversion utility. You must know the HACMP/ES version from which you are converting in order to run these utilities.

For more information on **cl_convert** and **clconvert_snapshot**, refer to the respective man pages, or to the above mentioned Appendix A, HACMP for AIX Commands.

Upgrading the Concurrent Resource Manager

To install the concurrent access feature on cluster nodes, complete the procedure in this section.

Note: In HACMP/ES Version 4.4 environments, concurrent access is available using only an IBM 7135-110 or 210 Disk Array, an IBM 7137 Disk Array, an IBM 2105 Versatile Storage Server (Models B09 and 100), or an IBM 7133 SSA disk subsystem. RAID devices from other manufacturers may not support concurrent access. Use the **smitty install_selectable_all** fastpath to load the concurrent access install image on a node. See the section for a list of software images to install. Depending on the AIX level installed on your system, not all images are required.

To install the concurrent access software on a server:

1. Insert the installation media and enter:

```
smit install_selectable_all
```

2. Enter the device name of the installation media or Install Directory in the **INPUT device / directory for software** field and press Enter.

If you are unsure about the input device name or about Install Directory, press F4 to list available devices. Then select the proper media or directory and press Enter. The correct value is entered into the **INPUT device/directory** field as the valid input device.

3. Press Enter. SMIT refreshes the screen.

4. Enter field values as follows:

SOFTWARE to install Change the value in this field to include **cluster.es.clvm** and **cluster.es.hc**. Note that the run-time executables for the HACMP/ES software and associated images are automatically installed when you select these images.

OVERWRITE same or newer versions? Leave this field set to **no**. Set it to **yes** only if you are reinstalling or reverting to Version 4.4 from a newer version of the HACMP/ES software.

AUTOMATICALLY Install requisite software Set this field to **no** if the prerequisite software for Version 4.4 is installed or if the **OVERWRITE same or newer versions?** field is set to **yes**; otherwise, set this field to **yes** to install required software.

5. Enter values for other fields appropriate for your site.

6. Press Enter when you are satisfied with the entries.

SMIT responds: **are you sure?**

7. Press Enter again.

You are then instructed to read the HACMP/ES 4.4 **release_notes** file in the **/usr/es/lpp/cluster/doc** directory for further information.

Problems During the Installation

If you experience problems during the installation, the installation program automatically performs a cleanup process. If, for some reason, the cleanup is not performed after an unsuccessful installation:

1. Enter the **smit install** software to display the Installation and Maintenance menu.
2. Select **Install and Update Software**.
3. Select **Clean Up After a Interrupted Installation**.
4. Review the SMIT output (or examine the **/smit.log** file) for the interruption's cause.
5. Fix any problems and repeat the installation process.

Making Additional Changes to the Cluster

Make any further changes or additions to the system as planned for in the worksheets. For example, you may want to include newly added nodes in the resource chain for one or more resource groups. Consult the appropriate chapters in Part 3, “Administering an HACMP/ES Cluster,” for those changes.

Modifying Cluster Snapshots

After you have upgraded your HACMP/ES software version to Version 4.4, you may want to restore one or more of the previous version cluster snapshots you created using the Cluster Snapshot utility. The default directory path for storage and retrieval of a snapshot is **/usr/es/sbin/cluster/snapshots**; however, you may have specified an alternate path using the `SNAPSHOTPATH` environment variable. Check in these locations before using the **/usr/es/sbin/cluster/diag/clconvert_snapshot** utility to convert the snapshot for use in a Version 4.4 cluster environment.

The **clconvert_snapshot** utility updates HACMP ODM classes with new fields for Version 4.4, initializes these classes to either zeros or empty strings, and adds new events in the HACMPevent ODM.

To convert and apply a cluster snapshot, enter:

```
clconvert_snapshot -v version -s <snapshot file name>
```

where the **-s** flag is used with the snapshot file name you want to update and apply and the **-v** flag is used with the version of the saved snapshot.

See Chapter 26, Saving and Restoring Cluster Configurations, for more information about creating, restoring, and applying cluster snapshots in a cluster environment.

Upgrading Clinfo Applications to HACMP/ES 4.4

Check the *HACMP for AIX Programming Client Applications* guide for updated information on the Clinfo C and C++ API routines. Make any changes necessary to your applications; then recompile and link the applications using the Clinfo library.

Chapter 16 Configuring Clinfo Scripts and Files

This chapter describes how to edit Clinfo-related files and scripts.

Note: In version 4.4, the directory `/usr/sbin/cluster` and subdirectories have symbolic links to the `/usr/es/sbin/cluster` directory and subdirectories. Files in those directories are *not* linked, as they were in releases prior to 4.3.1.

Editing the /usr/es/sbin/cluster/etc/clhosts File

For the Clinfo daemon (**clinfo**) to get the information it needs, you must edit the `/usr/es/sbin/cluster/etc/clhosts` file on each cluster node. This file should contain hostnames (addresses) of any HACMP/ES nodes with which **clinfo** can communicate, including servers from clusters accessible through logical connections.

Note: If a client is located in a network that has both HACMP and HACMP/ES clusters, the following rules apply to the **clhosts** file:

- If the client has HACMP software installed, the **clhosts** file must be configured with only HACMP hostnames. If the file contains any HACMP/ES hostnames, **clinfo** will abort.
- If the client has HACMP/ES software installed, the **clhosts** file may contain both HACMP and HACMP/ES hostnames.

As installed, the `/usr/es/sbin/cluster/etc/clhosts` file on an HACMP/ES node contains a loopback address. The **clinfo** daemon first attempts to communicate with a **clsmuxpd** process locally. If it succeeds, **clinfo** then acquires an entire cluster map, including a list of all HACMP/ES server interface addresses. From then on, **clinfo** uses this list rather than the provided loopback address to recover from a **clsmuxpd** communication failure.

If **clinfo** does not succeed in communicating with a **clsmuxpd** process locally, however, it can only continue trying to communicate with the local address. For this reason, you should replace the loopback address with all HACMP/ES service addresses accessible through logical connections to this node. The loopback address is provided only as a convenience.

An example `/usr/es/sbin/cluster/etc/clhosts` file follows:

```
n0_c183      # n0 service
n2_c183      # n2 service
n3_c183      # n3 service
```

Warning: Do not include standby addresses in the **clhosts** file, and do not leave this file empty. If either of these two conditions exist, neither **clinfo** nor the `/usr/es/sbin/cluster/clstat` utility will work properly.

Editing the `/usr/es/sbin/cluster/etc/clinfo.rc` Script

The `/usr/es/sbin/cluster/etc/clinfo.rc` script, executed whenever a cluster event occurs, updates the system's ARP cache. If you are not using the hardware address swapping facility, a copy of the `/usr/es/sbin/cluster/etc/clinfo.rc` script must be present on each node and client in the cluster for all ARP caches to be updated and synchronized.

The HACMP/ES software is distributed with a template version of the `/usr/es/sbin/cluster/etc/clinfo.rc` script. You can use the script as distributed, you can add new functionality to the script, or you can replace it with a custom script.

Note: If you are not using hardware address swapping, the ARP functionality must remain.

The format of the `clinfo` call to `clinfo.rc`:

```
clinfo.rc {join, fail, swap} interface_name
```

When `clinfo` gets a `cluster_stable` event, or when it connects to a new `clsmuxpd`, it receives a new map. `clinfo` then checks for changed states of interfaces.

- If a new state is UP, `clinfo` calls `clinfo.rc join interface_name`.
- If a new state is DOWN, it calls `clinfo.rc fail interface_name`.
- If `clinfo` receives a `node_down_complete` event, it calls `clinfo.rc` with the fail parameter for each interface currently UP.
- If `clinfo` receives a `fail_network_complete` event, it calls `clinfo.rc` with the fail parameter for all associated interfaces.
- If `clinfo` receives a `swap_complete` event, it calls `clinfo.rc swap interface_name`.

Chapter 17 Installing and Configuring Clients

This chapter describes how to install the HACMP/ES for AIX, Version 4.4 Licensed Program Product (LPP) on clients, and how to configure clients for an HACMP/ES cluster.

Prerequisites

- Read Chapter 9, Planning for HACMP/ES Clients.
- Read the Release Notes for HACMP/ES, version 4.4 in `/usr/es/lpp/cluster/doc/release_notes` for additional information on installing the HACMP/ES software.
- Install the HACMP/ES, Version 4.4 Licensed Program Product (LPP) on the cluster nodes. (See Chapter 14, Installing the HACMP/ES Software.)

Note: The directory `/usr/sbin/cluster` and subdirectories have symbolic links to the `/usr/es/sbin/cluster` directory and subdirectories. Files in those directories are *not* linked, as they were in releases prior to 4.3.1.

Overview

Installing the HACMP/ES software on each client that runs the Clinfo daemon enables the clients to receive messages about events and actions taken by the high availability software running in the cluster. The client can take predefined, automatic steps in response to some situations handled by the high availability software, and it can print messages to inform users logged in to a client of the cluster state and thus make them aware of actions required to maintain connectivity.

Installing and configuring the HACMP/ES software on each client consists of the following steps:

1. Install the base high availability system software on all clients.
2. Edit the `/usr/es/sbin/cluster/etc/clhosts` file on each client to suit your configuration.
3. Edit the `/usr/es/sbin/cluster/etc/clinfo.rc` script on each client to suit your configuration.

If a client is located in a network that has both HACMP and HACMP/ES clusters, see the related note on page 17-2.

4. Reboot each client in your cluster topology.

Note: If an SP node is used as a client, the SP Switch network can allow client access. For information on installing the HACMP/ES, Version 4.4 software on cluster nodes, see Chapter 14, Installing the HACMP/ES Software.

Installing the Base System Client Images

For a new installation, the `/usr` directory should have a minimum of **eight** megabytes (MB) of space available.

To install the base high availability software on a client:

1. Place the HACMP/ES tape into the tape drive and enter:

```
smit install_selectable_all
```

SMIT displays the **Install Selectable All** screen. If you are not sure of the name of the input device, press F4 to list the available devices. Select the proper drive and press Enter. That value is entered into the **INPUT device/directory** field as the valid input device.

2. Press Enter. SMIT refreshes the screen.

3. Enter field values as follows:

SOFTWARE to install Press F4 for a software listing. A popup window appears, listing all installed software.

Use the arrow keys to locate client software modules associated with the following Version 4.4 cluster images: **rsct.clients**, **rsct.core**, **cluster.es**, **cluster.es.cspoc**, and **cluster.adt.es**.

Next press F7 to select at least one client module from each image you want to install. Press Enter after making all selections. Your selections appear in this field.

Note that if you select at least one client module associated with an installable image, all other required client modules are installed automatically.

4. Enter values for other fields as appropriate for your site.

5. Press Enter when you are satisfied with the entries. SMIT responds:

```
Are you sure?
```

6. Press Enter again.

You are then instructed to read the HACMP/ES 4.4 **release_notes** file in the `/usr/lpp/cluster/doc` directory for further instructions.

Editing the `/usr/es/sbin/cluster/etc/clhosts` File on Clients

As installed, the **clhosts** file on an HACMP/ES client node contains no hostnames or addresses. You must provide the HACMP/ES server addresses at installation time. This file should contain all boot and service names (or addresses) of HACMP/ES servers accessible through logical connections to this client node. Upon startup, **clinfo** uses these names to attempt communication with a **clsmuxpd** process executing on an HACMP/ES server

Note: If a client is located in a network that has both HACMP and HACMP/ES clusters, the following rules apply to the **clhosts** file:

- If the client has HACMP software installed, the **clhosts** file must be configured with only HACMP hostnames. If the file contains any HACMP/ES hostnames, **clinfo** will abort.
- If the client has HACMP/ES software installed, the **clhosts** file may contain both HACMP and HACMP/ES hostnames.

Warning: Do not include standby addresses in the **clhosts** file, and do not leave this file empty. If either of these two conditions exist, neither **clinfo** nor **clstat** works properly.

An example `/usr/es/sbin/cluster/etc/clhosts` file follows:

```
n0_cl83      # n0 service
n2_cl83      # n2 service
n3_cl83      # n3 service
```

Editing the `/usr/es/sbin/cluster/etc/clinfo.rc` Script

The `/usr/es/sbin/cluster/etc/clinfo.rc` script, executed whenever a cluster event occurs, updates the system's ARP cache. If you are not using the hardware address swapping facility, a copy of the `/usr/es/sbin/cluster/etc/clinfo.rc` script must exist on each node and client in the cluster in order for all ARP caches to be updated and synchronized.

The HACMP/ES software is distributed with a template version of the `/usr/es/sbin/cluster/etc/clinfo.rc` script. You can use the script as distributed, add new functionality to the script, or replace it with a custom script.

Note: If you are not using hardware address swapping, the ARP functionality must remain.

The format of the **clinfo** call to **clinfo.rc**:

```
clinfo.rc {join,fail,swap} interface_name
```

When **clinfo** gets a `cluster_stable` event, or when it connects to a new **clsmuxpd**, it receives a new map. **clinfo** then checks for changed states of interfaces.

- If a new state is UP, **clinfo** calls `clinfo.rc join interface_name`.
- If a new state is DOWN, it calls `clinfo.rc fail interface_name`.
- If **clinfo** receives a **node_down_complete** event, it calls **clinfo.rc** with the fail parameter for each interface currently UP.
- If **clinfo** receives a **fail_network_complete** event, it calls `clinfo.rc` with the fail parameter for all associated interfaces.
- If **clinfo** receives a **swap_complete** event, it calls `clinfo.rc swap interface_name`.

See Chapter 8, Cluster Events: Tailoring and Creating, for complete information on cluster events and tailoring scripts.

See the *HACMP for AIX Programming Client Applications* guide for a sample client application that uses the Clinfo C API within the context of a customized **clinfo.rc** script.

Note: If you have written applications that use the Clinfo API and plan to use symmetric multi-processors, you may need to make changes to your application. See the *HACMP for AIX Programming Client Applications Guide* for updated information on the library routines. Then recompile and link your application.

Updating Non-Clinfo Clients

This section suggests two methods for updating the ARP cache of non-Clinfo clients.

Pinging Non-Clinfo Clients

On clients that are not running Clinfo, update the client's local ARP cache by pinging the client from the cluster node. Add the hostname or IP address of a client host you want to notify to the PING_CLIENT_LIST variable in the **clinfo.rc** script. Now, whenever a cluster event occurs, **clinfo.rc** executes the following command for each host specified in PING_CLIENT_LIST:

```
/etc/ping hostname 1024 2
```

Updating the ARP Cache on Non-Clinfo Clients after IP Address Takeover

When IP address takeover occurs, routers and other systems that are not running Clinfo must have their ARP tables updated to use the service IP address relocated to the surviving node. To perform this update, write a shell script with the following commands on the surviving node. Use the post-processing facility to call this shell script after the `acquire_takeover_addr` event.

To update the ARP cache on non-Clinfo clients:

1. Add a route to the router or system you wish to update, using the relocated interface:

```
route add router_IP_address -interface failed_node_service_IP_address
```

This creates a route that specifically uses the interface.
2. Delete the possible arp entry for the router:

```
arp -d router_IP_address
```
3. Ping the router or system you wish to update again:

```
ping -c1 router_IP_address
```

This updates the ARP cache.
4. Delete the route you previously added:

```
route delete router_IP_address failed_node_service_IP_address
```

This procedure forces the ping to go out over the interface with the relocated IP address. The address resolution triggered by the ping will provide the router or system you are pinging with the new hardware address now associated with this IP address.

Rebooting the Clients

The final step in installing the HACMP/ES software on a client is to reboot each client in your cluster topology.

Installing and Configuring Clients
Rebooting the Clients

Chapter 18 Configuring an HACMP/ES Cluster

This chapter describes how to configure an HACMP/ES cluster.

Overview

Complete the following procedures to define the cluster configuration:

- Define the cluster topology
- Define application servers and application monitors if desired
- Configure cluster resources
- Configure cluster security
- Verify the cluster environment
- Customize cluster events
- Customize cluster log files

Each procedure is described below.

Defining the Cluster Topology

Complete the following procedures to define the cluster topology. You only need to perform these steps on one node. When you synchronize the cluster topology, its definition is copied to the other node.

1. Give the cluster an ID and a name on the **Add a Cluster Definition** screen.
2. Define the nodes on the **Add Cluster Nodes** screen.
3. Define the network adapters on the **Add an Adapter** screen.
4. (optional)View or change the cluster topology services and group services setup. (Changes or additions are not likely to be needed.)
5. Copy the HACMP/ES ODM entries to each node in the cluster, using the **Synchronize Cluster Topology** option on the Configure Cluster menu.

Each procedure is described below.

Defining the Cluster ID and Name

The cluster ID and name uniquely identify each cluster in an HACMP/ES cluster environment. Complete the following steps to define the cluster ID and name. Refer to your completed network planning worksheets for the values.

To define the cluster ID and name:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **Cluster Configuration** and press Enter.

3. On the Cluster Configuration screen, select **Cluster Topology** and press Enter.
4. On the Cluster Topology screen, select **Configure Cluster** and press Enter.
5. On the Configure Cluster screen, select **Add a Cluster Definition** and press Enter to display the following screen.
6. Enter field values as follows:

Cluster ID	Enter a positive integer unique to your site (in the range 1 to 99,999).
Cluster Name	Enter an ASCII text string that identifies the cluster. The cluster name can include alpha and numeric characters and underscores. Use no more than 31 characters. It can be different from the hostname.

7. Press Enter.
The HACMP/ES software uses this information to create the cluster entries for the ODM.
8. Press F3 until you return to the Cluster Topology screen, or F10 to exit SMIT.

Defining Nodes

After defining the cluster name and ID, define the cluster nodes.

To define the cluster nodes:

1. Select **Configure Nodes** from the Cluster Topology menu and press Enter.
2. On the Configure Nodes screen, select **Add Cluster Nodes** and press Enter to display the following screen.
3. Enter the name for each cluster node in the **Node Names** field. Names cannot exceed 31 characters. They can include alpha and numeric characters and underscores. Leave a space between names. If you duplicate a name, you get an error message. The information is effective when you synchronize both nodes.
4. Press F3 until you return to the Cluster Topology screen, or F10 to exit SMIT.

Defining Adapters

To define the adapters, first consult your planning worksheets for both TCP/IP and serial networks, and then complete the following steps:

1. Select **Configure Adapters** from the Cluster Topology menu and press Enter.
2. On the Configure Adapters screen, select **Add an Adapter** and press Enter to display the Add an Adapter screen.

3. Enter field values as follows:

Adapter IP Label

Enter the IP label (the name) of the adapter you have chosen as the service address for this adapter. This name can be up to 31 characters long and can include alphanumeric characters, hyphens, and underscores.

Each adapter that can have its IP address taken over must have a boot adapter (address) label defined for it. Use a consistent naming convention for boot adapter labels. You will choose the Add an Adapter option again to define the boot adapter when you finish defining the service adapter.

Network Type

Indicate the type of network to which this adapter is connected. Pre-installed network modules are listed on the pop-up pick list. Supported types include: Ethernet, Token-Ring, ATM, FDDI, HPS, RS232, tmssa, and tmscsi.

Note: Choose **hps** for the SP Switch network.

Network Name

Enter the network name connected to this adapter. This name can be up to 31 characters long and can include alphanumeric characters and underscores.

The network name is arbitrary, but must be used consistently. If several adapters share the same physical network, make sure you use the same network name for each of these adapters.

Network Attribute

Indicate whether the network is public, private, or serial. Press TAB to toggle the values.

Ethernet, Token-Ring, and FDDI are *public* networks. The SP Switch and ATM are *private* networks. RS232 lines, target mode SSA loops, and target mode SCSI-2 buses are *serial* networks.

Adapter Function

Indicate whether the adapter's function is service, standby, or boot. Press TAB to toggle the values.

A node has a single service adapter for each public or private network. A serial network has a single service adapter. A node can have one or more standby adapters for each public network. Serial, private, and HPS networks do not have standby adapters. ATM is an exception; it does use standby adapters.

Note: In an HACMP/ES SP network, integrated Ethernet adapters cannot be used, and no standby adapters are configured. If a takeover occurs, the service address is aliased onto another node's service address.

In an ATM network, the adapter function should be listed as **svc_s** to indicate that the interface is used by HACMP/ES servers. Keep in mind that the netmask for all adapters in an HACMP/ES network must be the same to avoid communication problems between standby adapters after an adapter swap and after the adapter is reconfigured with its original standby address.

Adapter Identifier

Enter the IP address in dotted decimal format or a device file name.

IP address information is required for non-serial network adapters only if the node's address cannot be obtained from the domain name server or the local **/etc/hosts** file (using the adapter IP label given).

You must enter device file names for serial network adapters. RS232 serial adapters must have the device file name **/dev/tty n** . Target mode SCSI serial adapters must have the device file name **/dev/tm $scsin$** . Target mode SSA serial adapters must have the device file name **/dev/tm $ssan$** .

Adapter Hardware Address

This field is optional. Enter a hardware address for the adapter. The hardware address must be unique within the physical network. Enter a value in this field only if you are currently defining a service adapter, the adapter has a boot address, and you want to use hardware address swapping. The hardware address is 12 digits for Ethernet, Token-Ring and FDDI; and 14 digits for ATM.

Node Name

Assign the adapter to a node. Service, boot, and standby adapters are associated with a particular node.

4. Press Enter. The system adds these values to the HACMP/ES ODM and displays the Configure Adapters menu.
5. Define all the adapters, then press F3 until you return to the Cluster Configuration screen.

Configuring Global Networks

You can group multiple HACMP/ES networks of the same type under one global network name. This reduces the probability of network partitions that can cause the cluster nodes on one side of the partition to go down. You should always configure a global network when SP administrative ethernet adapters are included in the HACMP/ES configuration.

On large SP systems, each SP frame of nodes is usually set up as a different subnet on the SP administrative ethernet. Each of these subnets is then defined as a different HACMP/ES network. Defining a global network that includes all these SP administrative ethernets will avoid network partitions.

You define global networks by assigning a character string name to each HACMP/ES network that you want to include as a member of the global network. All members of a global network must be of the same type (all Ethernet, for example).

To define global networks, complete the following steps:

1. Select **Configure Global Networks** from the Cluster Topology menu and press Enter.
2. SMIT displays a pick list of defined HACMP/ES networks. Select one of these networks and press Enter.
3. SMIT displays the **Change/Show a Global Network** screen. The name of the network you selected is entered as the local network name. Enter the name of the global network (character string) and press Enter.
4. Repeat these steps to define all the HACM/ES networks to be included in each global network you want to define.
5. Define all the global networks, then press F3 until you return to the Cluster Topology screen.

Checking Topology Services and Group Services

The topology services and group services include the settings for the length of the Topology and Group services logs. The default settings are highly recommended. The SMIT screen contains entries for heartbeat settings, but these are not operable. The heartbeat rate is now set for each network module.

To view the screen for Topology and Group Services settings:

1. From the Cluster Topology menu, select **Configure Topology Services and Group Services > Change/Show Topology and Group Services Configuration** and press Enter. SMIT displays the Change/Show Topology and Group Services Configuration screen.

2. Enter field values as follows:

Interval between Heartbeats (seconds) No longer operable.

Fibrillate Count No longer operable.

Topology Services log length (lines) The default is 5000. This is usually sufficient.

Group Services log length (lines) The default is 5000. This is usually sufficient.

3. Press F3 to return to the **Cluster Topology** menu, or press Enter if you make any changes to field values, then press F3 to return to the **Cluster Topology** menu.

Configuring Cluster Performance Tuning

Cluster nodes sometimes experience extreme performance problems, such as large I/O transfers, excessive error logging, or lack of memory. When this happens, the Cluster Manager can be starved for CPU time. It might not reset the “deadman switch” within the time allotted. Misbehaved applications running at a priority higher than the cluster manager can also cause this problem.

The deadman switch is the AIX kernel extension that halts a node when it enters a hung state that extends beyond a certain time limit. This enables another node in the cluster to acquire the hung node’s resources in an orderly fashion, avoiding possible contention problems. If the deadman switch is not reset in time, it can cause a system panic and dump under certain cluster conditions.

Setting the following tuning parameters correctly may avoid some of the performance problems noted above. It is highly recommended to set the two AIX parameters; this may preclude having to change the HACMP Network Modules Failure Detection Rate.

- AIX high and low watermarks for I/O pacing
- AIX **syncd** frequency rate
- HACMP/ES Network Module Failure Detection Rate (Custom)
 - HACMP cycles to failure
 - HACMP heartbeat rate.

In HACMP/ES 4.4, you can configure these related parameters directly from HACMP SMIT.

Note: You must set the two AIX parameters on each cluster node. Network module settings are propagated to all nodes when you set them on one node and then synchronize the cluster topology.

Setting I/O Pacing

Although the most efficient high- and low-water marks vary from system to system, an initial high-water mark of **33** and a low-water mark of **24** provides a good starting point. These settings only slightly reduce write times and consistently generate correct fallover behavior from the HACMP/ES software.

See the *AIX Performance Monitoring & Tuning Guide* for more information on I/O pacing.

To change the I/O pacing settings:

1. Enter `smitty hacmp > Cluster Configuration > Advanced Performance Tuning Parameters > Change/Show I/O Pacing`
2. Configure the entry fields with the recommended HIGH and LOW watermarks:

HIGH water mark for pending write I/Os per file	33 is recommended for most clusters. Possible values are 0 to 32767.
---	--

LOW watermark for pending write I/Os per file	24 is recommended for most clusters. Possible values are 0 to 32766.
---	--

Setting Syncd Frequency

The **syncd** setting determines the frequency with which the I/O disk-write buffers are flushed. Frequent flushing of these buffers reduces the chance of deadman switch time-outs.

The AIX default value for **syncd** as set in `/sbin/rc.boot` is 60. It is recommended to change this value to 10. Note that the I/O pacing parameters setting should be changed first.

To change the **syncd** frequency setting:

1. Enter `smitty hacmp > Cluster Configuration > Advanced Performance Tuning Parameters > Change/Show syncd frequency`
2. Configure the entry fields with the recommended **syncd** frequency:

syncd frequency in seconds	10 is recommended for most clusters. Possible values are 0 to 32767.
----------------------------	---

Changing the Failure Detection Rate of a Network Module

Warning: I/O pacing must be enabled before changing the failure detection rate of a Network Module; it regulates the number of I/O data transfers. Also keep in mind that the setting for the Failure Detection Rate is network specific, and may vary. You should also try adjusting the **syncd** rate and test the system thoroughly before changing the attributes of a network module.

Be sure to consult Chapter 4, Planning Cluster Network Connectivity, for information on heartbeat settings for each type of network module and how these settings interact with the deadman switch before changing the defaults.

To change the attributes of a network module:

1. Stop cluster services on all cluster nodes.
2. Enter `smitty hacmp`
3. Select **Cluster Configuration > Advanced Performance Tuning Parameters > Change/Show Network Modules** and press Enter.
SMIT displays a list of defined network modules.
4. Select the network module you want to change and press Enter.
SMIT displays the attributes of the network module, with the current values.

Network Module Name	Name of network type, for example, ether.
Description	Descriptive text, for example, Ethernet Protocol
Grace Period	Number of seconds topology services waits before declaring a node down once all adapters on the node go down, for example, 30.
Failure Detection Rate	The default is Normal . Choices are Fast , Slow , and Custom .

Failure Cycle The current setting is the default for the network module selected. (Default for ether is 2.) This is the number of successive heartbeats that can be missed before the interface is considered to have failed. The failure cycle and the heartbeat interval determine how soon a failure can be detected. The time needed to detect a failure can be calculated using this formula: (heartbeat interval) * (failure cycle) * 2 seconds. If you choose **Custom** for the **Failure Detection Rate**, you can enter a number from 1 to 21474.

Heartbeat Rate The current setting is the default for the network module selected. This parameter tunes the interval (in tenths of a second) between heartbeats for the selected network module. If you select the **Custom** option in the Failure Detection Rate field, you can enter a number from 1 to 21474.

5. To change the heartbeat rate, you must first select **Custom** for the Failure Detection Rate field. Make the desired changes and press Enter. SMIT executes the command to modify the values of these attributes in the ODM.

6. On the local node, synchronize the cluster topology. Return to the SMIT **Cluster Topology** menu and select the **Synchronize Cluster Topology** option.

The configuration data stored in the DCD on each cluster node is updated and the changed configuration becomes the active configuration when you do a topology DARE.

Synchronizing the Cluster Definition across Nodes

Complete the following steps to synchronize a cluster definition across nodes:

1. On the Cluster Topology screen, select **Synchronize Cluster Topology** and press Enter. SMIT displays the Synchronize Cluster Topology screen.

2. Enter field values as follows:

Ignore Cluster Verification Errors By choosing **yes**, the result of the cluster verification is ignored and the configuration is synchronized even if verification fails. By choosing **no**, the synchronization process terminates; view the error messages in the system error log to determine the configuration problem.

Emulate or Actual If you set this field to **Emulate**, the synchronization is an emulation and does not affect the Cluster Manager. If you set this field to **Actual**, the synchronization actually occurs, and any subsequent changes affect the Cluster Manager. **Actual** is the default value.

Skip Cluster Verification

By default, this field is set to **no** and the cluster topology verification program is run. To save time in the cluster synchronization process, you can toggle this entry field to **yes**. By doing so cluster verification will be skipped.

Cluster verification is optional only when a cluster is *inactive*. Even if one node is active, **clverify** will be run.

3. Press Enter.

The cluster definition (including all node, adapter, and network module information) is copied to the other nodes.

4. Press F10 to exit SMIT.

Configuring Resource Groups and Resources

You now configure resources and set up the node environment. This involves:

- Configuring resource groups and node relationships to behave as desired
- Defining application servers
- Defining application monitors if desired
- Adding individual resources, including application servers, to each resource group
- Setting up run-time parameters per node
- Synchronizing the node environment.

Each step is explained below.

Adding Resource Groups

To create resource groups:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **Cluster Configuration > Cluster Resources > Define Resource Groups > Add a Resource Group** and press Enter.
3. Enter the field values as follows:

Resource Group Name

Enter the desired name. Use no more than 31 characters. You can use alpha or numeric characters and underscores. Duplicate entries are not allowed.

Node Relationship

Toggle the entry field between Cascading, Rotating, and Concurrent.

Participating Node Names

Enter the names of the nodes that can own or take over this resource group. Enter the node with the higher priority first, followed by the node with the lower priority. Leave a space between node names.

4. Press Enter to add the resource group information to the HACMP/ES ODM.

5. Press F3 after the command completes until you return to the Cluster Resources screen, or F10 to exit SMIT.

Configuring Application Servers

An *application server* is a cluster resource used to control an application that must be kept highly available. Configuring an application server does the following:

- Associates a meaningful name with the server application. For example, you could give the Image Cataloger demo supplied with the HACMP/ES software a name such as *imserv*. You then use this name to refer to the application server when you define it as a resource during node environment definition. When you set up the node environment, you define an application server as a cascading, concurrent, or rotating resource.
- Points the cluster event scripts to the scripts that they call to start and stop the server application.
- Allows you to then configure application monitoring for that application server.

Note that this section does not discuss writing the start and stop scripts. See the vendor documentation for specific product information on starting and stopping a particular application.

Defining an Application Server

You can define an application server dynamically. You can define an application server and add it to a resource group in a single dynamic reconfiguration operation.

Complete the following steps to create an application server on any cluster node.

1. Enter the fastpath **smit hacmp** and select the following options: **Cluster Configuration > Cluster Resources > Define Application Servers > Add an Application Server**.

When you press Enter, SMIT displays the Add an Application Server screen.

2. Enter field values as follows:

Server Name Enter an ASCII text string that identifies the server. You will use this name to refer to the application server when you define resources during node configuration. The server name can include alphabetic and numeric characters and underscores. Use no more than 31 characters.

Start Script Enter the pathname of the script (followed by arguments) called by the cluster event scripts to start the application server. This script must be in the same location on each cluster node that might start the server. The contents of the script, however, may differ.

Stop Script Enter the pathname of the script called by the cluster event scripts to stop the server. This script must be in the same location on each cluster node that may start the server. The contents of the script, however, may differ.

3. Press Enter to add this information to the ODM on the local node. Press the F3 key to return to previous HACMP/ES SMIT screens to perform other configuration tasks.

Configuring Applications Integrated with HACMP: AIX Fast Connect, AIX Connections, and CS/AIX

The applications AIX Fast Connect for Windows, AIX Connections, and CS/AIX are integrated with HACMP/ES already so you can configure them, via the SMIT interface, as highly available resources in resource groups. These applications do not need to be associated with application servers or special scripts.

Refer to your planning worksheets as you prepare to configure any of these applications as resources.

Configuring AIX Fast Connect

AIX Fast Connect allows client PCs running Windows, DOS, and OS/2 operating systems to request files and print services from an AIX server. Fast Connect supports the transport protocol NetBIOS over TCP/IP. You can configure AIX Fast Connect resources using the SMIT interface.

Prerequisites

Before you can configure Fast Connect resources in HACMP/ES, make sure these steps have been taken:

- Install the Fast Connect Server on all nodes in the cluster.
- Make sure AIX print queue names match for all nodes in the cluster if Fast Connect printshares are to be highly available.
- For cascading and rotating resource groups, assign the *same* netBIOS names to each node when the Fast Connect Server is configured. This action will minimize the steps needed for the client to connect to the server after fallover.
- For concurrently configured resource groups, assign *different* netBIOS names across nodes.
- Configure on the Fast Connect Server those files and directories on the AIX machine that you want shared.

Converting from AIX Connections to AIX Fast Connect

If you previously configured the AIX Connections application as a highly available resource, and you now wish to switch to AIX Fast Connect, you should take care to examine your AIX Connections planning and configuration information before removing it from the resource group. Remember that you cannot have both of these applications configured at the same time in the same resource group, so you must unconfigure all AIX Connections realm/service pairs before configuring Fast Connect fileshares and print queues.

The following instructions are repeated in Chapter 24, Changing the Cluster Configuration.

Keep in mind that AIX Fast Connect does not handle the AppleTalk and NetWare protocols that AIX Connections is able to handle. Fast Connect is primarily for connecting with clients running Windows operating systems. Fast Connect uses NetBIOS over TCP/IP.

Follow these steps when converting from AIX Connections to Fast Connect:

Configuring an HACMP/ES Cluster

Configuring Applications Integrated with HACMP: AIX Fast Connect, AIX Connections, and CS/AIX

1. Refer to your original planning worksheet for AIX Connections, where you listed the participating nodes and the realm/service pairs you planned to configure. Compare this information to your Fast Connect planning worksheet so you can be sure you are not leaving anything out.

If you do not have your planning sheet, note the information in the AIX Connections field when you go into SMIT to remove the AIX Connections resources from the resource group.

2. Start the Fast Connect server on each node and verify that you can connect to the shared directories and files on each node in turn.
3. In SMIT, go to the Change/Show Resource Groups screen, as described in the section below.
4. Select AIX Connections Resources, and remove all specified realm/service pairs.
5. Select Fast Connect Services, and specify the resources you wish to configure. Make sure if you are specifying fileshares, you have defined their filesystems in the Filesystems field earlier in the SMIT screen.
6. Synchronize the cluster as usual after you have made all changes. Instructions for synchronizing resources begin on page 18-30.

Configuration Notes for Fast Connect

When configuring Fast Connect as a cluster resource in HACMP/ES, keep the following points in mind:

- When starting cluster services, the Fast Connect server should be stopped on all nodes, so that HACMP/ES can take over the starting and stopping of Fast Connect resources properly.
- In concurrent configurations, the Fast Connect server should have a second, non-concurrent, resource group defined that does not have Fast Connect on it. Having a second resource group configured in a concurrent cluster keeps the AIX filesystems used by Fast Connect cross-mountable and highly available in the event of a node failure.
- Fast Connect cannot be configured in a mutual takeover configuration. Make sure there are no nodes participating in more than one Fast Connect resource groups at the same time.

For instructions on using SMIT to configure Fast Connect services as resources, see the section Configuring Resources in a Resource Group on page 18-24.

Verification of Fast Connect

After completing your resource configuration, you synchronize cluster resources. During this process, if Fast Connect resources are configured in HACMP/ES, the **clverify** utility verifies:

- That the Fast Connect server application exists on all participating nodes in a resource group.
- That the Fast Connect fileshares are in filesystems that have been defined as resources on all nodes in the resource group.
- That Fast Connect resources are not configured in a mutual takeover form; that is, there are no nodes participating in more than one Fast Connect resource group.
- That AIX Connections and Fast Connect resources are not configured in the same resource group or on the same node.

Configuring AIX Connections

AIX Connections software lets you share files, printers, applications, and other resources between AIX workstations and PC and Mac clients. You can still take advantage of AIX's multi-user and multi-tasking facilities, scalability, file and record locking features, and other security features. The AIX Connections application is integrated with HACMP/ES so that you can configure it as a resource in your HACMP/ES cluster, making the protocols handled by AIX Connections—IPX/SPX, Net BEUI, and AppleTalk—highly available in the event of node or adapter failure.

This section contains information you may need before adding AIX Connections services in the SMIT screen.

Configuration Notes for AIX Connections

Keep these considerations in mind as you configure AIX Connections:

- Make sure you have copied the AIX Connections installation information to all nodes on which the program might be active.
- On start-up, HACMP/ES starts all AIX Connections services and the network protocols specified in resource groups active on the local node. HACMP/ES does not stop those active on remote nodes, however. You need to see that any AIX Connections services that shouldn't be active on node start-up are not.
- Before you boot up, comment out any **inststart** commands for HACMP/ES in the initialization scripts (*/etc/rc.lsserver*, */etc/rc.nwserver*, or */etc/rc.macserver*), or modify the */etc/inittab* file to not call these scripts. This prevents the possibility of two nodes broadcasting the same name in shared disk volumes.
- You must configure a realm's AIX Connections to use the service adapter, not the standby.

For instructions on configuring AIX Connections resources in SMIT, see the section *Configuring Resources in a Resource Group* on page 18-24.

Adapter Failure Considerations with AIX Connections

Keep the following considerations in mind regarding AIX Connections and adapter failures:

- You must define one interface for each network allowing adapter swap. Each IPX/SPX and AppleTalk interface references a physical network adapter card. Each NetBIOS interface, on the other hand, references a Local Area Network Adapter (LANA), and each LANA references a physical network adapter card.
- When an IPX/SPX or AppleTalk protocol is running, it broadcasts all services over all interfaces. You need to remove an interface from the configuration to deactivate it. Again, LANAs work differently. You can define a LANA and still deactivate it.
- Even though you need to set up the interfaces for AIX Connections to work, because of the limitations of the IPX/SPX and AppleTalk protocols, HACMP/ES always moves the interfaces to the service adapter on the network.
- The only way to change a protocol's network adapter is to change the configuration file and restart the protocol.
- A configuration file may not report accurately on a protocol's network adapters for either of two reasons: the configuration file either might have been changed without restarting the protocol, or it might not report the way the user originally entered the information.

Configuring an HACMP/ES Cluster

Configuring Applications Integrated with HACMP: AIX Fast Connect, AIX Connections, and CS/AIX

For instructions on adding AIX Connections services to a resource group, see the section Configuring Resources in a Resource Group on page 18-24.

Verification of AIX Connections Configuration

After you have configured your resources, you synchronize cluster resources across all nodes. During this process, the **clverify** command checks the following AIX Connections configuration information:

- Realm/service pairs are configured correctly on all nodes in the resource group.
- Volume references of realm/service pairs are configured on all nodes in the resource group.
- Printer references of realm/service pairs are configured on all nodes in the resource group.
- Attach points of realm/service pairs are configured on all nodes in the resource group.
- AIX Connections initialization files (**rc.lsserver**, **rc.macserver**, and **rc.nwserver**) contain no **inststart** commands that aren't commented out. (See the Configuration Notes section above for more information.)

If problems are detected, the realm/service configuration check produces an error, the rest warnings.

To save time, **clverify** can be skipped during synchronization of cluster resources. To skip this process, choose **Yes** at the **Skip Cluster Verification** field in the **Synchronize Cluster Resources** or **Synchronize Cluster Topology** SMIT screen.

Note: If the cluster is active, cluster verification cannot be skipped.

Shell Commands for AIX Connections

For information about AIX Connections-specific commands, see the appendix on HACMP for AIX Commands, in the *HACMP for AIX Administration Guide*.

Configuring CS/AIX Communications Links

CS/AIX communication links are integrated with HACMP/ES already so you can configure them, via the SMIT interface, as highly available resources in resource groups. They do not need to be associated with application servers or special scripts.

An HACMP/ES communication link contains CS/AIX configuration information which is specific to a given node and network adapter. This configuration information enables an RS/6000 computer to participate in an SNA network that includes mainframes, PCs and other workstations. You can configure CS/AIX Communications Links resources using the SMIT interface.

See Chapter 3, Initial Cluster Planning, for installation considerations, supported networks, CS/AIX protocols, and CS/AIX product versions

Creating a CS/AIX Communications Link

These steps describe how to configure a highly available CS/AIX communications link. After completing these steps on a single node, the HACMP/ES software copies the information to all cluster nodes when you synchronize the nodes.

To configure a CS/AIX Communications Link:

1. Enter the following to start HACMP/ES system management:

```
smit hacmp
```

2. Select **Cluster Configuration > Cluster Resources > Define Highly Available Communications Links > Define Communication Links**.

This brings you to the main menu for CS/AIX system configuration. If you need information on any entries, use the F1 help. A valid CS/AIX configuration must exist before a CS/AIX DLC profile can be made highly available.

3. Press F3 to return to the **Define Highly Available Communications Links** screen.
4. Select **Make Communications Links Highly Available > Add a Highly Available Communications Link**.
5. Enter field values as follows:

DLC Name	Identify the CS/AIX DLC profile to be made highly available. Pick F4 to see a list of the DLC names.
Port	Enter ASCII text strings for the names of any CS/AIX ports to be started automatically.
Link Station	Enter ASCII text strings for the names of the CS/AIX link stations. This field is only available for CS/AIX Version 5.0.
Service	Enter the full pathname of an application start script. This start script starts any application layer processes that use the communication link. This field is optional.

6. Press Enter to add this information to the HACMP/ES ODM.
7. Press F10 after the command completes to leave SMIT and return to the command line.
8. Once you have defined a highly available CS/AIX communications link, you must then configure it as a resource in a resource group. See *Configuring Resources in a Resource Group* on page 18-24 for information on how to do this.

Changing a CS/AIX Communications Link

To change or remove a highly available CS/AIX communications link see Chapter 24, *Changing the Cluster Configuration*, for instructions.

CS/AIX Communications Links as Highly Available Resources

CS/AIX connections are protected during adapter and node failures. This section describes the HACMP/ES actions that take place for each of these failures.

Configuring an HACMP/ES Cluster

Configuring Applications Integrated with HACMP: AIX Fast Connect, AIX Connections, and CS/AIX

Note: HACMP/ES handles the stopping of DLC profiles during an adapter or node failure differently depending on whether CS/AIX Version 5.0 or 4.2 is used, as follows:

- CS/AIX Version 5.0 allows individual DLC profiles to be stopped. This allows HACMP/ES to only stop the DLC profiles which are using the adapter being taken out of service. Communication on other DLC profiles is unaffected.
- CS/AIX Version 4.2 does not allow stopping of individual DLC profiles. HACMP/ES must stop all of the active DLC profiles, on all active adapters.

Adapter Fallover and Recovery

When a service adapter over which CS/AIX is running fails, HACMP/ES will take the following actions: Stop the LU2 or LU6.2 sessions; stop the link stations; if CS/AIX version 4.2, stop the CS/AIX server; modify the DLC profiles to use an available standby adapter; verify CS/AIX; restart CS/AIX if stopped; start the link stations; start the sessions; and start the applications.

Depending on the number of DLC profiles, this process may take several minutes. CS/AIX may be unavailable during the time it takes to recover from an adapter failure. Clients or applications connected before the failure may have to reconnect.

Node Fallover and Recovery

CS/AIX profiles are defined to HACMP/ES as resources in a resource group. When a node fails, the resource group is taken over in the normal fashion, and the CS/AIX DLC profiles are restarted on the takeover node. Any identified resources of that DLC profile, such as link stations and service applications, are started on the takeover node.

Network Fallover and Recovery

Network failures are handled as they would in a non-CS/AIX environment. When a network failure occurs, HACMP/ES detects an IP network down and logs an error message in the `/tmp/hacmp.out` file. Even though the CS/AIX network is independent of the IP network, it is assumed that an IP network down event indicates that the CS/AIX network is also down. Recovery is achieved through user customization, as is consistent with HACMP/ES network fallover strategy.

Verification of CS/AIX Communications Links

The `clverify` command checks the consistency of DLC profiles between nodes which have a fallover relationship. An error message is generated if a DLC profile is not available on a node participating in a resource group containing that profile. There is no checking for invalid CS/AIX configuration information; it is assumed that the system administrator has properly configured CS/AIX.

Configuring Application Monitoring

HACMP/ES can monitor specified applications and automatically take action to restart them upon detecting process death or other application failures.

Overview

When you configure an application monitor, you choose either a process monitor or a user-defined monitor. You then specify through the SMIT interface which application is to be monitored, and then define various parameters such as time intervals, retry counts, and action to be taken in the event the application cannot be restarted. You control the application restart process through the Notify Method, Cleanup Method, and Restart Method SMIT fields, and by adding pre- and post-event scripts to any of the failure action or restart events you choose.

You can temporarily suspend and then resume an application monitor in order to perform cluster maintenance.

When an application monitor is defined, each node's ODM is aware of all monitored applications and their configuration data. This data is propagated to all nodes during cluster synchronization, and is backed up when a cluster snapshot is created. In addition, the **clverify** utility checks that any user-specified methods exist and are executable on all nodes.

Process and User-Defined Monitoring

You can choose either of two application monitoring methods:

- **Process application monitoring** detects the death of one or more processes of an application, using RSCT Event Management.
- **User-defined application monitoring** checks the health of an application with a custom monitor method at user-specified polling intervals.

Process monitoring is easier to set up, as it uses the built-in monitoring capability provided by RSCT and requires no custom scripts. However, process monitoring may not be an appropriate option for all applications. User-defined monitoring can monitor more subtle aspects of an application's performance and is more customizable, but it takes more planning, as you must create the custom scripts.

Fallover and Notify Actions

In both methods, process or user-defined monitoring, when a problem is detected by the monitor, HACMP/ES attempts to restart the application on the current node and continues the attempts until a specified retry count is exhausted.

When an application cannot be restarted within the retry count, HACMP/ES takes one of two actions, which you specify when configuring the application monitor:

- Choosing **failover** causes the resource group containing the application to fall over to the node with the next-highest priority according to the resource policy. (See the note on page 18-18 regarding a possible effect of this action on resource group availability.)
- Choosing **notify** causes HACMP/ES to generate a `server_down` event, similar to a `network_down` event, to inform the cluster of the failure.

Retry Count and Restart Interval

The restart activity depends on two configurable parameters, the *retry count* and the *restart interval*.

- The retry count specifies how many times HACMP/ES should try restarting before considering the application failed and taking subsequent fallover or notify action.
- The restart interval dictates the number of seconds that the restarted application must remain stable before the retry count is reset to zero, thus completing the monitor activity until the next failure occurs.

If the application successfully starts up before the retry count is exhausted, the restart interval comes into play. By resetting the restart count, it prevents unnecessary fallover action that could occur when applications fail several times over an extended time period. For example, a monitored application with a restart count set to three (the default) could fail to restart twice, and then successfully start and run cleanly for a week before failing again. This “third” failure should be counted as a new failure with three new restart attempts before invoking the fallover policy. The restart interval, set properly, would ensure the correct behavior: it would have reset the count to zero when the application was successfully started and in a stable state after the earlier failure.

Be careful not to set the restart interval too short. If it is too short, the count could be reset to zero too soon, before the immediate next failure, and the fallover or notify activity will never occur.

See the instructions for setting the retry count and restart intervals later in this chapter for additional details.

Application Monitoring Prerequisites and Considerations

Keep the following in mind when planning and configuring application monitoring:

- Any application to be monitored must be defined to an application server in an existing cluster resource group.
- Only one application server per resource group can be monitored. When planning your cluster, be sure to put applications you plan to monitor in separate resource groups.
- A monitored application server may not be a member of more than one resource group.
- A monitored application cannot be under SRC control.
- Application monitoring is not supported in mixed-version HACMP/ES clusters; synchronization will fail in a mixed cluster with an application monitor configured.

Note on the Fallover Option and Resource Group Availability

Be aware that if you choose the **fallover** option of application monitoring—which could cause a resource group to migrate from its original node—the possibility exists that while the highest priority node is up, the resource group remains down. This situation occurs when an **rg_move** event moves a resource group from its highest priority node to a lower priority node, and then you bring the lower priority node down with a graceful shutdown. Unless you bring the resource group up manually, it remains in an inactive state.

Configuring a Process Application Monitor

Process application monitoring takes advantage of the RSCT Event Management functionality to detect the death of a process and generate an event. This section covers how to configure process application monitoring, in which you specify one or more processes of a single application to be monitored.

Remember that HACMP/ES can monitor only one application per resource group, though you can specify multiple *processes* of that application for process monitoring.

Note: Process monitoring may not be the appropriate solution for all applications. See the section Configuring a User-defined Application Monitor on page 18-21 for details on the other method of monitoring applications. For instance, you cannot monitor a shell script with a process application monitor. If you wish to monitor a shell script, configure a user-defined monitor.

Identifying Correct Process Names

For process monitoring, it is very important that you list the correct process names in the SMIT field. You should use processes that are listed in response to the **ps -el** command, and not **ps -f**.

If there is any doubt about the correct names, here is a recommended short procedure to identify all the process names for your list:

1. Type `ps -el | cut -c72-80 | sort > list1`
2. Run the application server.
3. Type `ps -el > | cut -c72-80 | sort > list2`
4. Compare the two lists by typing `diff list1 list2 | grep\>`

The result should be a list of all the processes spawned by the application server. You may choose not to include all of them in your process list, but you now have a complete and accurate list of possible processes to monitor.

Steps for Configuring a Process Application Monitor

Remember that an application must be defined to an application server before you set up the monitor. Set up your process application monitor as follows:

1. Type `smit hacmp` to open a SMIT session.
2. Choose **Cluster Configuration > Cluster Resources > Configure Application Monitoring > Define Process Application Monitor > Add Process Application Monitor**

A list of application servers, if previously defined, appears.

3. Choose the application server for which you want to add a process monitor.
4. In the Add Process Application Monitor screen, fill in the field values as follows.

Application Server Name	(This field is already filled with the name of the application server you selected, and cannot be edited here.)
Processes to Monitor	Specify the process(es) to monitor. You can type more than one process name. Use spaces to separate the names. Note: To be sure you are using correct process names, use the names as they appear from the ps -el command (NOT ps -f), as explained in the section Identifying Correct Process Names on page 18-19.
Process Owner	Specify the user id of the owner of the processes specified above, for example <i>root</i> . Note that the process owner must own all processes to be monitored.
Instance Count	Specify how many instances of the application to monitor. The default is 1 instance. Note: this number <i>must</i> be 1 if you have specified more than one process to monitor.
Stabilization Interval	Specify the time (in seconds) to wait before beginning monitoring. For instance, with a database application, you may wish to delay monitoring until after the start script and initial database search have been completed. You may need to experiment with this value to balance performance with reliability. Note: In most circumstances, this value should <i>not</i> be zero.
Restart Count	Specify the restart count, denoting the number of times to attempt to restart the application before taking any other actions. The default is 3 . Note: Make sure you enter a Restart Method below if your Restart Count is any non-zero value.
Restart Interval	Specify the interval (in seconds) that the application must remain stable before resetting the restart count. Do not set this to be shorter than (Restart Count) x (Stabilization Interval). The default is 10% longer than that value. If the restart interval is too short, the restart count will be reset too soon and the desired fallover or notify action may not occur when it should. See page 18-18 for additional notes on the importance of the Restart Interval and the Restart Count.

Action on Application Failure	Specify the action to be taken if the application cannot be restarted within the restart count. You can keep the default choice notify , which runs an event to inform the cluster of the failure, or choose fallover , in which case the resource group containing the failed application moves over to the cluster node with the next-highest priority for that resource group. See the note on page 18-18 regarding how resource group availability could be affected by application monitor fallover action.
Notify Method	(Optional.) Define a notify method that will run when the application fails. This user-defined method runs during the restart process and during notify activity.
Cleanup Method	(Optional) Specify an application cleanup script to be invoked when a failed application is detected, before invoking the restart method. The default is the application server stop script defined when the application server was set up. Note: With application monitoring, since the application is already stopped when this script is called, the server stop script may fail.
Restart Method	(Required if Restart Count is not zero.) The default restart method is the application server start script defined previously, when the application server was set up. You can specify a different method here if desired.

5. Press enter when you have entered the desired information.

The values are then checked for consistency and entered into the ODM. When the resource group comes online, the application monitor starts.

When you synchronize the cluster, the **clverify** utility verifies that all methods you have specified exist and are executable on all nodes.

Configuring a User-defined Application Monitor

User-defined application monitoring allows you to write a monitor method to test for conditions other than process death. For example, if an application sometimes becomes unresponsive even though it is still running, a custom monitor method could test the application at defined intervals and report when the application's response is too slow. Also, some applications (shell scripts, for example) cannot be registered with RSCT, so process monitoring cannot be configured for them. A user-defined application monitor method can monitor these types of applications.

For instructions on defining a process application monitor, which requires no custom monitor method, refer back to the section beginning on page 18-19

Notes on Defining a Monitor Method

Unlike process monitoring, user-defined application monitoring requires you to provide a script to test the health of the application. You must also decide on a suitable polling interval.

When devising your custom monitor method, keep the following points in mind:

- The monitor method must be an executable program (it can be a shell script) that tests the application and exits, returning an integer value that indicates the application's status. The return value must be zero if the application is healthy, and must be a non-zero value if the application has failed.
- HACMP/ES cannot pass arguments to the monitor method.
- The method can log messages to `/tmp/clappmond.<application monitor name>.monitor.log` by simply printing messages to the standard output (**stdout**) file. The monitor log file is overwritten each time the application monitor runs.
- Since the monitor method is set up to be killed if it does not return within the specified polling interval, do not make the method overly complicated.

Steps for Configuring a User-Defined Application Monitor

To set up a user-defined application monitoring method:

1. Type `smit hacmp` to reach the main menu.
2. Choose **Cluster Configuration > Cluster Resources > Configure Application Monitoring > Define Custom Application Monitor > Add Custom Application Monitor**.
3. From the list of defined application servers, choose the one for which you want to add a monitoring method.
4. In the Add Custom Application Monitor screen, fill in field values as follows.

Note that the Monitor Method and Monitor Interval fields require you to supply your own scripts and specify your own preference for polling interval.

Application Server Name	(This field is already filled in with the name of the selected application server, and cannot be edited here.)
Monitor Method	Enter a script or executable for custom monitoring of the health of the specified application. Do not leave this field blank. Note that the method must return a zero value if the application is healthy, and a non-zero value if a problem is detected. You can have the monitor log messages to the log file <code>/tmp/clappmond.<application monitor name>.monitor.log</code> by having it print messages to the stdout file. The messages are automatically redirected to the monitor log.
Monitor Interval	Enter the polling interval (in seconds) for checking the health of the application. If the monitor does not respond within this interval, it is considered "hung."

Hung Monitor Signal	Enter a signal to kill the user-defined monitor method if it does not return within the monitor interval. The default signal is kill -9 .
Stabilization Interval	Specify the time (in seconds) to wait before beginning monitoring. For instance, with a database application, you may wish to delay monitoring until after the start script and initial database search have been completed. You may need to experiment with this value to balance performance with reliability. Note: In most circumstances, this value should <i>not</i> be zero.
Restart Count	Specify the restart count, denoting the number of times to attempt to restart the application before taking any other actions. The default is 3 .
Restart Interval	Specify the interval (in seconds) that the application must remain stable before resetting the restart count. Do not set this to be shorter than (Restart Count) x (Stabilization Interval + Monitor Interval). The default is 10% longer than that value. If the restart interval is too short, the restart count will be reset too soon and the desired failure response action may not occur when it should. See page 18-18 for additional notes on the importance of the Restart Interval and the Restart Count.
Action on Application Failure	Specify the action to be taken if the application cannot be restarted within the restart count. You can keep the default choice notify , which runs an event to inform the cluster of the failure, or choose fallover , in which case the resource group containing the failed application moves over to the cluster node with the next-highest priority for that resource group.
Notify Method	(Optional.) Define a notify method that will run when the application fails. This user-defined method runs during the restart process and during notify activity.
Cleanup Method	(Optional) Specify an application cleanup script to be invoked when a failed application is detected, before invoking the restart method. The default is the application server stop script defined when the application server was set up. Note: With application monitoring, since the application may be already stopped when this script is called, the server stop script may fail. For more information on stop scripts, see Appendix C, Applications and HACMP.

Restart Method (Required if Restart Count is not zero.) The default restart method is the application server start script defined previously, when the application server was set up. You can specify a different method here if desired.

5. Press enter when you have filled in all desired information.

The values are then checked for consistency and entered into the ODM. When the resource group comes online, the application monitor starts.

When you synchronize the cluster, the **clverify** utility verifies that all methods you have specified exist and are executable on all nodes.

Suspending, Changing, and Removing Application Monitors

You can temporarily suspend an application monitor in order to perform cluster maintenance. You should not change the application monitor configuration while in a suspended state.

For instructions on temporarily suspending application monitoring, changing the configuration, or permanently deleting a monitor, see *Changing or Removing Application Monitors* on page 24-21 in Volume 2 of this manual.

Configuring Resources in a Resource Group

Once you have defined a resource group, you assign resources to it. You can configure a resource group even if a node is powered down; however, SMIT cannot list possible shared resources for the node (making configuration errors more likely).

Resource Configuration Considerations

Keep the following in mind as you prepare to define the resources in your resource group:

- You cannot configure a resource group until you have completed the information on the **Add a Resource Group** screen. If you need to do this, refer back to the instructions under *Adding Resource Groups* on page 18-9.
- If you configure a cascading resource group with an NFS mount point, you must also configure the resource to use IP Address Takeover. If you do not do this, takeover results are unpredictable. You should also set the field value **Filesystems Mounted Before IP Configured** to **true** so that the takeover process proceeds correctly.
- When setting up a cascading resource with an IP Address takeover configuration, each cluster node should be configured in no more than $(N + 1)$ resource groups on a particular network. Here, N is the number of standby adapters on a particular node and network.
- If you configure application monitoring, remember that HACMP/ES can monitor only one application in a given resource group, so you should put applications you intend to have HACMP/ES monitor in separate resource groups.

HANFS Functionality Added to HACMP/ES 4.4

As you enter resource information, you can specify some items related to NFS.

Prior to version 4.4, HACMP for AIX included a separate product subsystem called High Availability for Network File System for AIX (HANFS for AIX). HANFS for AIX provided reliable NFS server capability by allowing a backup processor to recover current NFS activity should the primary NFS server fail. The HANFS special functionality extended the HACMP architecture to include highly available modifications and locks on NFS filesystems. HANFS clusters could have a maximum of two nodes.

In version 4.4, the HANFS functionality has been added to the basic HACMP architecture. The following enhancements are included in version 4.4 of the HACMP/ES product subsystem:

- You can use the Reliable NFS server capability that preserves locks and dupcache (2-node clusters only).
- You can now specify a network for NFS mounting.
- You can define NFS exports and mounts at the directory level.
- You can specify export options for NFS-exported directories and filesystems.

Entering Resource Information in SMIT

To configure resources in SMIT:

1. On the Cluster Resources menu, select **Change/Show Resources/Attributes for a Resource Group** and press Enter to display a list of defined resource groups.
2. Select the resource group you want to configure and press Enter. SMIT returns the following screen with the **Resource Group Name**, **Node Relationship**, and **Participating Node Names** fields filled in.

If the participating nodes are powered on, you can press F4 to list the shared resources. If a resource group/node relationship has not been defined, or if a node is not powered on, F4 displays the appropriate warnings.

SMIT displays the Configure a Resource Group screen.

3. Enter the field values as follows:

Service IP Label	List the IP labels to be taken over when this resource group is taken over. Press F4 to see a list of valid IP labels. These include addresses which rotate or may be taken over.
Filesystems	Identify the file systems to include in this resource group. Press F4 to see a list of the file systems.
Filesystems Consistency Check	Identify the method of checking consistency of file systems, fsck (default) or logredo (for fast recovery).

Filesystems Recovery Method	<p>Identify the recovery method for the file systems, parallel (for fast recovery) or sequential (default).</p> <p>Do <i>not</i> set this field to parallel if you have shared, nested file systems. These must be recovered sequentially. (Note that the cluster verification utility, clverify, does not report file system and fast recovery inconsistencies.)</p>
Filesystems/Directories to Export	<p>Identify the filesystems or directories to be NFS exported. The filesystems should be a subset of the filesystems listed above. The directories should be contained in one of the filesystems listed above. Press F4 for a list.</p>
Filesystems/Directories to NFS Mount	<p>Identify the filesystems or directories to NFS mount. All nodes in the resource chain will attempt to NFS mount these filesystems or directories while the owner node is active in the cluster.</p>
Network for NFS Mount	<p>(This field is optional.)</p> <p>Choose a previously defined IP network where you want to NFS mount the filesystems. The F4 key lists valid networks.</p> <p>This field is relevant only if you have filled in the previous field. The Service IP Label field should contain a service label which is on the network you choose.</p> <p>Note: You can specify more than one service label in the Service IP Label field. It is highly recommended that at least one entry be an IP label on the network chosen here.</p> <p>If the network you have specified is unavailable when the node is attempting to NFS mount, it will seek other defined, available IP networks in the cluster on which to establish the NFS mount.</p>
Volume Groups	<p>Identify the shared volume groups that should be varied on when this resource group is acquired or taken over. Press F4 to see a list of shared volume groups.</p> <p>If you have previously entered values in the Filesystems field, the appropriate volume groups are already known to the HACMP/ES software.</p>
Concurrent Volume Groups	<p>Identify the shared volume groups that can be accessed simultaneously by multiple nodes. Press F4 to see a list of shared concurrent volume groups.</p>

Raw Disk PVIDs	<p>Press F4 for a listing of the PVIDs and associated hdisk device names.</p> <p>If you have previously entered values in the Filesystems or Volume groups fields, the appropriate disks are already known to the HACMP/ES software.</p> <p>If you are using an application that directly accesses raw disks, list the raw disks here.</p>
AIX Connections Services	<p>Press F4 to choose from a list of all realm/service pairs that are common to all nodes in the resource group. You can also type in realm/service pairs. Use % as a divider between service name and service type; do not use a colon. <i>Note that you cannot configure both AIX Connections and AIX Fast Connect in the same resource group.</i></p>
AIX Fast Connect Services	<p>Press F4 to choose from a list of Fast Connect resources common to all nodes in the resource group, specified during the initial configuration of Fast Connect. If you are adding Fast Connect fileshares, make sure you have defined their filesystems in the resource group. <i>Note that you cannot configure both AIX Connections and AIX Fast Connect in the same resource group.</i></p>
Application Servers	<p>Indicate the application servers to include in the resource group. Press F4 to see a list of application servers. See the section Configuring Resources in a Resource Group on page 18-24 for information on defining application servers.</p>
Highly Available Communications Links	<p>Indicate the communications links to include in the resource group. Press F4 to see a list of communications links. See the section Configuring CS/AIX Communications Links on page 18-14 for information on defining communications links.</p>
Miscellaneous Data	<p>A string you want to place into the topology, along with the resource group information. It is accessible by the scripts, for example, <i>Database1</i>.</p>

**Inactive Takeover
Activated**

Set this variable to control the *initial acquisition* of a resource group by a node when the node/resource relationship is cascading. This variable does not apply to rotating or concurrent resource groups.

If Inactive Takeover is **true**, then the first node in the resource group to join the cluster acquires the resource group, regardless of the node's designated priority.

If Inactive Takeover is **false**, the first node to join the cluster acquires only those resource groups for which it has been designated the highest priority node.

The default is **false**.

**Cascading without
Fallback**

Set this variable to determine the fallback behavior of a cascading resource group.

Note: You may find it useful to review the definitions of *failover* and *fallback* in the section, Defining the Node Relationship of Your Cluster on page 3-6

When the CWOFF variable is set to **false**, a cascading resource group will fallback as a node of higher priority joins or reintegrates into the cluster.

When CWOFF is **true**, a cascading resource group will *not* fallback as a node of higher priority joins or reintegrates into the cluster. It migrates from its owner node only if the owner node fails. It will not fallback to the owner node when the owner node reintegrates into the cluster.

The default for CWOFF is **false**.

9333 Disk Fencing Activated

Set as required for your configuration.

SSA Disk Fencing Activated

Set as required for your configuration.

**Filesystems Mounted Before
IP Configured**

This field specifies whether, on takeover, HACMP/ES takes over volume groups and mounts a failed node's filesystems before or after taking over the failed node's IP address or addresses.

The default is **false**, meaning the IP address is taken over first. Similarly, upon reintegration of a node, the IP address is acquired before the filesystems.

Set this field to **true** if the resource group contains filesystems to export. This is so that the filesystems will be available once NFS requests are received on the service address.

4. Press Enter to add the values to the HACMP/ES ODM.
5. Press F3 until you return to the Cluster Resources menu, or F10 to exit SMIT.

NFS Exporting Filesystems and Directories

The process of NFS-exporting filesystems and directories in HACMP/ES is different from that in AIX. Remember the following when NFS-exporting in HACMP:

Specifying Filesystems and Directories to NFS Export

While in AIX, you list filesystems and directories to NFS-export in the `/etc/exports` file; in HACMP/ES, you must put these in a resource group.

Specifying Export Options for NFS Exported Filesystems and Directories

If you want to specify special options in for NFS-exporting in HACMP, you can create a `/usr/sbin/cluster/etc/exports` file. This file has the same format as the regular `/etc/exports` file used in AIX.

Note: Use of this alternate exports file is optional. HACMP/ES checks the `/usr/sbin/cluster/etc/exports` file when NFS-exporting a filesystem or directory. If there is an entry for the filesystem or directory in this file, HACMP/ES will use the options listed. If the filesystem or directory for NFS-export is not listed in the file, or, if the user has not created the `/usr/sbin/cluster/etc/exports` file, the filesystem or directory will be NFS-exported with the default option of root access for all cluster nodes.

Configuring the Optional `/usr/sbin/cluster/etc/exports` File

In this step you add the directories of the shared filesystems to the exports file. Complete the following steps for each filesystem you want to add to the exports file. Refer to your NFS-Exported Filesystem Worksheet.

Note: Remember that this alternate exports file does not specify *what* will be exported, only *how* it will be exported. To specify what to export, you must put it in a resource group.

1. Enter the `smit mknfsexp` fastpath to display the **Add a Directory to Exports List** screen.
2. In the **EXPORT directory now, system restart or both** field, enter `restart`.
3. In the **PATHNAME of alternate Exports file** field, enter `/usr/sbin/cluster/etc/exports`. This step will create the alternate exports file which will list the special NFS export options.
4. Add values for the other fields as appropriate for your site, and press Enter. Use this information to update the `/usr/sbin/cluster/etc/exports` file.
5. Press F3 to return to the **Add a Directory to Exports List** screen, or F10 to exit SMIT.
6. Repeat steps 1 through 4 for each filesystem or directory listed in the **FileSystems/Directories to Export** field on your planning worksheets.

Configuring Run-Time Parameters

Run-time parameters include settings for the debug level and name serving for each node.

To define the run-time parameters for a node:

1. On the Cluster Resources menu, select **Change/Show Run Time Parameters** to list the node names. You define run-time parameters individually for each node.

2. Select a node name and press Enter. SMIT returns the following screen with the node name displayed.
3. Enter field values as follows:

Debug Level

Cluster event scripts have two levels of logging. The low level logs errors encountered while the script executes. The high level logs all actions performed by the script. The default is **high**.

Host uses NIS or Name Server

If the cluster uses Network Information Services (NIS) or name serving, set this field to **true**. The HACMP/ES software then disables these services before entering reconfiguration, and enables them after completing reconfiguration. The default is **false**.

4. Press Enter to add the values into the HACMP/ES ODM.
5. Press F3 to return to the **Cluster Resources** menu or Press F10 to exit SMIT.

Synchronizing Cluster Resources

Synchronizing the cluster resources sends the information contained on the local node to the remote nodes.

If all cluster nodes do not have the same setting for the security mode, the HACMP/ES software generates a run-time error at cluster startup.

If a node attempts to join a cluster with a node environment which is out-of-sync with the active node, it is denied. You must synchronize the node environment to the joining member.

To synchronize the cluster resources for all nodes:

1. Select **Synchronize Cluster Resources** from the Cluster Resources menu.
2. Enter field values as follows:

Ignore Cluster Verification Errors?

The default is **no**. Only choose **yes** if you want to force all nodes to accept the definition on the local node, despite verification errors.

Un/Configure Cluster Resources

The default is **yes** (enter all additions, changes, or deletions).

Emulate or Actual

If you set this field to **Emulate**, the synchronization will be an emulation and will not affect the Cluster Manager. If you set this field to **Actual**, the synchronization will actually occur, and any subsequent changes will be made to the Cluster Manager. **Emulate** is the default value.

Skip Cluster Verification

By default, this field is set to **no** and the verification of cluster resources program is run.

To save time in the synchronization process, you can toggle this entry field to **yes**.

3. Press Enter to synchronize the resource group configuration and node environment across the cluster.
4. Press F3 until you return to the HACMP/ES menu, or F10 to exit SMIT.

Note: Log files which are no longer stored in a default directory, but a directory of choice instead, should undergo verification by the **clverify** utility, which checks that each log file has the same pathname on every node in the cluster and reports an error if this is not the case.

Configuring Cluster Security

Kerberos is a network authentication protocol used on the SP. Based on a secret-key encryption scheme, Kerberos offers a secure authentication mechanism for client/server applications.

By centralizing command authority via one authentication server, normally configured to be the SP control workstation, Kerberos eliminates the need for the traditional TCP/IP access control lists (**.rhosts** files) that were used in earlier HACMP/ES security implementations. Rather than storing hostnames in a file (the **.rhosts** approach), Kerberos issues dually encrypted *authentication tickets*. Each ticket contains two encryption keys: One key is known to both the client user and to the ticket-granting service, and one key is known to both the ticket-granting service and to the target service that the client user wants to access. For a more detailed explanation of Kerberos and the security features of the SP system, refer to the *IBM Parallel System Support Programs for AIX Administration Guide*.

In addition, PSSP 3.2 provides the option of running an RS/6000 SP system with an enhanced level of security, and as of AIX 4.3.1, you can use DCE authentication rather than Kerberos 4 authentication. However, these options may affect your HACMP/ES functionality. Please read the following sections before planning your cluster security.

Configuring Kerberos Security with HACMP/ES for AIX

By setting up all network IP labels in your HACMP/ES configuration to use Kerberos authentication, you reduce the possibility of a single point of failure. You can configure Kerberos for a cluster automatically by running a setup utility called **cl_setup_kerberos**. Alternatively, you can perform the process manually. Because the utility-based approach is faster and less prone to error, it is usually preferable to the manual method.

To configure Kerberos on the SPs within an HACMP/ES cluster, you must perform these general steps (detailed procedures appear in the following sections):

Step	What you do...
1	Make sure that HACMP/ES has been properly installed on all nodes in the cluster. For more information, see Chapter 14, Installing the HACMP/ES Software.
2	<p>Configure the HACMP/ES cluster topology information on one node in the cluster. Note that because the cl_setup_kerberos utility needs an initial Kerberized rcmd path to each node in the cluster and to the control workstation, you must include the SP Ethernet as part of the configuration.</p> <p>Note that on the SP setup_authent is usually used to configure Kerberos on the entire SP system. setup_authent creates rcmd (used for rsh and rcp) service principals for all network IP labels listed in the System Data Repository (SDR). The SDR does not allow multiple IP labels to be defined on the same interface. However, HACMP/ES requires that multiple IP labels be defined for the same interface during IPAT configurations. HACMP/ES also requires that godm (Global ODM) service principals be configured on all IP labels for remote ODM operations. For these reasons, each time the nodes are customized after the SP setup_authent script is run (via setup_server or alone), you must rerun the cl_setup_kerberos script or manually reconfigure the systems to use Kerberos.</p>
3	Create new Kerberos service principals and configure all IP labels for Kerberos authentication. You can choose to perform these tasks automatically (see Configuring Kerberos Automatically on page 18-32) or manually (see Configuring Kerberos Manually on page 18-33).
4	Set the cluster security mode to Enhanced , then synchronize the cluster topology. See PSSP 3.2 Enhanced Security Options on page 18-36.
5	Delete (or at least edit) the cl_krb_service file, which contains the Kerberos service principals password you entered during the configuration process. At the very least, you should edit this file to prevent unauthorized users from obtaining the password and possibly changing the service principals.
6	Consider removing unnecessary .rhosts files. With Kerberos configured, HACMP/ES does not require the traditional TCP/IP access control lists provided by these files (but other applications might). You should consult your cluster administrator before removing any version of this file.

Configuring Kerberos Automatically

The **cl_setup_kerberos** utility automatically creates new Kerberos service principals in the Kerberos Authentication Database by copying the IP labels from the **cl_krb_service** file. It extracts the service principals and places them in a new Kerberos services file, **cl_krb-srvtab**; creates a **cl_klogin** file that contains additional entries required by the **.klogin** file; updates the **.klogin** file on the control workstation and on all nodes in the cluster; concatenates the **cl_krb-srvtab** file to each node's **/etc/krb-srvtab** file.

To run the **cl_setup_kerberos** utility:

Note: Make sure that you have already installed HACMP/ES on at least one node, and that you have configured the topology information before you perform this procedure.

1. Verify that there is a valid **.k** file on the control workstation. This file stores the Kerberos authentication password so that batched commands can be run. If the **.k** file is not present, issue the following command locally on the control workstation:

```
/usr/lpp/ssp/kerberos/etc/kstash
```

2. Run **cl_setup_kerberos** from the configured node. (The utility is found in the **/usr/es/sbin/cluster/sbin** directory.)

Note: You must be *within* the directory to run this command successfully. It is not sufficient to define the PATH correctly; the only way to run the **cl_setup_kerberos** command correctly is from within the **/usr/es/sbin/cluster/sbin** directory.

cl_setup_kerberos extracts the HACMP/ES IP labels from the configured node and creates a file, **cl_krb_service**, that contains all of the IP labels and additional format information required by the **add_principal** Kerberos utility. It also creates the **cl_adapters** file that contains a list of the IP labels required to extract the service principals from the authentication database.

3. When prompted, enter a Kerberos password for the new principals:

```
Password:
```

Note: This password is added to the **cl_krb_service** file. This can be the same as the Kerberos Administration Password, but doesn't have to be. Follow your site's password security procedures.

Configuring Kerberos Manually

To properly configure Kerberos on all HACMP-configured networks, you must perform the following general steps:

Step	What you do
1	Add an entry for each new Kerberos service principal to the Kerberos Authentication Database. See Adding New Service Principals to the Authentication Database on page 18-34.
2	Update the krb-srvtab file by extracting each newly added instance from the Kerberos Authentication Database. See Updating the krb-srvtab File on page 18-34.
3	Add the new service principals to each node's .klogin file. See Adding Kerberos Principals to Each Node's .klogin File on page 18-35.
4	Add the new service principals to each node's /etc/krb.realms file. See Adding Kerberos Principals to Each Node's /etc/krb.realms File on page 18-36.

Adding New Service Principals to the Authentication Database

To add new service principals to the Kerberos Authentication Database for each network interface:

1. On the control workstation, start the **kadmin** utility

```
kadmin
```

A welcome message appears.

2. At the `admin:` prompt type the **add_new_key** command with the name and instance of the new principal:

```
admin: ank service_name.instance
```

where

service_name is the service (**godm** or **rcmd**) and *instance* is the address label to be associated with the service. Thus, using the service **godm** and address label **il_sw** the command is:

```
admin: ank godm.il_sw
```

3. When prompted, enter the Kerberos Administration Password.

```
Admin password: password
```

4. When prompted, enter a Kerberos password for the new principal.

```
Password for service_name.instance: password
```

Note: The password can be the same as the Kerberos Administration Password, but doesn't have to be. Follow your site's password security procedures.

5. Verify that you have indeed added the new principals to the Kerberos database.

```
kdb_util dump /tmp/testdb
```

```
cat /tmp/testdb
```

Remove this copy of the database when you have finished examining it.

```
rm /tmp/testdb
```

Updating the krb-srvtab File

To update the **krb-srvtab** file and propagate new service principals to the HACMP/ES cluster nodes:

1. Extract each new service principal for each instance you added to the Kerberos Authentication Database for those nodes you want to update. (This operation creates a new file in the current directory for each instance extracted.)

```
usr/lpp/ssp/kerberos/etc/ext_srvtab -n il_sw il_en il_tr
```

2. Combine these new files generated by the **ext_srvtab** utility into one file called *node_name-new-srvtab*:

```
cat il_sw-new-srvtab il_en-new-srvtab il_tr-new-srvtab  
> node_name-new-srvtab
```

The new file appears in the directory where you typed the command.

Note: Shared labels (used for rotating resource groups) need to be included in every **krb-srvtab** file (for nodes in that rotating resource group), so you must concatenate each shared-label srvtab file into each *node_name-new-srvtab* file.

3. Copy each *node_name-new-srvtab* file to its respective node.
4. Make a copy of the current **/etc/krb-srvtab** file so that it can be reused later if necessary:

```
cp /etc/krb-srvtab /etc/krb-srvtab-date
```

(where *date* is the date you made the copy).
5. Replace the current **krb-srvtab** file with the new *node_name-new-srvtab* file:

```
cp node_name-new-srvtab /etc/krb-srvtab
```
6. Verify that the target node recognizes the new principals by issuing the following command on it:

```
ksrvutil list
```

You should see all the new principals for each network interface on that node; if not, repeat this procedure.

Adding Kerberos Principals to Each Node's .klogin File

To add the new Kerberos principals to the **/.klogin** file on each HACMP/ES cluster node:

1. Edit the **/.klogin** file on the control workstation to add the principals that were created for each network instance:

```
vi /.klogin
```

Here is an example of the **/.klogin** file for two nodes, i and j. ELVIS_IMP is the name of the realm that will be used to authenticate service requests. Each node has the SP Ethernet, a Token Ring service, and an Ethernet service adapter.

```
root.admin@ELVIS_IMP  
rcmd.i1@ELVIS_IMP  
rcmd.i1_ensvc@ELVIS_IMP  
rcmd.i1_trsvc@ELVIS_IMP  
rcmd.j1@ELVIS_IMP  
rcmd.j1_ensvc@ELVIS_IMP  
rcmd.j1_trsvc@ELVIS_IMP  
godm.i1@ELVIS_IMP  
godm.i1_ensvc@ELVIS_IMP  
godm.i1_trsvc@ELVIS_IMP  
godm.j1@ELVIS_IMP  
godm.j1_ensvc@ELVIS_IMP  
godm.j1_trsvc@ELVIS_IMP
```

2. Copy the **/.klogin** file from the control workstation to each node in the cluster.

To verify that you set this up correctly, issue a Kerberized **rsh** command on all nodes using one of the newly defined interfaces. For example:

```
/usr/lpp/ssp/rcmd/bin/rsh i1_ensvc date
```

To eliminate single points of failure, you should add Kerberos **rcmd** and **godm** principals for every interface configured in HACMP/ES.

Adding Kerberos Principals to Each Node's /etc/krb.realms File

To add the new Kerberos principals to the /etc/krb.realms file on each HACMP/ES cluster node:

1. Edit the /etc/krb.realms file on the control workstation and add the principals that were created for each network instance.

```
vi /etc/krb.realms
```

Here is an example of the **krb.realms** file for two nodes, i and j. ELVIS_IMP is the name of the realm that will be used to authenticate service requests. Each node has the SP Ethernet, a Token-Ring service, and an Ethernet service adapter.

```
root.admin ELVIS_IMP
i1 ELVIS_IMP
i1_ensvc ELVIS_IMP
i1_trsvc ELVIS_IMP
j1 ELVIS_IMP
j1_ensvc ELVIS_IMP
j1_trsvc ELVIS_IMP
i1 ELVIS_IMP
i1_ensvc ELVIS_IMP
i1_trsvc ELVIS_IMP
j1 ELVIS_IMP
j1_ensvc ELVIS_IMP
j1_trsvc ELVIS_IMP
```

2. Copy the /etc/krb.realms file from the control workstation to each node in the cluster.

PSSP 3.2 Enhanced Security Options

PSSP 3.2 provides the option of running your RS/6000 SP system with an enhanced level of security. This function removes the dependency PSSP has to internally issue **rsh** and **rcp** commands as a root user from a node. When this function is enabled, PSSP does not automatically grant authorization for a root user to issue **rsh** and **rcp** commands from a node. Be aware that if you enable this option, some procedures may not work as documented. To run HACMP, an administrator must grant the authorizations for a root user to issue **rsh** and **rcp** commands that PSSP would otherwise grant automatically. See the redbook *Exploiting RS/6000 SP Security: Keeping it Safe*, SG24-5521-00, for a description of this function and a complete list of limitations.

If the enhanced security feature is enabled, the administrator could authorize a root user to issue **rsh** and **rcp** commands using the following steps:

1. Prior to the altering, synchronizing, or verifying an HACMP/ES cluster configuration, the administrator of each node in the HACMP/ES cluster must create (or update) the root user's **rsh** authorization file (either **/.klogin** or **/.rhosts**) to allow the root users on the other nodes in the cluster to issue **rsh** commands to that node. Also, the appropriate AIX remote command authentication method would have to be enabled on the nodes of the HACMP/ES cluster (if the method was not already enabled).
2. Perform any desired alteration, verification, and synchronization of the HACMP/ES cluster configuration.
3. (Optional step if desired by the node administrators): Remove the authorization file entries added in step 1. Disable the authentication method (if enabled in step 1).

DCE Authentication

As of AIX 4.3.1, you are allowed the option of using DCE authentication rather than Kerberos 4 authentication. If you do this, you will not be able to alter, synchronize, or verify the HACMP/ES configuration. Note that you will not be able to explicitly move a resource group, since that is a type of reconfiguration. If DCE (i.e. only Kerberos V5) is enabled as an authentication method for the AIX remote commands, you can still use HACMP, but must perform the following steps:

1. Prior to configuring the HACMP/ES cluster, enable Kerberos V4 or Standard AIX as an authentication method for the AIX remote commands on the cluster nodes, and create the remote command authorization files (either **/.klogin** or **/.rhosts** files) for the root user on those nodes. This provides the ability for root to **rsh** among those nodes.
2. Configure the HACMP/ES cluster.
3. Remove the remote command authorization files created in step 1 on the HACMP/ES cluster nodes.
4. Disable the Kerberos V4 or Standard AIX authentication method enabled in step 1 on the HACMP/ES cluster nodes.

Setting a Cluster's Security Mode

You can set or change the security mode of all nodes in the cluster from the Change/Show Cluster Security SMIT screen. Because the cluster security mode is part of the HACMPcluster ODM, any changes you make are viewed as topology changes. This means that you must synchronize topology to propagate your security mode changes to all other nodes in the cluster. (You also verify the security setting by synchronizing/verifying the cluster topology.) For the same reason, you cannot dynamically reconfigure a cluster's topology and resource configuration simultaneously.

To set the security mode of all nodes in a cluster to **Enhanced**:

1. From the SMIT Cluster Configuration screen, select **Cluster Security > Change/Show Cluster Security**.
The Change/Show Cluster Security SMIT screen appears.
2. Set the security mode to **Enhanced**.
3. Synchronize the cluster topology. For more information, see Chapter 24, Changing the Cluster Configuration.

Verifying the Cluster Environment

This section describes how to verify the cluster environment, including the cluster and node configurations. This process ensures that all nodes agree on cluster topology, security, and assignment of resources.

Verifying Cluster and Node Environment

After defining the cluster and node environment, run the cluster verification procedure on one node to check that all nodes agree on the assignment of HACMP/ES cluster resources.

To verify the cluster and node configuration:

1. Enter the following command:

```
smit hacmp
```

2. From the main menu, select **Cluster Configuration > Cluster Verification > Verify Cluster** and press Enter.

The Verify Cluster screen appears.

Fill in the fields as follows:

Base HACMP Verification Modules

By default, both the cluster topology and resources verification programs are run. You can toggle this entry field to run either program, or you can select **none** to specify a custom-defined verification module in the Define Custom Verification Module field.

Define Custom Verification Module

Enter the name of a custom-defined verification module. You can also press F4 for a list of previously defined verification modules. By default, if no modules are selected, the `clverify` utility also will not check the base verification modules, and it generates an error message.

The order in which verification modules are listed determines the sequence in which selected modules are run. This sequence remains the same for subsequent verifications until different modules are selected.

Error Count

By default, the program will run to the end, even if it finds many errors. To cancel the program after a specific number of errors, type the number in this field.

Log File to store output

Enter the name of an output file in which to store verification output. By default, verification output is stored the `smit.log` file.

3. Press Enter.

SMIT runs the `clverify` utility. The output from the verification is displayed in the SMIT Command Status window. If you receive error messages, make the necessary changes and run the verification procedure again.

For example, the following figure illustrates the return from a failed verification.

Output from a Failed Verification

```
COMMAND STATUS

Command: failed          stdout: yes             stderr: no

Before command completion, additional instructions may appear below.

[TOP]
Contacting node cav ...
HACMPnode ODM on node cav verified.

Contacting node orion ...
HACMPnode ODM on node orion verified.

Contacting node sigmund ...
HACMPnode ODM on node sigmund verified.

Contacting node cav ...
HACMPnetwork ODM on node cav verified.

Contacting node orion ...
[MORE...203]
```

To synchronize all cluster nodes, use the **Synchronize Cluster Topology** option on the **Cluster Topology** SMIT screen. See the section on synchronizing the cluster topology in Chapter 24, *Changing the Cluster Configuration*, for more information.

Checking Cluster Topology

Run the following command to verify that all nodes agree on the cluster topology:

```
clverify cluster topology check
```

When the program finishes, check the output. If a problem exists with the cluster topology, a message similar to the following appears:

```
ERROR: Could not read local configuration
ERROR: Local Cluster ID XXX different from Remote Cluster ID XXX.
ERROR: Nodes have different numbers of networks
```

Forcing the Synchronization of Cluster Topology

If you are sure you want to define the cluster as it is defined on the local node, you can force agreement on cluster topology by choosing to ignore verification errors.

To synchronize a cluster definition across nodes:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **Cluster Configuration > Cluster Topology > Synchronize Cluster Topology** and press Enter.
3. On the resulting screen, choose whether to skip cluster verification in order to save time, select **yes** to Ignore Cluster Verification Errors and **Actual** to synchronize rather than emulate synchronization. Press Enter.

The cluster definition (including all node, adapter, and network module information) is copied from the local node to the other nodes.

4. Press F10 to exit SMIT.

Also see the **clverify** man page for further details on the utility.

Customizing Cluster Log Files

You can redirect a cluster log from its default directory to a directory of your choice. Should you redirect a log file to a directory of your choice, keep in mind that the requisite (upper limit) disk space for most cluster logs is 2MB. 14MB is recommended for **hacmp.out**.

Note: Logs should not be redirected to shared filesystems or NFS filesystems. Though this may be desirable in rare cases, such action may cause problems if the filesystem needs to unmount during a fallover event.

To redirect a cluster log from its default directory to another destination, take the following steps:

1. Enter
`smitty hacmp`
2. Select **Cluster System Management > Cluster Log Management > Change/Show Cluster Log Directory**
SMIT displays a picklist of cluster log files with a short description of each.

clstrmgr.debug	Generated by clstrmgr activity
cluster.mmdd	Cluster history files generated daily
cspoc.log	Generated by C-SPOC commands
dms_loads.out	Generated by deadman switch activity
emuhacmp.out	Generated by event emulator scripts
hacmp.out	Generated by event scripts and utilities
cluster.log	Generated by cluster scripts and daemons

3. Select a log that you want to redirect.
SMIT displays a screen with the selected log's name, description, default pathname, and current directory pathname. The current directory pathname will be the same as the default pathname if you do not elect to change it.

The example below shows the **cluster.mmdd** log file screen. Edit the final field to change the default pathname.

Custom Log Name	cluster.mmdd
Cluster Log Description	Cluster history files generated daily
Default Log Destination Directory	/usr/es/sbin/cluster/history
Log Destination Directory	The default directory name appears here. To change the default, enter the desired directory pathname.

4. Press F3 to return to the screen to select another log to redirect, or return to the Cluster System Management screen to proceed to the screen for synchronizing cluster resources.
5. After you change a log directory, a prompt appears reminding you to synchronize cluster resources from this node (Cluster log ODMs must be identical across the cluster). The cluster log destination directories as stored on this node will be synchronized to all nodes in the cluster.

Log destination directory changes will take effect when you synchronize cluster resources, or if the cluster is not up, the next time cluster services are restarted.

Note: Existing log files will not be moved to the new location.

Configuring Cluster Events

The HACMP/ES system is event-driven. An event is a change of status within a cluster. When the Cluster Manager detects a change in cluster status, it executes the designated script to handle the event and initiates any user-defined customized processing.

To configure cluster events, you indicate the script that handles the event and any additional processing that should accompany an event, as described below. You can define multiple customized pre- and post-event scripts.

Configuring Custom Cluster Events

To define your customized cluster event scripts, take the following steps.

1. To start system management for HACMP/ES, enter:

```
smit hacmp
```
2. Select **Cluster Configuration > Cluster Custom Modification > Define Custom Cluster Events**. SMIT displays the menu choices for adding, changing, or removing a custom event.
3. Select **Add a Custom Cluster Event** from the menu.

4. Enter the field values as follows:

Cluster Event Name	The name can have a maximum of 32 characters.
Cluster Event Description	Enter a short description of the event.
Cluster Event Script Filename	Enter the full pathname of the user-defined script to execute.

5. Press Enter to add the information to HACMPcustom in the local ODM.
6. Synchronize your changes across all cluster nodes by selecting the **Synchronize Cluster Resources** option off the **Cluster Resources** SMIT screen. Press F10 to exit SMIT.

Note: Synchronizing does not propagate the actual new or changed scripts; you must add these to each node manually.

Configuring Cluster Event Processing

Complete the following steps to set up or change the processing for an event. In this step you indicate to the Cluster Manager to use your customized pre- or post-event processing. You only need to complete these steps on a single node. The HACMP/ES software propagates the information to the other nodes when you synchronize the nodes.

To configure cluster event processing:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. Select **Cluster Configuration > Cluster Resources > Cluster Events > Change/Show Cluster Events** to display a list of cluster events and subevents.
3. Select an event or subevent that you want to configure and press Enter. SMIT displays the screen with the event name, description, and default event command shown in their respective fields.
4. Enter field values as follows:

Event Name	The name of the cluster event to be configured.
Description	A brief description of the event's function. This information cannot be changed.
Event Command	The full pathname of the command that processes the event. The HACMP/ES software provides a default script. If additional functionality is required, it is strongly recommended that you make changes by adding pre-or post-event processing of your own design, rather than by modifying the default scripts or writing new ones.

Notify Command

(optional) Enter the full pathname of a user-supplied script to run both before and after a cluster event. This script can notify the system administrator that an event is about to occur or has occurred.

The arguments passed to the command are: the event name, one keyword (either *start* or *complete*), the exit status of the event (if the keyword was “complete”), and the same trailing arguments passed to the event command.

Pre-Event Command

(optional) If you have defined custom cluster events, press F4 for the list. Or, enter the name of a custom-defined event to run before the cluster event command executes. This command provides pre-processing before a cluster event occurs.

The arguments passed to this command are the event name and the trailing arguments passed to the event command.

Remember that the Cluster Manager will not process the event until this pre-event script or command has completed.

Post-Event Command

(optional) If you have defined custom cluster events, press F4 for the list. Or, enter the name of the custom event to run after the cluster event command executes successfully. This script provides post-processing after a cluster event.

The arguments passed to this command are the event name, event exit status, and the trailing arguments passed to the event command.

Recovery Command

(optional) *This field is optional.* Enter the full pathname of a user-supplied script or AIX command to execute to attempt to recover from a cluster event command failure. If the recovery command succeeds and the retry count is greater than zero, the cluster event command is rerun. The arguments passed to this command are the event name and the arguments passed to the event command.

Recovery Counter

Enter the number of times to run the recovery command. Set this field to zero if no recovery command is specified, and to at least one (1) if a recovery command is specified.

5. Press Enter to add this information to the HACMP/ES ODM.
6. Return to the **Cluster Resources** screen, by pressing the F3 key, and synchronize your event customization by selecting the **Synchronize Cluster Resources** option. Press the F10 key to exit SMIT.

Note: All HACMP/ES event scripts are maintained in the `/usr/es/sbin/cluster/events` directory. The parameters passed to a script are listed in the script's header.

See Chapter 21, Monitoring an HACMP/ES Cluster, for a discussion of event emulation, which lets you emulate HACMP/ES event scripts without actually affecting the cluster.

Sample Custom Scripts

Two situations where it is useful to run custom scripts are illustrated here:

- Making **cron** jobs highly available
- Making print queues highly available.

Making cron jobs Highly Available

To help maintain the HACMP/ES environment, you need to have certain **cron** jobs execute only on the cluster node that currently holds the resources. If a **cron** job executes in conjunction with a resource or application, it is useful to have that **cron** entry fallover along with the resource. It may also be necessary to remove that **cron** entry from the **cron** table if the node no longer possesses the related resource or application.

The following example shows one way to use a customized script to do this:

The example cluster is a two node hot standby cluster where node1 is the primary node and node2 is the backup.

Node1 normally owns the shared resource group and application. The application requires that a **cron** job be executed once per day but only on the node that currently owns the resources.

To ensure that the job will be run even if the shared resource group and application have fallen over to node2, create two files as follows:

1. Assuming that the root user is executing the **cron** job, create a file `root.resource` and another called `root.noresource` in a directory on a non-shared filesystem on node1. Make these files resemble the **cron** tables that reside in the directory `/var/spool/crontabs`.

The `root.resource` table should contain all normally executed system entries and entries pertaining to the shared resource or application.

The `root.noresource` table should contain all normally executed system entries but no entries pertaining to the shared resource or application.

2. Copy the files to the other node so that both nodes have a copy of the two files.
3. On both systems, the following command should be executed at system startup:

```
crontab root.noresource
```

This will ensure that the **cron** table for root has only the “no resource” entries at system startup.

4. You can use either of two methods to activate the `root.resource` **cron** table. The first method is the simpler of the two.

- Execute *crontab root.resource* as the last line of the application start script. In the application stop script, the first line should then be *crontab root.noresource*. By executing these commands in the application start and stop scripts, you are ensured that they will activate and deactivate on the proper node at the proper time.
- Execute the **crontab** commands as a post_event to node_up_complete and node_down_complete.
 - Upon node_up_complete on the primary node, execute *crontab root.resources*.
 - On node_down_complete execute *crontab root.noresources*.

The takeover node must also use the event handlers to execute the correct **cron** table. Logic must be written into the node_down_complete event to determine if a takeover has occurred and to execute the *crontab root.resources* command. On a reintegration, a pre-event to node_up must determine if the primary node is coming back into the cluster and then execute a *crontab root.noresource* command.

Making Print Queues Highly Available

In the event of a fallover, the print jobs currently queued can be saved and moved over to the surviving node.

The print spooling system consists of two directories: **/var/spool/qdaemon** and **/var/spool/lpd/qdir**. One directory contains files containing the data (content) of each job. The other contains the files consisting of information pertaining to the print job itself. When jobs are queued, there are files in each of the two directories. In the event of a fallover, these directories do not normally fallover and thus the print jobs are lost.

The solution for this problem is to define two filesystems on a shared volume group. You might call these filesystems **/prtjobs** and **/prtdata**. When HACMP/ES starts, these filesystems are mounted over **/var/spool/lpd/qdir** and **/var/spool/qdaemon**.

Write a script to perform this operation as a post event to node_up. The script should do the following:

- Stop the print queues
- Stop the print queue daemon
- Mount **/prtjobs** over **/var/spool/lpd/qdir**
- Mount **/prtdata** over **/var/spool/qdaemon**
- Restart the print queue daemon
- Restart the print queues.

In the event of a fallover, the surviving node will need to do the following:

- Stop the print queues
- Stop the print queue daemon
- Move the contents of **/prtjobs** into **/var/spool/lpd/qdir**
- Move the contents of **/prtdata** into **/var/spool/qdaemon**
- Restart the print queue daemon
- Restart the print queues.

To do this, write a script called as a post-event to node_down_complete on the takeover. The script needs to determine if the node_down is from the primary node.

Part 3

Volume 1 Appendixes

These appendixes include worksheets to help you plan your cluster, information about planning for applications and HACMP, and installation instructions for cluster monitoring with Tivoli.

Appendix A, Planning Worksheets

Appendix B, Using the Online Cluster Planning Worksheet Program

Appendix C, Applications and HACMP

Appendix D, Installing and Configuring Cluster Monitoring with Tivoli

Appendix A Planning Worksheets

This appendix contains the following worksheets:

Worksheet	Purpose	Page
TCP/IP Networks	Use this worksheet to record the TCP/IP network topology for a cluster. Complete one worksheet per cluster.	A-3
TCP/IP Network Adapter	Use this worksheet to record the TCP/IP network adapters connected to each node. You need a separate worksheet for each node defined in the cluster, so begin by photocopying a worksheet for each node and filling in a node name on each worksheet.	A-5
Serial Networks	Use this worksheet to record the serial network topology for a cluster. Complete one worksheet per cluster.	A-7
Serial Network Adapter	Use this worksheet to record the serial network adapters connected to each node. You need a separate worksheet for each node defined in the cluster, so begin by photocopying a worksheet for each node and filling in the node name on each worksheet.	A-9
Shared SCSI-2 Differential or Differential Fast/Wide Disks	Use this worksheet to record the shared SCSI-2 Differential or Differential Fast/Wide disk configuration for the cluster. Complete a separate worksheet for each shared bus.	A-11
Shared IBM SCSI Disk Arrays	Use this worksheet to record the shared IBM SCSI disk array configurations for the cluster. Complete a separate worksheet for each shared SCSI bus.	A-13
Shared IBM Serial Storage Architecture Disk Subsystem	Use this worksheet to record the IBM 7131-405 or 7133 SSA shared disk configuration for the cluster.	A-15
Non-Shared Volume Group (Non-Concurrent Access)	Use this worksheet to record the volume groups and filesystems that reside on a node's internal disks in a non-concurrent access configuration. You need a separate worksheet for each volume group, so begin by photocopying a worksheet for each volume group and filling in a node name on each worksheet.	A-17
Shared Volume Group/Filesystem (Non-Concurrent Access)	Use this worksheet to record the shared volume groups and filesystems in a non-concurrent access configuration. You need a separate worksheet for each shared volume group, so begin by photocopying a worksheet for each volume group and filling in the names of the nodes sharing the volume group on each worksheet.	A-19
NFS-Exported Filesystem/Directory	Use this worksheet to record the filesystems and directories NFS-exported by a node in a non-concurrent access configuration. You need a separate worksheet for each node defined in the cluster, so begin by photocopying a worksheet for each node and filling in a node name on each worksheet.	A-21

Planning Worksheets

Worksheet	Purpose	Page
Non-Shared Volume Group (Concurrent Access)	Use this worksheet to record the volume groups and filesystems that reside on a node's internal disks in a concurrent access configuration. You need a separate worksheet for each volume group, so begin by photocopying a worksheet for each volume group and filling in a node name on each worksheet.	A-23
Shared Volume Group (Concurrent Access)	Use this worksheet to record the shared volume groups and filesystems in a concurrent access configuration. You need a separate worksheet for each shared volume group, so begin by photocopying a worksheet for each volume group and filling in the names of the nodes sharing the volume group on each worksheet.	A-25
Application	Use these worksheets to record information about applications in the cluster.	A-27
AIX Fast Connect	Use this worksheet to record Fast Connect resources	A-31
AIX Connections	Use this worksheet to record AIX Connections realm/service pairs	A-33
CS/AIX Communication Links	Use this worksheets to record information about CS/AIX Communications Links in the cluster.	A-35
Application Server	Use these worksheets to record information about application servers in the cluster.	A-37
Application Monitor (Process)	Use this worksheet to record information for configuring a process monitor for an application.	A-39
Application Monitor (User-defined)	Use this worksheet to record information for configuring a user-defined monitor method for an application.	A-41
Resource Group	Use this worksheet to record the resource groups for a cluster.	A-43
Cluster Event	Use this worksheet to record the planned customization for an HACMP/ES cluster event.	A-45

TCP/IP Networks Worksheet

Cluster ID _____

Cluster Name _____

Network Name	Network Type	Network Attribute	Netmask	Node Names
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____

Sample TCP/IP Networks Worksheet

Cluster ID 1

Cluster Name bivalves

Network Name	Network Type	Network Attribute	Netmask	Node Names
ether1	Ethernet	public	255.255.255.0	clam, mussel, oyster
token1	Token-Ring	public	255.255.255.0	clam, mussel, oyster
fddi1	FDDI	public	255.255.255.0	clam, mussel
socc1	SOCC	private	255.255.255.0	clam, mussel
atm1	ATM	private	255.255.255.0	clam, mussel

TCP/IP Network Adapter Worksheet

Node Name _____

Interface Name	Adapter IP Label	Adapter Function	Adapter IP Address	Network Name	Network Attribute	Boot Address	Adapter HW Address

Note: The SMIT Add an Adapter screen displays an **Adapter Identifier** field that correlates with the **Adapter IP Address** field on this worksheet.

Also, entries in the **Adapter HW Address** field should refer to the locally administered address (LAA), which applies only to the service adapter.

Sample TCP/IP Network Adapter Worksheet

Node Name nodea

Interface Name	Adapter Label	IP Address	Adapter Function	Adapter IP Address	Network Name	Network Attribute	Boot Address	Adapter HW Address
len0	nodea_en0		service	100.10.1.10	ether1	public		0x08005a7a7610
en0	nodea_boot1		boot	100.10.1.74	ether1	public		
en1	nodea_en1		standby	100.10.11.11	ether1	public		
tr0	nodea_tr0		service	100.10.2.20	token1	public		0x42005aa8b57b
tr0	nodea_boot2		boot	100.10.2.84	token1	public		
fi0	nodea_fi0		service	100.10.3.30	fddi1	public		
sl0	nodea_sl0		service	100.10.5.50	slip1	public		
css0	nodea_svc		service		hps1	private		
css0	nodea_boot3		boot		hps1	private		
at0	nodea_at0		service	100.10.7.10	atm1	private		0x0020481a396500
at0	nodea_boot1		boot	100.10.7.74	atm1	private		

Note: The SMIT Add an Adapter screen displays an **Adapter Identifier** field that correlates with the **Adapter IP Address** field on this worksheet.

Also, entries in the **Adapter HW Address** field should refer to the locally administered address (LAA), which applies only to the service adapter.

Serial Networks Worksheet

Cluster ID _____

Cluster Name _____

Network Name	Network Type	Network Attribute	Node Names
_____	_____	serial	_____
_____	_____	serial	_____
_____	_____	serial	_____
_____	_____	serial	_____
_____	_____	serial	_____
_____	_____	serial	_____
_____	_____	serial	_____
_____	_____	serial	_____
_____	_____	serial	_____

Note: RS232 serial lines, target mode SCSI-2 buses, and tmssa serial links do not use the TCP/IP protocol and do not require a netmask or an IP address.

Sample Serial Networks Worksheet

Cluster ID 1

Cluster Name clus1

Network Name	Network Type	Network Attribute	Node Names
rs232a	RS232	serial	nodea, nodeb
tm SCSI1	Target Mode SCSI	serial	nodea, nodeb
tmssa1	Target Mode SSA	serial	nodea, nodeb

Note: RS232 serial lines, target mode SCSI-2 buses, and tmssa serial links do not use the TCP/IP protocol and do not require a netmask or an IP address.

Serial Network Adapter Worksheet

Node Name _____

Slot Number	Interface Name	Adapter Label	Network Name	Network Attribute	Adapter Function
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service
_____	_____	_____	_____	serial	service

Note: Serial networks do not carry TCP/IP traffic; therefore, no boot addresses, adapter identifiers (IP addresses), or adapter hardware addresses are required to maintain keepalives and control messages between nodes.

Sample Serial Network Adapter Worksheet

Node Name	nodea				
Slot Number	Interface Name	Adapter Label	Network Name	Network Attribute	Adapter Function
SS2	/dev/tty1	nodea_tty1	rs232a	serial	service
08	scsi2	nodea_tm SCSI2	tm SCSI1	serial	service
01	tmssa1	nodea_tmssa1	tmssa1	serial	service

Note: Serial networks do not carry TCP/IP traffic; therefore, no boot addresses, adapter identifiers (IP addresses), or adapter hardware addresses are required to maintain keepalives and control messages between nodes.

Shared SCSI-2 Differential or Differential Fast/Wide Disks Worksheet

Note: Complete a separate worksheet for each shared SCSI-2 Differential bus or Differential Fast/Wide bus. Keep in mind that the IBM SCSI-2 Differential High Performance Fast/Wide adapter cannot be assigned SCSI IDs 0, 1, or 2. The SCSI-2 Differential Fast/Wide adapter cannot be assigned SCSI IDs 0 or 1.

Type of SCSI-2 Bus

SCSI-2 Differential _____ SCSI-2 Differential Fast/Wide _____

Host and Adapter Information

	Node A	Node B	Node C	Node D
Node Name	_____	_____	_____	_____
Slot Number	_____	_____	_____	_____
Logical Name	_____	_____	_____	_____

SCSI Device IDs on Shared Bus

	Node A	Node B	Node C	Node D
Adapter	_____	_____	_____	_____
First Shared Drive	_____			
Second Shared Drive	_____			
Third Shared Drive	_____			
Fourth Shared Drive	_____			
Fifth Shared Drive	_____			
Sixth Shared Drive	_____			

Shared Drives

Disk	Size	Logical Device Name			
		Node A	Node B	Node C	Node D
First	_____	_____	_____	_____	_____
Second	_____	_____	_____	_____	_____
Third	_____	_____	_____	_____	_____
Fourth	_____	_____	_____	_____	_____
Fifth	_____	_____	_____	_____	_____
Sixth	_____	_____	_____	_____	_____

Sample Shared SCSI-2 Differential or Differential Fast/Wide Disks Worksheet

Note: Complete a separate worksheet for each shared SCSI-2 Differential bus or Differential Fast/Wide bus. Keep in mind that the IBM SCSI-2 Differential High Performance Fast/Wide adapter cannot be assigned SCSI IDs 0, 1, or 2. The SCSI-2 Differential Fast/Wide adapter cannot be assigned SCSI IDs 0 or 1.

Type of SCSI-2 Bus

SCSI-2 Differential	SCSI-2 Differential Fast/Wide	X
----------------------------	--------------------------------------	---

Host and Adapter Information

	Node A	Node B	Node C	Node D
Node Name	nodea	nodeb		
Slot Number	7	7		
Logical Name	ascsi1	ascsi1		

SCSI Device IDs on Shared Bus

	Node A	Node B	Node C	Node D
Adapter	6	5		
First Shared Drive	3			
Second Shared Drive	4			
Third Shared Drive	5			
Fourth Shared Drive				
Fifth Shared Drive				
Sixth Shared Drive				

Shared Drives

Disk	Size	Logical Device Name			
		Node A	Node B	Node C	Node D
First	670	hdisk2	hdisk2		
Second	670	hdisk3	hdisk3		
Third	670	hdisk4	hdisk4		
Fourth					
Fifth					
Sixth					

Shared IBM SCSI Disk Arrays Worksheet

Note: Complete a separate worksheet for each shared SCSI disk array.

Host and Adapter Information

	Node A	Node B	Node C	Node D
Node Name	_____	_____	_____	_____
Slot Number	_____	_____	_____	_____
Logical Name	_____	_____	_____	_____

SCSI Device IDs on Shared Bus

	Node A	Node B	Node C	Node D
Adapter	_____	_____	_____	_____
First Array Controller	_____	_____	_____	_____
Second Array Controller	_____	_____	_____	_____
Third Array Controller	_____	_____	_____	_____
Fourth Array Controller	_____	_____	_____	_____

Shared Drives		Shared LUNs			
Size	RAID Level	Logical Device Name			
		Node A	Node B	Node C	Node D
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____

Array Controller and Path Information

	Array 1	Array 2
Array Controller Logical Name	_____	_____
Array Controller Logical Name	_____	_____
Disk Array Router Logical Name	_____	_____

Sample Shared IBM SCSI Disk Arrays Worksheet

This sample worksheet shows an IBM 7135 RAIDiant Disk Array configuration.

Note: Complete a separate worksheet for each shared SCSI disk array.

Host and Adapter Information

	Node A	Node B	Node C	Node D
Node Name	nodea	nodeb		
Slot Number	2	2		
Logical Name	ascsi1	ascsi1		

SCSI Device IDs on Shared Bus

	Node A	Node B	Node C	Node D
Adapter	14	15		
First Array Controller	3			
Second Array Controller	4			
Third Array Controller				
Fourth Array Controller				

Shared Drives		Shared LUNs			
Size	RAID Level	Logical Device Name			
		Node A	Node B	Node C	Node D
2GB	5	hdisk2	hdisk2		
2GB	3	hdisk3	hdisk3		
2GB	5	hdisk4	hdisk4		
2GB	5	hdisk5	hdisk5		

Array Controller and Path Information

RAIDiant 1	
Array Controller Logical Name	dac0
Array Controller Logical Name	dac1
Disk Array Router Logical Name	dar0

Shared IBM Serial Storage Architecture Disk Subsystems Worksheet

Host and Adapter Information

	Node A	Node B	Node C	Node D
Node Name	_____	_____	_____	_____
SSA Adapter Label	_____	_____	_____	_____
Slot Number	_____	_____	_____	_____
Dual-Port Number	_____	_____	_____	_____

SSA Logical Disk Drive

Logical Device Name

Node A	Node B	Node C	Node D
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

SSA Logical Disk Drive

Logical Device Name

Node A	Node B	Node C	Node D
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

Sample Shared IBM Serial Storage Architecture Disk Subsystems Worksheet

Host and Adapter Information

	Node A	Node B	Node C	Node D
Node Name	clam	mussel		
SSA Adapter Label	ha1, ha2	ha1, ha2		
Slot Number	2, 4	2, 4		
Dual-Port Number	a1, a2	a1, a2		

SSA Logical Disk Drive

		Logical Device Name			
Node A	Node B	Node C	Node D	Node C	Node D
hdisk2	hdisk2				
hdisk3	hdisk3				
hdisk4	hdisk4				
hdisk5	hdisk5				

SSA Logical Disk Drive

		Logical Device Name			
Node A	Node B	Node C	Node D	Node C	Node D
hdisk2	hdisk2				
hdisk3	hdisk3				
hdisk4	hdisk4				
hdisk5	hdisk5				

Non-Shared Volume Group Worksheet (Non-Concurrent Access)

Node Name _____

Volume Group Name _____

Physical Volumes _____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Filesystem Mount Point _____

Size (in 512-byte blocks) _____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Filesystem Mount Point _____

Size (in 512-byte blocks) _____

Sample Non-Shared Volume Group Worksheet (Non-Concurrent Access)

Node Name	clam
Volume Group Name	localvg
Physical Volumes	hdisk1

Logical Volume Name	locallv
Number of Copies of Logical Partition	1
On Separate Physical Volumes?	no
Filesystem Mount Point	/localfs
Size (in 512-byte blocks)	100000

Logical Volume Name	_____
Number of Copies of Logical Partition	_____
On Separate Physical Volumes?	_____
Filesystem Mount Point	_____
Size (in 512-byte blocks)	_____

Shared Volume Group/Filesystem Worksheet (Non-Concurrent Access)

	Node A	Node B	Node C	Node D
Node Names	_____	_____	_____	_____
Shared Volume Group Name	_____			
Major Number	_____	_____	_____	_____
Log Logical Volume Name	_____			
Physical Volumes	_____	_____	_____	_____
	_____	_____	_____	_____
	_____	_____	_____	_____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Filesystem Mount Point _____

Size (in 512-byte blocks) _____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Filesystem Mount Point _____

Size (in 512-byte blocks) _____

Sample Shared Volume Group/Filesystem Worksheet (Non-Concurrent Access)

	Node A	Node B	Node C	Node D
Node Names	trout	guppy		
Shared Volume Group Name		bassvg		
Major Number	24	24		
Log Logical Volume Name		bassloglv		
Physical Volumes	hdisk6	hdisk6		
	hdisk7	hdisk7		
	hdisk13	hdisk16		

Logical Volume Name basslv
Number of Copies of Logical Partition 3
On Separate Physical Volumes? yes
Filesystem Mount Point /bassfs
Size (in 512-byte blocks) 200000

Logical Volume Name _____
Number of Copies of Logical Partition _____
On Separate Physical Volumes? _____
Filesystem Mount Point _____
Size (in 512-byte blocks) _____

NFS-Exported Filesystem or Directory Worksheet (Non-Concurrent Access)

Resource Group _____

Network for NFS Mount _____

Filesystem Mounted Before IP Configured? _____

Filesystem/Directory to Export _____

Export Options (read-only, etc.) Refer to the *exports* man page for a full list of export options:

_____	_____	_____
_____	_____	_____
_____	_____	_____

Filesystem/Directory to Export _____

Export Options (read-only, etc.) Refer to the *exports* man page for a full list of export options:

_____	_____	_____
_____	_____	_____
_____	_____	_____

Filesystem/Directory to Export _____

Export Options (read-only, etc.) Refer to the *exports* man page for a full list of export options:

_____	_____	_____
_____	_____	_____
_____	_____	_____

Sample NFS-Exported Filesystem or Directory Worksheet (Non-Concurrent Access)

Resource Group rg1

Network for NFS Mount tr1

Filesystem Mounted Before IP Configured? true

Filesystem/Directory to Export /fs1

Export Options (read-only, root access, etc.) Refer to the *exports* man page for a full list of export options:

client access: client1 _____

root access: node 1, node 2 _____

mode: read/write _____

Filesystem/Directory to Export /fs 2

Export Options (read-only, root access, etc.) Refer to the *exports* man page for a full list of export options:

client access: client 2 _____

root access: node 3, node 4 _____

mode: read only _____

Non-Shared Volume Group Worksheet (Concurrent Access)

Node Name _____

Volume Group Name _____

Physical Volumes _____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Filesystem Mount Point _____

Size (in 512-byte blocks) _____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Filesystem Mount Point _____

Size (in 512-byte blocks) _____

Sample Non-Shared Volume Group Worksheet (Concurrent Access)

Node Name	clam
Volume Group Name	localvg
Physical Volumes	hdisk1

Logical Volume Name	locallv
Number of Copies of Logical Partition	1
On Separate Physical Volumes?	no
Filesystem Mount Point	/localfs
Size (in 512-byte blocks)	100000

Logical Volume Name	_____
Number of Copies of Logical Partition	_____
On Separate Physical Volumes?	_____
Filesystem Mount Point	_____
Size (in 512-byte blocks)	_____

Shared Volume Group/Filesystem Worksheet (Concurrent Access)

	Node A	Node B	Node C	Node D
Node Names	_____	_____	_____	_____
Shared Volume Group Name	_____			
Physical Volumes	_____	_____	_____	_____
	_____	_____	_____	_____
	_____	_____	_____	_____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Logical Volume Name _____

Number of Copies of Logical Partition _____

On Separate Physical Volumes? _____

Sample Shared Volume Group/Filesystem Worksheet (Concurrent Access)

	Node A	Node B	Node C	Node D
Node Names	trout	guppy		
Shared Volume Group Name		bassvg		
Physical Volumes	hdisk6	hdisk6		
	hdisk7	hdisk7		
	hdisk13	hdisk16		

Logical Volume Name basslv
Number of Copies of Logical Partition 3
On Separate Physical Volumes? yes
Filesystem Mount Point /bassfs
Size (in 512-byte blocks) 200000

Logical Volume Name _____
Number of Copies of Logical Partition _____
On Separate Physical Volumes? _____
Filesystem Mount Point _____
Size (in 512-byte blocks) _____

Logical Volume Name _____
Number of Copies of Logical Partition _____
On Separate Physical Volumes? _____

Application Worksheet

Application Name _____

Key Application Files

	Directory/Path	Filesystem	Location	Sharing
Executables:	_____	_____	_____	_____
Configuration Files:	_____	_____	_____	_____
Data Files/Devices:	_____	_____	_____	_____
Log Files/Devices:	_____	_____	_____	_____

Cluster Name: _____

Node Relationship:
(cascading/concurrent/
rotating)

Fallover Strategy: (P = primary; T = takeover)

Node: _____

Strategy: _____

Normal Start Commands/Procedures:

Verification Commands/Procedures:

Node Reintegration/Takeover Caveats:

Node	Reintegration/Takeover Caveats
_____	_____
_____	_____
_____	_____
_____	_____

Normal Stop Commands/Procedures:

Verification Commands/Procedures:

Node Reintegration/Takeover Caveats:

Node	Reintegration/Takeover Caveats

Sample Application Worksheet

Application Name _____

Key Application Files

	Directory/Path	Filesystem	Location	Sharing
Executables:	/app1/bin	/app1	internal	non-shared
Configuration Files:	/app1/config/one	/app1/config/one	external	shared
Data Files/Devices:	/app1lv1	NA	external	shared
Log Files/Devices:	/app1loglv1	NA	external	shared

Cluster Name: tetra

Node Relationship: cascading
(cascading/concurrent/
rotating)

Fallover Strategy: (P = primary; T = takeover)

Node: One Two Three Four

Strategy: P NA T1 T2

Normal Start Commands/Procedures:

- Verify that the app1 server group is running
- If the app1 server group is not running, as user app1_adm, execute app1 start -I One
- Verify that the app1 server is running
- If node Two is up, start (restart) app1_client on node Two

Verification Commands/Procedures:

- Run the following command: lssrc -g app1
- Verify from the output that daemon1, daemon2, and daemon3 are “Active”
- Send notification if not “Active”

Node Reintegration/Takeover Caveats:

Node	Reintegration/Takeover Caveats
One	NA
Two	NA
Three	Must restart the current instance of app1 with app1start -Ione -Ithree
Four	Must restart the current instance of app1 with app1start -Ione -Ifour

Sample Application Worksheet (continued)

Normal Stop Commands/Procedures:

- Verify that the app1 server group is running
- If the app1 server group is running, stop by app1stop as user app1_admin
- Verify that the app1 server is stopped
- If the app1 server is still up, stop individual daemons with the kill command

Verification Commands/Procedures:

- Run the following command: lssrc -g app1
- Verify from the output that daemon1, daemon2, and daemon3 are “Inoperative”

Node Reintegration/Takeover Caveats:

Node	Reintegration/Takeover Caveats
One	NA
Two	May want to notify app1_client users to log off
Three	Must restart the current instance of app1 with app1start -lthree
Four	Must restart the current instance of app1 with app1start -lfour

Note: In this sample worksheet, the server portion of the application, app1, normally runs on three of the four cluster nodes: nodes One, Three, and Four. Each of the three nodes is running its own app1 instance: one, three, or four. When a node takes over an app1 instance, the takeover node must restart the application server using flags for multiple instances. Also, because Node Two within this configuration runs the client portion associated with this instance of app1, the takeover node must restart the client when the client’s server instance is restarted.

AIX Fast Connect Worksheet

Cluster ID: _____

Cluster Name: _____

Resource Group	Nodes	Fast Connect Resources
----------------	-------	------------------------

_____	_____	_____

Resource Group	Nodes	Fast Connect Resources
----------------	-------	------------------------

_____	_____	_____

Sample AIX Fast Connect Worksheet

Cluster ID: 2

Cluster Name: cluster2

Resource Group	Nodes	Fast Connect Resources
rg1	NodeA, NodeC	FS1%f%/smbtest/fs1 LPT1%p%printq _____ _____ _____ _____ _____ _____

Resource Group	Nodes	Fast Connect Resources
rg2	Node B, Node D	FS2%f%/smbtest/fs2 LPT2%p%printq _____ _____ _____ _____ _____ _____

AIX Connections Worksheet

Cluster: _____

Resource Group	Nodes	Realm (NB,NW,AT)	Service Name	Service Type (file,print,term,nvt,atls)
_____	_____	_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____

Resource Group	Nodes	Realm (NB,NW,AT)	Service Name	Service Type (file,print,term,nvt,atls)
_____	_____	_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____

Sample AIX Connections Worksheet

Cluster: cluster1

Resource Group	Nodes	Realm (NB,NW,AT)	Service Name	Service Type (file,print,term,nvt,atls)
rg1	clam, mussel, oyster	NB	printnb	print
		NB	term1	term
		NB	clamnb	file
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____

Resource Group	Nodes	Realm (NB,NW,AT)	Service Name	Service Type (file,print,term,nvt,atls)
_____	_____	_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____
		_____	_____	_____

CS/AIX Communications Links Worksheet

Cluster ID _____

Cluster Name _____

**Communications Link
Name:**

Resource Group:

Nodes:

DLC Name:

Port:

Link Station:

Service:

**Communications Link
Name:**

Resource Group:

Nodes:

DLC Name:

Port:

Link Station:

Service:

Sample CS/AIX Communications Links Worksheet

Cluster ID __ 1 _____

Cluster Name __ cluster1 _____

Communications Link Link1

Name: _____

Resource Group: rg1 _____

Nodes: nodea, nodeb _____

DLC Name: profile1 _____

Port: port1 _____

Link Station: station1 _____

Service: /tmp/service1.sh _____

Communications Link

Name: _____

Resource Group: _____

Nodes: _____

DLC Name: _____

Port: _____

Link Station: _____

Service: _____

Application Server Worksheet

Cluster ID _____

Cluster Name _____

Note: Use full pathnames for all user-defined scripts.

Server Name: _____

Start Script: _____

Stop Script: _____

Server Name: _____

Start Script: _____

Stop Script: _____

Server Name: _____

Start Script: _____

Stop Script: _____

Server Name: _____

Start Script: _____

Stop Script: _____

Sample Application Server Worksheet

Cluster ID 1

Cluster Name clus1

Server Name: imagedemo

Start Script: /usr/es/sbin/cluster/utis/start_imagedemo

Stop Script: /usr/es/sbin/cluster/utis/stop_imagedemo

Server Name: _____

Start Script: _____

Stop Script: _____

Server Name: _____

Start Script: _____

Stop Script: _____

Server Name: _____

Start Script: _____

Stop Script: _____

Application Monitor Worksheet (Process Monitor)

Cluster ID

Cluster Name

Application Server Name

**Can Application Be
Monitored with Process
Monitor?***

Yes / No (*If no, go to User-Defined worksheet*)

Processes to Monitor

Process Owner

Instance Count

Stabilization Interval

Restart Count

Restart Interval

**Action on Application
Failure**

Notify Method

Cleanup Method

Restart Method

*Some applications, for instance shell scripts, cannot be monitored with a process monitor. For those cases, a user-defined monitor may be appropriate. See Chapter 18, Configuring an HACMP/ES Cluster, for full details on defining both types of application monitors.

Sample Application Monitor Worksheet (Process Monitor)

Cluster ID 1

Cluster Name Cluster1

Application Server Name imagedemo

Can this Application Be Monitored with Process Monitor? yes

Processes to Monitor imserv

Process Owner root

Instance Count 1

Stabilization Interval 30

Restart Count 3

Restart Interval 95

Action on Application Failure fallover

Notify Method /usr/es/sbin/cluster/events/notify_imagedemo

Cleanup Method /usr/es/sbin/cluster/utills//events/stop_imagedemo

Restart Method /usr/es/sbin/cluster/utills/events/start_imagedemo

Application Monitor Worksheet (User-Defined Monitor)

Cluster ID:

Cluster Name:

Application Server Name

Monitor Method

Monitor Interval

Hung Monitor Signal

Stabilization Interval

Restart Count

Restart Interval

**Action on Application
Failure**

Notify Method

Cleanup Method

Restart Method

Sample Application Monitor Worksheet (User-defined Monitor)

Cluster ID: 1

Cluster Name: Cluster1

Application Server Name	imagedemo
Monitor Method	/usr/es/sbin/cluster/events/utills/monitor_imagedemo
Monitor Interval	60
Hung Monitor Signal	9
Stabilization Interval	30
Restart Count	3
Restart Interval	280
Action on Application Failure	notify
Notify Method	/usr/es/sbin/cluster/events/utills/notify_imagedemo
Cleanup Method	/usr/es/sbin/cluster/events/utills/stop_imagedemo
Restart Method	/usr/es/sbin/cluster/events/utills/start_imagedemo

Resource Group Worksheet

Cluster ID _____ Cluster Name _____

Resource Group Name _____

Node Relationship _____

Participating Node Names _____

Filesystems _____

Filesystems to Export _____

Filesystems to NFS Mount _____

Volume Groups _____

Raw Disks _____

AIX Connections Services _____

AIX Fast Connect Services _____

Application Servers _____

Highly Available Communication Links _____

Inactive Takeover _____

Cascading without Fallback Enabled _____

Sample Resource Group Worksheet

Cluster ID: 1 **Cluster Name: clus1**

Resource Group Name rotgrp1

Node Relationship rotating

Participating Node Names clam, mussel, oyster

Filesystems /sharedfs1

Filesystems/Directories to Export

Filesystems/Directories to NFS Mount /sharedvg1

Volume Groups

Raw Disks

AIX Connections Services

AIX Fast Connect Resources

Highly Available Communication Links

Application Servers imagedemo

Inactive Takeover false

Cascade without Fallback Enabled false

Cluster Event Worksheet

Note: Use full pathnames for all Cluster Event Methods, Notify Commands, and Recovery commands.

Cluster ID	_____
Cluster Name	_____
Cluster Event Description	_____
Cluster Event Method	_____
Cluster Event Name	_____
Event Command	_____
Notify Command	_____
Pre-Event Command	_____
Post-Event Command	_____
Event Recovery Command	_____
Recovery Counter	_____
Cluster Event Name	_____
Event Command	_____
Notify Command	_____
Pre-Event Command	_____
Post-Event Command	_____
Event Recovery Command	_____
Recovery Counter	_____

Sample Cluster Event Worksheet

Note: Use full pathnames for all user-defined scripts.

Cluster ID	1
Cluster Name	bivalves
Cluster Event Name	node_down_complete
Event Command	
Notify Command	
Pre-Event Command	
Post-Event Command	/usr/local/wakeup
Event Recovery Command	
Recovery Counter	
Cluster Event Name	_____
Event Command	_____
Notify Command	_____
Pre-Event Command	_____
Post-Event Command	_____
Event Recovery Command	_____
Recovery Counter	_____

Appendix B Using the Online Cluster Planning Worksheet Program

This appendix covers how to use the web-based online worksheets provided with the HACMP/ES software in the `/usr/es/lpp/cluster/samples/worksheets` directory. The online worksheet program is a tool to aid you in planning and entering your cluster configuration and then applying the configuration directly to your cluster.

To use the online worksheets, you must transfer the worksheet program files to a PC-based system equipped with the appropriate web browser (see below).

This appendix contains instructions for using the online worksheets correctly, and references to the appropriate chapters in this book for detailed information about each stage of the cluster planning process. The worksheet program includes its own online help topics as well.

Note: Before you start, and while you are using the online worksheet program, refer to the planning information in *Enhanced Scalability Installation and Administration Guide, Vol. 1* for details about planning each component of your cluster.

Online Cluster Planning Worksheets—Overview

The HACMP online cluster planning “worksheets” are actually a series of panels presented to you on a PC through a web browser. Each panel contains fields in which you can specify the basic components of your HACMP cluster topology and resource configuration. As you progress through the panels, you can save information you have entered, create an HTML report to print out or e-mail, or go back to earlier panels and make changes. The online planning panels present logical default choices whenever possible. When appropriate, the worksheet program prevents you from entering nonfeasible configuration choices by presenting only the options that work based on what you have already entered, or by presenting a message if you make an invalid choice.

The panels are based in part on the paper worksheets in Appendix A, Planning Worksheets, but the information may be consolidated or ordered in a slightly different way than the paper worksheets. Therefore, the worksheet instructions in *Enhanced Scalability Installation and Administration Guide, Vol. 1* chapters apply only to the paper worksheets.

As soon as you complete all of the planning worksheets, you can actually apply the new configuration to your cluster nodes. To do this, you create an AIX file of your configuration data, and then transfer it via FTP to an AIX cluster node to apply the configuration to your cluster.

Installing and Viewing the Worksheet Program

The online planning worksheet program files are packaged with your other HACMP/ES filesets, in the samples directory. To use the worksheets, you must transfer the fileset to a PC-based system running an up-to-date version of the Microsoft Internet Explorer™ web browser.

Note: At the time of publication of this manual, *Microsoft Internet Explorer, version 4.0 or higher*, is the required browser for the online worksheets. Check the most recent HACMP/ES product README file for any updates to this requirement.

To begin using the online cluster planning worksheets:

1. Install the HACMP/ES software.
2. In the **usr/es/lpp/cluster/samples/worksheets** directory, find two files: **worksheets.html** and **worksheets.jar**.
3. Transfer these two files (via FTP or another method) to an appropriate folder of your choice on the PC-based system.

Note: You must specify binary mode for your FTP session.

4. Set the CLASSPATH variable to the **worksheets.jar** pathname, as follows:

Windows 95: You must edit the **autoexec.bat** file to include the CLASSPATH variable. The **autoexec.bat** file is usually located in the root (C:) directory. Open the file in an editor such as Notepad, or open a DOS window and type `edit autoexec.bat` at the prompt.

Edit the **autoexec.bat** file by adding the following line:

```
set CLASSPATH=c:\<folder name>\worksheets.jar
```

Windows NT: Go to Control Panel > System Properties > Environment. Add CLASSPATH as the *variable*, and the full pathname for **worksheets.jar** as the *value*.

5. In your browser, open **worksheets.html**.
You should see the first worksheet panel (shown on page B-3), with fields for entering the cluster name and ID.

Note: It is recommended that you view the worksheets in full screen mode.

The Online Worksheet Format

The online planning worksheet panels are arranged in a sequence that makes logical sense for planning your cluster topology and resources. In many cases, you must follow this sequence for correct configuration.

As you proceed through the worksheet panels, refer to the instructions on the following pages and on the worksheet panels, and refer to the specified chapters in Volume 1 of this guide for important information about how to plan each component. In addition, refer to the online help topics for details about each worksheet panel.

Panels, Topic Tabs, and Buttons

The first panel you see when you open the worksheet program is the Cluster panel shown below. In this initial panel, you see the tabs that take you to individual worksheet panels for data entry. You also see several buttons. Most of the tabs, buttons, and other worksheets are disabled until you enter the cluster name and ID or open an existing set of worksheets with data already in them.

The screenshot shows the 'Cluster Planning Worksheets' window. At the top right, it displays 'Name: cluster1' and 'Type: HACMP'. Below this is a tabbed interface with 'Cluster' selected. A text box contains instructions: 'To create a new set of cluster worksheets, enter in cluster information and press Add. To open an existing set of works... press Open Existing... and select the desired cluster's name.' The form includes fields for 'Cluster Name' (cluster1), 'Cluster ID' (12345), 'Cluster Type' (radio buttons for HACMP and HACMPRES), 'Author', 'Company', and 'Last Updated'. 'Add' and 'Open Existing...' buttons are at the bottom of the form. At the very bottom of the window are tabs for 'Topology', 'Disk', 'Resource', 'Cluster Notes', and 'Help', and buttons for 'Clear All', 'Print', 'Save', and 'Create Configurat...'.

Tabs

The tabs on the bottom and top of the panel indicate at all times which worksheet panels are enabled; the panel currently selected is indicated by bold type on the appropriate tabs. The tabs at the bottom of the screen indicate the major topics, and those at the top indicate subtopics within each major topic.

Bottom Tabs

The tabs at the bottom of your screen indicate three major planning topics. You should proceed with your configuration in the sequence of the tabs as follows:

1. **Topology**
2. **Disk**
3. **Resources**

The Cluster Notes tab brings to you a screen for miscellaneous user/system administrator notes.

The Help tab brings you to a directory of help topics for specific worksheets.

Top Tabs

When you select a major topic from the bottom tabs, a series of subtopic tabs at the top of your screen are enabled. Again, these tabs are presented from left to right in the order that works best for planning a cluster.

The entire set of planning sheets follows this hierarchy of topics and subtopics:

Topology:	Cluster, Node, Network, Global Network (<i>active for HACMP/ES type only</i>), Adapters
Disk:	Disks, Volume Groups, Logical Volumes, NFS Exports
Resource:	Cluster Events, Applications, Application Servers, Resource Groups, Resource Associations
Cluster Notes:	Space for administrator notes
Help:	All help topics

Worksheet Panel Buttons

Within each panel, buttons are used for the following actions:

Add	Updates the configuration with the information in the current entry field and makes the information available to subsequent panels.
Modify	Allows you to make changes in any entry you have made in the current worksheet panel. First, in the lower window, select what entry you wish to change. The data for that entry is displayed in the entry fields above. Make the desired changes and press the Modify button.
Delete	Deletes the selected entry field data.

Base Panel Buttons

On the base panel, which is visible at all times below the topic tabs, the following buttons are available:

Clear All	Deletes all entries in the entire set of worksheets you have open
Create Report	Saves to an HTML file all of the configuration data you have entered thus far for the set of worksheets currently open. You can then open this file in a browser and print it or distribute it via e-mail to others.
Save	Writes the cluster configuration entered thus far to the PC disk, to a specified worksheet (.ws) file.*

Create Configuration Creates the AIX file that you will transfer to an AIX cluster node to apply the configuration data to the cluster.

Select the **Help** tab at any time to get specific information about entering data in a given worksheet panel.

*Due to differences in older and newer versions of **msjava.dll**, the behavior of the Save option may vary. In newer versions of **msjava.dll**, the **.ws** files can be saved only to the desktop.

Entering Cluster Configuration Data

If you are familiar with the planning information in the earlier chapters of this guide, and have your cluster configuration preferences mapped out in a diagram or other format, you are ready to begin entering data into the online worksheet panels as follows.

Note: If you are not in full screen mode, you may find that scrolling down and back up may cause the top (subtopic) tabs of the panel to disappear from view. They reappear when you press the Add button, or if you go to another panel and back again. *Do not use the browser's Refresh button*, as this will clear the panel of the information you have entered.

Stage 1: Topology Planning

The first step is to define the topology of your cluster: the framework of cluster nodes and the networks that connect them.

Cluster

When you open the worksheet program, the first thing you see is the Cluster screen.

In the Cluster screen, you name your cluster and assign it an ID. You also specify which product subsystem you are using: HACMP or HACMP/ES. After you have entered the name and ID, you can choose from the subtopics on the top tabs. The Author and Company fields are optional.

You can also open the worksheets for an existing cluster you have already started and saved.

To complete the Cluster worksheet:

1. Fill in the cluster name. This can be any combination of 31 or fewer alphanumeric characters and underscores.
2. Fill in the cluster ID. The ID can be any positive integer up to 99999.

Note: Be sure the cluster name and ID do not duplicate those of another cluster at your site.

3. Check the appropriate box (HACMP or HACMP/ES) for the product type (subsystem) you are using.

4. Press the Add button.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Assigning a Name and ID to Your Cluster, page 3-5.
---	--

Note: If you want to change what you have entered in the Cluster panel, just change your text and press Add again, *except for Type* (HACMP or HACMP/ES). If you need to change the Type designation after you have hit Add, use the Clear All button to start over.

Nodes

In the Nodes screen, you specify all nodes that will participate in the cluster. The maximum number of nodes is eight (HACMP) or 32 (HACMP/ES). The minimum is two for either case.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Initial Cluster Planning, page 3-2
---	------------------------------------

To specify cluster nodes, enter each node name and press the Add button. When you have finished this panel, go on to the Networks panel.

Networks

In this screen, you define the networks in your cluster. The drop-down list for attributes presents only the attributes that are possible for the network type you select. For example, if your network type is Ethernet, you can choose either *public* or *private* as an attribute. For an RS232 serial network, only *serial* is presented.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 4, Planning Cluster Network Connectivity
---	--

Global Network

In this screen (HACMP/ES only), you define the global network. The dropdown menu presents the networks you entered in the preceding panel.

1. Specify a unique name for your global network.
2. Specify the networks to be included in the global network.

Note: The networks in a global network must be of the same type.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 4, Planning Cluster Network Connectivity
---	--

Adapters

In this screen, you enter information about the adapters for each network you have defined. The Networks drop-down list presents a list of the networks defined so far.

The worksheet program allows you to enter only the adapter functions and hardware addresses that work with the types of networks you have defined.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 4, Planning Cluster Network Connectivity
---	--

Disk Planning

Once you have defined the major components of your cluster topology, you plan your shared disk configuration.

Disks

In this screen, you enter information about the disks you want to use in your cluster.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 5, Planning Shared Disk Devices
Also see:	<i>HACMP for AIX: Concepts and Facilities</i>

Volume Groups

In this screen, you specify the volume groups for your cluster.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 6, Planning Shared LVM Components
---	---

Logical Volumes

In the Logical Volumes screen, you specify logical volumes and the volume groups with which they are associated.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 6, Planning Shared LVM Components
Also see:	<i>AIX System Management Guide: Operating System and Devices</i>

NFS Exports

Here, you specify which filesystems, if any, need to be exported, so other nodes can NFS mount them.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 6, Planning Shared LVM Components
---	---

Stage 3: Resource Planning

Now that you have planned your nodes, networks, and disk components for your cluster, you move on to planning event processing and configuring the cluster *resources* you wish to make highly available under HACMP/ES.

Cluster Events

In this screen, you specify events and associated scripts. You type in the command or script path for each event. You can enter multiple entries per event per node. You cannot enter a duplicate command for the same event.

In <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 8, Cluster Events: Tailoring and Creating
---	---

Applications

The Applications screen asks you to assign a name to each application and enter its directory path and filesystem.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 3, Initial Cluster Planning
---	-------------------------------------

Application Servers

Here, you specify the application server name, and then choose from the picklist of applications you specified in the previous screen. You then specify the full path locations of your application start and stop scripts. In addition, you can write notes about your scripts in the space provided.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 3, Initial Cluster Planning
---	-------------------------------------

Resource Groups

In this screen, you assign names to your resource groups and specify the participating nodes and their position (priority), if appropriate, in the resource chain.

The Increase Priority and Decrease Priority buttons become visible when you highlight the node name. Note that for concurrent and rotating resource groups, priority information is not valid.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 3, Initial Cluster Planning
---	-------------------------------------

Resource Associations

In this screen, you specify which individual resources—such as filesystems, IP labels, volume groups, and application servers—are to be part of each resource group you defined in the previous screen.

In the <i>Enhanced Scalability Installation and Administration Guide, Vol. 1</i> , see:	Chapter 3, Initial Cluster Planning
---	-------------------------------------

Cluster Notes

In this screen, you have space to note any special details about the cluster or your cluster planning process for future reference.

Applying Worksheet Data to AIX Cluster Nodes

When you have completed all the worksheets, are satisfied with your decisions, and have saved the worksheets, you can create an AIX file to configure your actual cluster. You then transfer the configuration file, via FTP, to your AIX cluster nodes.

Prerequisites

Before you can transfer your new file to a cluster node, the following conditions must be met:

- Your HACMP/ES software must be installed.
- All hardware devices that you specified for your cluster configuration must be in place.
- If you are replacing an existing configuration, any current cluster ODM information should be retained in a snapshot.
- Cluster services must be stopped on all nodes.

Creating the AIX Configuration File

To create an AIX file of your cluster configuration data:

1. Press the Create Configuration button located on the base panel.
(Note that this button is enabled only after you have defined the cluster nodes.)
The Create AIX Cluster Configuration dialog box appears.
2. Save your configuration file as the default (**cluster.conf**) or any name you choose.

Transferring the AIX File to the Cluster

You now transfer the configuration data file from the PC-based system to one of the AIX cluster nodes, in order to apply the configuration to the cluster.

1. FTP your configuration (.conf) file to one of your cluster nodes.
2. On the cluster node, run the **cl_opsconfig** command as follows:

```
/usr/es/sbin/cluster/utilities/cl_opsconfig <your configuration file>
```

The **cl_opsconfig** utility automatically performs a synchronization, including verification, of the configuration. During this process, you see a series of messages indicating the events taking place and any warnings or errors.

Boot Adapter Warnings

The **cl_opsconfig** utility adds all boot adapters before adding any service adapters. Because of this, you see a series of warnings indicating “There is no service interface” for each boot adapter as it is added. You can ignore these warnings if you have configured the proper service adapters, as **cl_opsconfig** will add them later in the process, resolving the false error.

Viewing Error Messages

You can view the **cl_opsconfig** error messages on the screen, or redirect them to a log file. If you wish to redirect the standard error information to another file, add the symbols `2>` and specify an output file, as in the following:

```
/usr/es/sbin/cluster/utilities/cl_opsconfig <configuration file> 2> <output file>
```

Note that redirecting *all* output (standard output and standard error) is not recommended.

If errors are detected, go back to the worksheet program to fix the problems, and repeat the process of creating the configuration file, transferring it to the cluster node, and running **cl_opsconfig**.

Where You Go From Here

When your configuration file has been transferred and **cl_opsconfig** has run successfully without reporting configuration errors, your cluster has a basic working HACMP/ES configuration.

You can now proceed to additional configuration and customization tasks—for example, configuring AIX Error Notification, run time parameters, custom scripts, IP address takeover, and so on. For more information, refer to the appropriate chapters in the installation and administration sections of this guide.

Appendix C Applications and HACMP

This appendix addresses some of the key issues to consider when making your applications highly available under HACMP. The information here is general enough to be useful to HACMP and HACMP/ES users.

For details on the planning for and configuring of applications, see Chapter 3, Initial Cluster Planning, and Chapter 18, Configuring an HACMP/ES Cluster.

For a general discussion of issues in keeping a cluster running on a 7x24 basis, also see Appendix H, 7x24 Maintenance, in Volume 2 of this guide.

Overview

Besides understanding the hardware and software needed to make a cluster highly available, you will need to spend some time on *application* considerations when planning your HACMP environment. The goal of clustering is to keep your important applications available despite any single point of failure. To achieve this goal, it is important to consider the aspects of an application that make it recoverable under HACMP.

There are few hard and fast *requirements* that an application must meet to recover well under HACMP. For the most part, there are simply good practices that can head off potential problems. Some required characteristics, as well as a number of suggestions, are discussed here. These are grouped according to key points that should be addressed in all HACMP environments. This appendix covers the following application considerations:

- *Automation*—making sure your applications start and stop without user intervention
- *Dependencies*—knowing what factors outside HACMP affect the applications
- *Interference*—knowing that applications themselves can hinder HACMP's functioning
- *Robustness*—choosing strong, stable applications
- *Implementation*—using appropriate scripts, file locations, and cron schedules

At the end of this appendix, you will find two examples of popular applications—Oracle Database™ and SAP R/3™—and some issues to consider when implementing these applications in an HACMP environment.

Application Automation: Minimizing Manual Intervention

One key requirement for an application to function successfully under HACMP is that the application be able to start and stop without any manual intervention.

Application Start Scripts

You should create a start script that completely starts the application. Configure HACMP to call this script at cluster startup to initially bring the application online. Since the cluster daemons call the start script, there is no option for interaction. Additionally, upon an HACMP failover, the recovery process calls this script to bring the application on line on a standby node. This allows for a fully automated recovery.

Keep in mind that this application start script may need to take additional action to prepare the cluster to bring the application online. The start script will be called by HACMP as the “root” user. It may be necessary to change to a different user in order to start the application. The **su** command can accomplish this. Also, it may be necessary to run **nohup** on commands that are started in the background and have the potential to be terminated upon exit of the shell.

For example, an HACMP cluster node may be a client in a Network Information Service (NIS) environment. If this is the case, and you need to use the **su** command to change user id, there must be a route to the NIS master at all times. In the event that a route doesn't exist, and the **su** is attempted, the application script hangs. You can avoid this by enabling the HACMP cluster node to be an NIS slave. That way a cluster node has the ability to access its own NIS map files to validate a user ID.

Another good practice in application start scripts is to check the return code upon exiting a script. If the return code is not zero, an error may have occurred in starting that should be addressed. If a non-zero return code is passed back to HACMP, the *event_error* event is run and the cluster enters an error state. This check alerts administrators that the cluster is not functioning properly.

The start script should also check for the presence of required resources or processes. This will ensure an application can start successfully. If the necessary resources are not available, a message can be sent to the administration team to correct this and restart the application.

Keep in mind that the start script may be run after a primary node has failed. There may be recovery actions necessary on the backup node in order to restart an application. This is common in database applications. Again, the recovery must be able to run without any interaction from administrators.

Application Stop Scripts

The most important aspect of an application stop script is that it completely stop an application. Failure to do so may prevent HACMP from successfully completing a takeover of resources by the backup nodes. In stopping, the script may need to address some of the same concerns the start script addresses, such as NIS and the **su** command.

The application stop script should use a phased approach. The first phase should be a graceful attempt to stop the processes and release any resources. If processes refuse to terminate, the second phase should be used to forcefully ensure all processing is stopped. Finally, a third phase can use a loop to repeat any steps necessary to ensure that the application has terminated completely.

Be sure that your application stop script exits with the value 0 (zero) when the application has been successfully stopped. In particular, examine what happens if you run your stop script when the application is already stopped. Your script must exit with zero in this case as well. If your stop script exits with a different value, that tells HACMP that the application is still running, though possibly in a damaged state. The `event_error` event will be run and the cluster will enter an error state. This check alerts administrators that the cluster is not functioning properly.

Note: Keep in mind that HACMP allows six minutes by default for events to complete processing. A message indicating the cluster has been in reconfiguration too long appears until the cluster completes its reconfiguration and returns to a stable state. This warning may be an indication that a script is hung and requires manual intervention. If this is a possibility, you may wish to consider stopping an application manually before stopping HACMP.

If desired, you can alter the time period before the `config_too_long` event is invoked. See the *HACMP for AIX Troubleshooting Guide* for more information.

Application Tier Issues

Often, applications are of a multi-tier architecture. The first tier may be a database, the second tier an application/login tier and the third a client. You must consider all tiers of an architecture if one or more is made highly available through the use of HACMP.

For example, if the database is made highly available, and a failover occurs, consider whether actions should be taken at the higher tiers in order to automatically return the application to service. If so, it may be necessary to stop and restart application or client tiers. This can be facilitated in one of two ways. One way is to run **clinfo** on the tiers, the other is to use the **rsh** command.

Using the Clinfo API

clinfo is the cluster information daemon. You can write a program using the Clinfo API to run on any tiers that would stop and restart an application after a failover has completed successfully. In this sense, the tier, or application, becomes “cluster aware,” responding to events that take place in the cluster. See the manual *HACMP for AIX: Programming Client Applications* for more detail on the Clinfo API.

Using Pre- and Post-Event Scripts

Another way to address the issue of multi-tiered architectures is to use pre- and post-event scripts around a cluster event. These scripts would call the **rsh** command to stop and restart the application. Keep in mind that the use of the **rsh** command may require a loosening of security that is unacceptable for some environments.

Another way to address the **rsh** security issue is by using Kerberos as an authentication method. If you choose this method, Kerberos must be in place prior to the HACMP installation and configuration. IBM supports Kerberos on IBM Scalable POWERParallel (RS/6000 SP)

systems. For more information about configuring Kerberos, see Appendix F of the *HACMP for AIX Installation Guide* (for HACMP for AIX on an RS/6000 SP) or Configuring Kerberos Security with HACMP/ES for AIX on page 18-31 of this manual.

Application Dependencies

In many cases, applications depend on more than data and an IP address. For the success of any application under HACMP, it is important to know what the application should *not* depend upon in order to function properly. This section outlines many of the major dependency issues. Keep in mind that these dependencies may come from outside the HACMP and application environment. They may be incompatible products or external resource conflicts. Look beyond the application itself to potential problems within the enterprise.

Locally Attached Devices

Locally attached devices can pose a clear dependency problem. In the event of a fallover, if these devices are not attached and accessible to the standby node, an application may fail to run properly. These may include a CD-ROM device, a tape device, or an optical juke box. Consider whether your application depends on any of these and if they can be shared between cluster nodes.

Hard Coding

Anytime an application is hard coded to a particular device in a particular location, there is the potential for a dependency issue. For example, the console is typically assigned as `/dev/tty0`. Although this is common, it is by no means guaranteed. If your application assumes this, ensure that all possible standby nodes have the same configuration.

Hostname Dependencies

Some applications are written to be dependent on the AIX hostname. They issue a command in order to validate licenses or name filesystems. The hostname is not an IP address label. The hostname is specific to a node and is not failed over by HACMP. It is possible to manipulate the hostname, or use hostname aliases, in order to trick your application, but this can become cumbersome when other applications, not controlled by HACMP, also depend on the hostname.

Software Licensing

Another possible problem is software licensing. Software can be licensed to a particular CPU ID. If this is the case with your application, it is important to realize that a fallover of the software will not successfully restart. You may be able to avoid this problem by having a copy of the software resident on all cluster nodes. Know whether your application uses software that is licensed to a particular CPU ID.

Application Interference

Sometimes an application or an application environment may interfere with the proper functioning of HACMP. An application may execute properly on both the primary and standby nodes. However, when HACMP is started, a conflict with the application or environment could arise that prevents HACMP from functioning successfully.

Software Using IPX/SPX Protocol

A conflict may arise between HACMP and any software that binds a socket over a network interface. An example is the IPX/SPX protocol. When active, it binds an interface and prevents HACMP from properly managing the interface. Specifically, for ethernet and token ring, it inhibits the hardware address takeover from completing successfully. A “device busy” message appears in the HACMP logs. The software using IPX/SPX must be either completely stopped or not used in order for hardware address takeover to work.

Products Manipulating Network Routes

Additionally, products that manipulate network routes can keep HACMP from functioning as it was designed. These products can find a secondary path through a network that has had an initial failure. This may prevent HACMP from properly diagnosing a failure and taking appropriate recovery actions.

AIX Connections, AIX Fast Connect, and CS/AIX

You can reduce the problem of conflict with certain protocols, and the need for manual intervention, if you are using AIX Connections, AIX Fast Connect, or Communications Server for AIX (CS/AIX) to share resources. The protocols handled by these applications can easily be made highly available because of their integration with HACMP.

AIX Connections is a network operating software that enables sharing of resources between AIX workstations and clients running other operating systems such as Windows NT, OS/2, and Macintosh. AIX Connections is already integrated with HACMP so the IPX/SPX, NetBEUI, and AppleTalk protocols handled by AIX Connections can be easily configured as highly available resources in the cluster. The protocols can then be taken over in the event of node or adapter failure. For example, in the case of a NetWare client using the IPX/SPX protocol, if AIX Connections is configured, HACMP steps up in the event of an adapter failure: it stops communications on the port, frees up the bind on the socket—thereby allowing the address takeover to proceed—and restarts the communication.

AIX Fast Connect software is integrated with HACMP in a similar way, so that it can be configured as a highly available resource. AIX Fast Connect allows you to share resources between AIX workstation and PCs running Windows, DOS, and OS/2 operating systems. Fast Connect supports the NetBIOS protocol over TCP/IP.

Communications Server for AIX enables an RS/6000 computer to participate in an SNA network that includes mainframes, PCs and other workstations. You can configure CS/AIX data link profiles as HACMP communications links resources. HACMP handles the CS/AIX protocol stopping and starting in the event of an adapter or node failure. The LU6.2 or LU2 connections, link stations, and CS/AIX servers are automatically stopped and restarted during failure and recovery.

For more information on configuring these applications for HACMP, see Chapter 3, Initial Cluster Planning and the section Configuring Applications Integrated with HACMP: AIX Fast Connect, AIX Connections, and CS/AIX on page 18-11. You'll also find the appropriate planning worksheets in Appendix A, Planning Worksheets.

Robustness of Application

Of primary importance to the success of any application is the health, or robustness, of the application. If the application is unstable or crashing intermittently, you should be sure these issues are resolved prior to placing it in a high availability environment.

Beyond basic stability, an application under HACMP should meet other robustness characteristics, such as the following.

Successful Start After Hardware Failure

A good application candidate for HACMP should be able to restart successfully after a hardware failure. Run a test on an application prior to putting in under HACMP. Run the application under a heavy load and fail the node. What does it take to recover once the node is back on line? Can this recovery be completely automated? If not, the application may not be a good candidate for high availability.

Survival of Real Memory Loss

For an application to function well under HACMP it should be able to survive a loss of the contents of real memory. It should be able to survive the loss of the kernel or processor state. When a node failure occurs, these are lost. Applications should also regularly check-point the data to disk. In the event that a failure occurs, the application will be able to pick up where it last check-pointed data, rather than starting completely over.

Application Implementation Strategies

There are a number of aspects of an application that you should consider as you plan for implementing it under HACMP. You must consider characteristics such as time to start, time to restart after failure, and time to stop. Your decisions in a number of areas, including those discussed in this section—scriptwriting, file storage, **/etc/inittab** file and **cron** schedule issues—can improve the probability of successful application implementation.

Writing Effective Scripts

Writing smart application start scripts can also help reduce the likelihood of problems when bringing applications online.

A good practice for start scripts is to check prerequisite conditions before starting an application. These may include access to a filesystem, adequate paging space and free filesystem space. The start script should exit and run a command to notify system administrators if the requirements are not met.

When starting a database it is important to consider whether there are multiple instances within the same cluster. If this is the case, you must be careful to start only the instances applicable for each node. Certain database startup commands read a configuration file and start all known databases at the same time. This may not be a desired configuration for all environments.

Considering File Storage Locations

You should also give thought to where the configuration files reside. They could either be on shared disk, and thus potentially accessed by whichever node has the volume group varied on, or on each node's internal disks. This holds true for all aspects of an application. Certain files must be on shared drives. These include data, logs, and anything that could be updated by the execution of the application. Files such as configuration files or application binaries could reside in either location.

There are pros and cons to storing optional files in either location. Having files stored on each node's internal disks implies that you have multiple copies of, and potentially multiple licenses for, the application. This could require additional cost as well as maintenance in keeping these files synchronized. However, in the event that an application needs to be upgraded, the entire cluster need not be taken out of production. One node could be upgraded while the other remains in production. The "best" solution is the one that works best for a particular environment.

Considering /etc/inittab and cron Table Issues

You must also give thought to applications, or resources needed by an application, that either start out of the **/etc/inittab** file or out of the **cron** table. The **inittab** starts applications upon boot up of the system. If cluster resources are needed for an application to function, they will not become available until after HACMP is started. It is better to use the HACMP application server feature which allows the application to be a resource that is started only after all dependent resources are online.

In the **cron** table, jobs are started according to a schedule set in the table and the date setting on a node. This information is maintained on internal disks and thus cannot be shared by a standby node. You must synchronize these **cron** tables so that a standby node can perform the necessary action at the appropriate time. You must also ensure the date is set the same on the primary node and any of its standby nodes.

Examples: Oracle Database™ and SAP R/3™

Here are two examples illustrating issues to consider in order to make the applications Oracle Database and SAP R/3 function well under HACMP.

Example 1: Oracle Database

The Oracle Database, like many databases, functions very well under HACMP. It is a robust application that handles failures well. It can roll back uncommitted transactions after a fallover and return to service in a timely manner. There are, however, a few things to keep in mind when using Oracle Database under HACMP.

Starting Oracle

Oracle must be started by the Oracle user ID. Thus, the start script should contain an **su - oracleuser**. The dash (-) is important since the **su** needs to take on all characteristics of the Oracle user and reside in the Oracle user's home directory. The command would look something like this:

```
su - oracleuser -c "/apps/oracle/startup/dbstart"
```

Commands like **dbstart** and **dbshut** read the **/etc/oratabs** file for instructions on which database instances are known and should be started. In certain cases it is inappropriate to start all of the instances, because they might be owned by another node. This would be the case in the mutual takeover of two Oracle instances. The **oratabs** file typically resides on the internal disk and thus cannot be shared. If appropriate, consider other ways of starting different Oracle instances.

Stopping Oracle

The stopping of Oracle is a process of special interest. There are several different ways to ensure Oracle has completely stopped. The recommended sequence is this: first, implement a graceful shutdown; second, call a shutdown immediate, which is a bit more forceful method; finally, create a loop to check the process table to ensure all Oracle processes have exited.

Oracle File Storage

The Oracle product database contains several files as well as data. It is necessary that the data and redo logs be stored on shared disk so that both nodes may have access to the information. However, the Oracle binaries and configuration files could reside on either internal or shared disks. Consider what solution is best for your environment.

For more information about keeping your Oracle applications highly available, see the IBM Redbook #SG24-4788, *Bullet-Proofing Your Oracle Database with HACMP: A Guide to Implementing AIX Databases with HACMP*.

Example 2: SAP R/3, a Multi-Tiered Application

SAP R/3 is an example of a three-tiered application. It has a database tier, an application tier, and a client tier. Most frequently, it is the database tier that is made highly available. In such a case, when a failover occurs and the database is restarted, it is necessary to stop and restart the SAP application tier. You can do this in one of two ways: by using the **rsh** command, or by making the application tier nodes "cluster aware."

Using the rsh Command

The first way to stop and start the SAP application tier is to create a script that performs an **rsh** to the application nodes. The application tier of SAP is stopped and then restarted. This is done for every node in the application tier. Use of the **rsh** command requires a method of allowing the database node access to the application node. Certain methods, such as the use of **.rhosts** files, pose a security risk and may not be desirable.

As mentioned earlier, another way to address the **rsh** security issue is by using Kerberos as an authentication method. If you choose this method, Kerberos must be in place prior to the HACMP installation and configuration. IBM supports Kerberos on IBM Scalable POWERParallel (RS/6000 SP) systems. For more information about configuring Kerberos, see Appendix F of the *HACMP for AIX Installation Guide*.

Making Application Tier Nodes “Cluster Aware”

A second method for stopping and starting the application tier is to make the application tier nodes “cluster aware.” This means that the application tier nodes are aware of the clustered database and know when a fallover occurs. You can implement this by making the application tier nodes either HACMP servers or clients. If the application node is a server, it runs the same cluster events as the database nodes to indicate a failure; pre- and post-event scripts could then be written to stop and restart the SAP application tier. If the application node is an HACMP client, it is notified of the database fallover via SNMP through the cluster information daemon (**clinfo**). A program could be written using the Clinfo API to stop and restart the SAP application tier.

Consult the manual *HACMP for AIX: Programming Client Applications* for more detail on the Clinfo API.

Applications and HACMP

Examples: Oracle Database™ and SAP R/3™

Appendix D Installing and Configuring Cluster Monitoring with Tivoli

This appendix contains instructions for making an HACMP/ES cluster known to Tivoli in order to monitor the cluster through the Tivoli management console.

Overview

You can monitor the state of an HACMP/ES cluster and its components through your Tivoli Framework enterprise management system. Using various windows of the Tivoli interface, you can monitor the following aspects of your cluster:

- Cluster state and substate
- Configured networks and network state
- Participating nodes and node state
- Resource group location and state
- Individual resource location (not state)

In order to set up this monitoring, you must do a number of installation and configuration steps in order to make Tivoli aware of the HACMP/ES cluster and to ensure proper monitoring of IP address takeover.

For more information on using Tivoli to monitor your cluster once installation is complete, see Chapter 21, *Monitoring an HACMP/ES Cluster*.

Installing and Configuring Cluster Monitoring with Tivoli

The rest of this appendix covers prerequisites and procedures for setting up your cluster to be monitored by Tivoli Framework.

Prerequisites and Considerations

When planning and configuring Cluster Monitoring with Tivoli, keep the following points in mind:

- The Tivoli Management Region (TMR) should be located on an AIX node outside the cluster.
- The HACMP/ES cluster nodes must be configured as managed nodes in Tivoli.
- The Tivoli Framework, Distributed Monitoring, and AEF components must be installed on the Tivoli Management Region node and on each cluster node. (See *Memory and Disk Requirements for Cluster Monitoring with Tivoli* on page D-2).
- To ensure proper monitoring of IP address takeover activity, the ideal configuration is to have a separate network dedicated to communication between the TMR and the cluster nodes. If you do not have a separate network dedicated to Tivoli, you must take additional

steps to ensure that IPAT is monitored properly. These steps include defining an extra subnet and an alias IP address. This additional subnet is used for the TMR node IP address and for the cluster node's alias IP address. (See note below on subnet considerations.)

Memory and Disk Requirements for Cluster Monitoring with Tivoli

The memory required for individual Distributed Monitors for cluster components varies depending on the size of the cluster and the number of components being monitored. Consult your Tivoli documentation for more information.

Installation of the **hativoli** filesets requires 400 KB of disk space. Check your Tivoli documentation for additional disk space requirements.

Subnet Considerations for Cluster Monitoring with Tivoli

In order to ensure the proper monitoring of IP address takeover in a Tivoli-monitored cluster, you must create an alias to the standby adapter of each cluster node. You must include this alias in the `/etc/hosts` file, the `ipaliases.conf` file, and also in the Tivoli `/etc/wlocalhost` file.

The subnet of this alias must be *different* than the cluster node's service and standby adapters, and the *same* as the subnet of the Tivoli Management Region (server) node.

Here is an example of what you might insert into the `/etc/hosts` file for a Tivoli-monitored cluster node named HACMPnode and a Tivoli server node named TMRnode. HACMPnode has service, standby, and alias IP addresses; TMRnode has a service IP address.

The netmask for this example network is 255.255.255.0

Adapter Label	Address
HACMPnode_svc	10.50.20.88
HACMPnode_stby	10.50.25.88
HACMPnode_alias	10.50.21.89
TMRnode	10.50.21.10

You can see in this example that the alias address and the TMR address are on the same subnet, and this subnet is *in addition to* the two already used for the cluster node's service and standby adapters.

Steps for Installing and Configuring Cluster Monitoring with Tivoli

Preparing to monitor a cluster with Tivoli involves several stages and prerequisite tasks.

The table below provides an overview of all of the steps you will take. Use this table to familiarize yourself with the "big picture" of the installation and configuration steps. Then refer to the sections that follow for details on each step.

This sequence of steps assumes an environment in which:

- Tivoli has already been installed and set up.
- The Tivoli configuration is being modified to monitor an HACMP/ES cluster for the first time.
- You do not have a separate network for monitoring the HACMP/ES cluster.

Step	Details on page...
<p>1 Ensure that Tivoli software is installed and running on the TMR node and on cluster nodes.</p> <p>Note: If you are doing a fresh installation of Tivoli, see the steps below related to creating an alias to each node's standby adapter. You may want to perform these steps as you install Tivoli to avoid unnecessary work later.</p>	D-4
<p>2 On the TMR, create a Policy Region and Profile Manager for HACMP/ES monitoring.</p>	D-4
<p>3 Define the cluster nodes as Tivoli clients (managed nodes).</p>	D-4
<p>4 Define other necessary managed resources.</p>	D-5
<p>5 Subscribe the cluster nodes to the Tivoli Profile Manager.</p>	D-5
<p>6 If you haven't done so already, install the HACMP/ES software, and install the three cluster.hativoli filesets on the TMR and the cluster nodes.</p>	D-5
<p>7 If you haven't done so already, configure the HACMP/ES cluster and synchronize.</p>	D-5
<p>8 Define the IP address alias and enter it in the /etc/hosts file on each node.</p> <p>Note: At this point, you may need to change the IP address of the TMR so that the TMR can communicate with the alias IP address on the cluster nodes. Refer to your Tivoli documentation or customer support for additional help.</p>	D-6
<p>9 Verify that the Tivoli /etc/wlocalhost file exists and add the IP address alias to it.</p>	D-6
<p>10 Create the ipaliases.conf file and include the network name connecting Tivoli with the cluster, and the name of each cluster node with its alias label.</p>	D-6
<p>11 Run the ifconfig command to define the alias to network interface.</p>	D-6
<p>12 Start the Tivoli oserv process on each node.</p>	D-6
<p>13 If you have prior customizations of node properties in Tivoli, make sure they are saved.</p>	D-6
<p>14 Run the /usr/sbin/hativoli/bin/install script.</p>	D-7
<p>15 Re-synchronize cluster resources from the same node you used in the previous step.</p>	D-8

	Step	Details on page...
16	Run the <code>/usr/sbin/hativoli/AEF/install</code> script on the TMR.	D-8
17	Run the <code>/usr/sbin/hativoli/bin/install_aef_client</code> script on all cluster nodes.	
18	Start cluster services on the cluster nodes.	
19	Start Tivoli on the TMR node (if not already running).	

The following sections provide further details about each of the installation steps.

Installing the Required Tivoli Software

The following Tivoli software must be installed before installing the Tivoli-related HACMP/ES filesets:

- Tivoli Framework 3.6 (on TMR and cluster nodes)
- Tivoli Application Extension Facility (AEF) 3.6 (on TMR only)
- Tivoli TME 10 Distributed Monitoring 3.5 (on TMR and cluster nodes)
- Tivoli TME 10 Distributed Monitoring 3.5.1 (on TMR and cluster nodes)

Creating a Cluster Policy Region and Profile Manager

The first step is to create a Policy Region and Profile Manager to handle the HACMP/ES cluster information.

Consult your Tivoli documentation or online help if you need instructions for performing these Tivoli tasks.

Defining HACMP/ES Cluster Nodes as Tivoli Managed Nodes

You also must configure each HACMP/ES cluster node as a subscriber (client) node to an HACMP/ES Profile on the Tivoli Management Region (TMR). Each configured node is then considered a “managed node” that appears in the Tivoli Policy Region window. Each managed node maintains detailed node information in its local Tivoli database, which the TMR accesses for updated node information.

Note that since the TMR does not recognize HACMP/ES automatically, you must enter the name of an adapter known to the cluster node you are defining as a client. Do this in the Add Clients window.

Note: If you already have Tivoli configured, and want to configure IP address takeover, you must configure an alias to the standby adapter of each cluster node. In addition, you must change the IP address of the TMR node to match the subnet of the IP address aliases you assigned for the standby adapters of the cluster nodes. See page D-6 for details.

Follow the procedure you would follow to install any nodes for Tivoli to manage. Refer to Tivoli documentation and online help for instructions.

Defining Administrators

Define the cluster nodes as Login Names in the Administrators screen. Consult your Tivoli documentation or online help if you need instructions for performing Tivoli tasks.

Defining Other Managed Resources

At this stage, you define some other resources to be managed in addition to the cluster nodes, such as the profile manager and indicator collection, as follows:

1. In the TME Desktop initial window, click on the newly-created policy region.
The Policy Region window appears.
2. From the Policy Region window, select **Properties > Managed Resources**.
The Set Managed Resources window appears.
3. From the Available Resources list, double-click on the following items to move them to the Current Resources list:
 - **ManagedNode**
 - **IndicatorCollection**
 - **ProfileManager**
 - **SentryProfile**
 - **TaskLibrary**
4. Click **Set & Close** to continue.

Adding Nodes as Subscribers to the Profile Manager

1. Double-click on the new Profile Manager icon.
The Profile Manager window appears.
2. Select **ProfileManager > Subscribers...**
3. In the Subscribers window, move your cluster node names from the Available to become Subscribers list to the Current Subscribers list.
4. Click **Set Subscriptions & Close**.
5. Return to the main TME Desktop window.

Installing the HACMP/ES Cluster Monitoring (hativoli) Filesets

Your HACMP/ES software includes three Tivoli-related optional filesets named **cluster.hativoli.client**, **cluster.hativoli.server**, and **cluster.msg.en_US.hativoli** that you can select in the SMIT Install screen. Verify that these are installed on both the Tivoli server node and the HACMP/ES cluster nodes.

Configure the HACMP/ES Cluster

You must have your HACMP/ES software installed and the cluster configured at this point.

Setting up IP Aliasing and Checking the /etc/hosts and /etc/wlocalhosts Files

You must define an alias to the IP address of each cluster node's standby adapter with a netmask that differs from the other adapter addresses and matches that of the TMR adapter. Make sure this alias is included in the `/etc/hosts` file, the `ipaliases.conf` file (see next section), and the Tivoli `/etc/wlocalhost` files. Note that if the `/etc/wlocalhost` file was not created earlier in Tivoli, you should create it now.

Refer back to the section Subnet Considerations for Cluster Monitoring with Tivoli on page D-2 for more details and an example showing what the IP addresses might look like.

To create the alias, use the `ifconfig` command as follows:

```
ifconfig <standby interface> alias <alias name> netmask <subnet mask>
```

where the *standby interface* is the interface (e.g. en2) of the adapter over which you want to communicate, and *alias* is a chosen IP address or host name (e.g. NodeA_alias).

Creating the ipaliases.conf File

If you are using IPAT without a dedicated network, you must create a file called `/usr/sbin/hativoli/ipaliases.conf` and copy it to each cluster node. This file must contain the network name you will be using for the IP aliasing, and the name of each cluster node with its alias label. For example:

```
network=token21  
node1 node1_alias  
node2 node2_alias  
node3 node3_alias
```

Starting the oserv Process

Start the Tivoli `oserv` process on all nodes. Note that the `oserv` process will not start if the alias IP address is not configured.

Note: The Tivoli `oserv` process must be running at all times in order to update the cluster information accurately. It is recommended that you set up a way to monitor the state of the `oserv` process. For information on defining an application monitor in HACMP/ES, see the section Configuring Application Monitoring on page 18-17.

To start `oserv`, run the following command on each node:

```
/etc/Tivoli/oserv.rc start
```

Saving Prior Node Properties Customizations

If you previously customized the node properties displayed in the Tivoli Cluster Managed Node window, they will be lost when the `hativoli` scripts are installed.

HACMP/ES automatically saves a copy of your parent dialog. If you need to restore earlier customizations, find the saved file in `/usr/sbin/hativoli/ParentDialog.dsl.save`.

Running the Additional hativoli Install Scripts

You now run three additional install scripts as follows. Note the node(s) on which you run each script, and note that you must synchronize cluster resources after step one.

1. Run `/usr/sbin/hativoli/bin/install` on any *ONE* cluster node

You are prompted to select the Region, the Profile Manager, and the Indicator Collection, which you set up earlier on the TMR.

There is a delay of up to 10 minutes while the system creates and distributes profiles and indicators, and adds custom post-events to your HACMP/ES configuration on this node.

2. *Important:* From the same node, re-synchronize cluster resources so that the custom post-event scripts are added to all cluster nodes.
3. Run `/usr/sbin/hativoli/AEF/install` on the *TMR* node
4. Run `/usr/sbin/hativoli/AEF/install_aef_client` on *ALL* cluster nodes

Added Post-event Scripts

Running the first **hativoli** install script and then synchronizing cluster resources automatically adds on each node a set of post-event scripts. These scripts are necessary to enable IP address takeover to work as needed. Post-events are added to the following HACMP/ES events:

- `swap_adapter_complete`
- `node_down_complete`
- `node_up_complete`
- `reconfig_resource_complete`
- `reconfig_topology_complete`
- `fail_standby`

The post-event scripts will be appended to any other post-event scripts you may have configured previously for these events.

Starting Cluster Services

Start cluster services on each cluster node.

Starting Tivoli

If Tivoli is not already running, start Tivoli by performing these steps on the TMR node

1. Make sure access control has been granted to remote nodes by running the **xhost** command with the plus sign (+) or with specified nodes. This will allow you to open a SMIT window from Tivoli.

If you want to grant access to all computers in the network, type:

```
xhost +
```

or, if you want to grant access to specific nodes only:

```
xhost <computers to be given access>
```

2. Also to ensure later viewing of SMIT windows, set `DISPLAY=<TMR node>`.

3. Run the command `./etc/Tivoli/setup_env.sh` if it was not run earlier.
4. Type **tivoli** to start the application.

The Tivoli graphical user interface appears, showing the initial TME Desktop window.

Note that there may be a delay as Tivoli adds the indicators for the cluster.

Deinstalling Cluster Monitoring with Tivoli

To discontinue cluster monitoring with Tivoli, you must perform the following steps to delete the HACMP-specific information from Tivoli.

Perform the following steps:

1. Run a deinstall through the SMIT interface, deinstalling the three **hativoli** filesets on all cluster nodes and the TMR.
2. If it is not already running, invoke Tivoli on the TMR:
 1. type `./etc/Tivoli/setup_env.sh`
 2. Type **tivoli**
3. In the Policy Region for the cluster, go to HATivoli Properties.
4. Select the Modify Properties task.

A window appears containing task icons.
5. Choose **Edit > Select All** to select all tasks, and then **Edit > Delete** to delete.

The Operations Status window at the left shows the progress of the deletions.
6. Return to the Properties window and delete the Modify Properties task icon.
7. Open the Profile Manager.
8. Choose **Edit > Profiles > Select All** to select all HACMP/ES Indicators.
9. Choose **Edit > Profiles > Delete** to delete the Indicators.
10. Unsubscribe the cluster nodes from the Profile Manager:
 1. In the Profile Manager window, choose Subscribers.
 2. Highlight each HACMP/ES node on the left, and click to move it to the right side.
 3. Click **Set & Close** to unsubscribe the nodes.

Where You Go From Here

If the installation procedure has been completed successfully, Tivoli can now begin monitoring your cluster.

See the Chapter 21, Monitoring an HACMP/ES Cluster in Volume 2 of this manual for information on how to proceed with monitoring your HACMP/ES cluster through the Tivoli management console.

HACMP/ES Master Index

This Master Index contains entries for topics covered in the two volumes of the *Enhanced Scalability Installation and Administration Guide* documentation set. Each index entry identifies the HACMP for AIX, Version 4.4 manual in which the topic is covered, using one of the following abbreviations:

For index entry page number preceded by...

Vol 1 + chapter no. 1-18, appendix letter A-D

Vol 2 + chapter no. 19-32, appendix letter E-I

Look in this HACMP for AIX, Version 4.4 manual:

Enhanced Scalability Installation and Administration Guide, Vol. 1

Enhanced Scalability Installation and Administration Guide, Vol. 2

+-* /

/.rhosts file

security setting *Vol 2* 19-4

tailoring AIX *Vol 1* 13-3

/etc/filesystems file *Vol 1* 12-5

/etc/firstboot file *Vol 1* 14-17

/etc/ha/cfg directory

EMCDB version string files *Vol 2* 31-7

run-time EMCDB files *Vol 2* 31-7

/etc/hosts file *Vol 2* H-7, H-11

and boot address *Vol 1* 13-2

editing *Vol 1* 13-2

name resolution *Vol 2* 19-4

/etc/inetd.conf file *Vol 2* 19-4

godm entry *Vol 2* 19-4

/etc/innitab file

entry for HACMP/ES startup *Vol 2* 20-5

IPAT modifications *Vol 2* 19-4

/etc/rc.net file *Vol 1* 13-4, *Vol 2* 19-5

tailoring AIX *Vol 1* 13-4

/etc/rc.net script

for cluster startup *Vol 2* 19-7

/etc/resolv.conf file *Vol 1* 4-17, *Vol 2* H-7

/etc/services file *Vol 2* 19-5

use by Event Management *Vol 2* 31-4

use by Group Services *Vol 2* 30-4

/etc/snmpd.conf file *Vol 2* 19-5

/etc/snmpd.peers file *Vol 2* 19-5

/etc/syslog.conf file *Vol 2* 19-6

/etc/tcp.clean file *Vol 2* 19-6

/etc/trcfmt file *Vol 2* 19-6

/tmp/clresmgrd.log file *Vol 2* 21-39

/tmp/clstrmgr.debug log file *Vol 2* 21-39, 29-4

/tmp/cspoc.log file *Vol 2* 20-11, 21-39

recommended use *Vol 2* 29-4

/tmp/dms_loads.out file *Vol 2* 29-4

/tmp/emuhacmp.out file *Vol 2* 21-39

/tmp/hacmp.out file *Vol 2* 21-38, 29-3

changing name or placement *Vol 2* 29-12

message formats *Vol 2* 29-8

recommended use *Vol 2* 29-3

selecting verbose script output *Vol 2* 29-12

understanding messages *Vol 2* 29-8

/usr/adm/cluster.log file *Vol 2* 21-38

/usr/es/adm/cluster.log file *Vol 2* 21-38

/usr/lib/libha_em.a *Vol 2* 31-2

/usr/lib/libha_em_r.a *Vol 2* 31-2

/usr/lib/libha_gs.a *Vol 2* 30-3

/usr/lib/libha_gs_r.a *Vol 2* 30-3

/usr/sbin/cluster/clinfo daemon *Vol 2* 20-2, 20-9, 20-15

/usr/sbin/cluster/clsmuxpd daemon *Vol 2* 20-9 starting *Vol 2* 20-9

/usr/sbin/cluster/clstat utility *Vol 2* 21-24

/usr/sbin/cluster/clstrmgr daemon *Vol 2* 20-2

/usr/sbin/cluster/diag/clconvert_snapshot utility converting cluster snapshots *Vol 1* 15-7

/usr/sbin/cluster/etc/clhosts file *Vol 1* 16-1

maintaining on clients *Vol 2* 20-14

maintaining on nodes *Vol 2* 20-15

on client *Vol 1* 17-2

/usr/sbin/cluster/etc/clinfo.rc script *Vol 1* 17-3

/usr/sbin/cluster/etc/clstop script *Vol 2* 20-5

/usr/sbin/cluster/etc/exports file *Vol 1* 6-10

/usr/sbin/cluster/etc/harc.net file

automounter daemon *Vol 1* 13-5

/usr/sbin/cluster/etc/rc.cluster script *Vol 2* 19-7

starting clients *Vol 2* 20-14

/usr/sbin/cluster/events/network scripts *Vol 2* 19-8

/usr/sbin/cluster/events/node scripts *Vol 2* 19-7

/usr/sbin/cluster/events/rules/hacmprd file *Vol 1* 8-4

/usr/sbin/cluster/events/swap_adapters script
Vol 2 19-8

/usr/sbin/cluster/events/utills directory *Vol 2 E-1*

/usr/sbin/cluster/events/utills/cl_deactivate_nfs utility
Vol 1 6-13 Vol 2 28-3, 28-4

/usr/sbin/cluster/events/utills/cl_nfskill command
Vol 1 6-13 Vol 2 28-3, 28-4

/usr/sbin/cluster/godm daemon *Vol 1 13-3*

/usr/sbin/cluster/history/cluster.mmdd file
Vol 2 21-38, 29-3

/usr/sbin/cluster/snapshots/active.x.odm file
 dynamic reconfiguration backup file *Vol 2 24-15*

/usr/sbin/cluster/utilities/cl_clstop script *Vol 2 19-7*

/usr/sbin/cluster/utilities/cl_rc.cluster script
Vol 2 19-7

/usr/sbin/cluster/utilities/clexit.rc script *Vol 2 19-7, 20-7*

/usr/sbin/cluster/utilities/clstart script *Vol 2 19-6, 20-3*

/usr/sbin/cluster/utilities/clstop script *Vol 2 19-7*

/usr/sbin/rsct/lib/libha_gs.a *Vol 2 30-3*

/usr/sbin/rsct/bin/grpsvcsctrl *Vol 2 30-4*

/usr/sbin/rsct/bin/haemd *Vol 2 31-2*

/usr/sbin/rsct/bin/hagsd *Vol 2 30-3*

/usr/sbin/rsct/bin/hatsd daemon *Vol 2 32-2*

/usr/sbin/rsct/bin/topsvcsctrl script *Vol 2 32-4*

/usr/sbin/rsct/lib *Vol 2 30-3*

/usr/sbin/rsct/lib/libha_em.a *Vol 2 31-3*

/usr/sbin/rsct/lib/libha_em_r.a *Vol 2 31-3*

/usr/sbin/rsct/lib/libha_gs_r.a *Vol 2 30-3*

/var file system
 and Event Management tracing *Vol 2 31-9*
 and Group Services tracing *Vol 2 30-6*

/var/ha/lck directory
 Group Services lock files *Vol 2 30-4*

/var/ha/lck/haem directory
 Event Management lock files *Vol 2 31-5*

/var/ha/log directory
 Event Management log files *Vol 2 31-6*
 Group Services log files *Vol 2 30-5*

/var/ha/log/grpplsm *Vol 2 21-39*

/var/ha/log/grpsvcs *Vol 2 21-39*
 recommended use *Vol 2 29-5*

/var/ha/log/topsvcs *Vol 2 21-39, 29-5*

/var/ha/run directory
 Event Manager daemon working files *Vol 2 31-6*
 Group Services daemon working files *Vol 2 30-5*
 registration cache *Vol 2 31-6*
 resource monitor working directories *Vol 2 31-6*

/var/ha/soc directory
 Event Management socket files *Vol 2 31-5*
 Group Services socket files *Vol 2 30-4*

/var/spool/cron/crontab/root file *Vol 2 19-6*

/var/spool/lpd/qdir *Vol 1 18-45*

/var/spool/qdaemon *Vol 1 18-45*

0,1,2...

2105 Versatile Storage Server (VSS) *Vol 1 5-5*

2520 messages *Vol 2 G-1*

2522 messages *Vol 2 G-28*

2523 messages *Vol 2 G-48*

2525 messages *Vol 2 G-70*

6214 SSA adapter *Vol 1 5-14*

6216 SSA adapter *Vol 1 5-14*

7 X 24 environment *Vol 2 H-1*

7013-S70 *Vol 1 4-8*

7015-S70 *Vol 1 4-8*

7017-S70 *Vol 1 4-8*

7133 SSA disk subsystem
 cluster support *Vol 1 1-4, 5-5*

7135 RAIDiant Disk Array
 cluster support *Vol 1 1-3, 5-3*
 planning considerations *Vol 1 5-3*

7137 Disk Arrays
 cluster support *Vol 1 1-4, 5-5*
 planning considerations *Vol 1 5-5*

A

activating
 volume groups
 in concurrent access mode *Vol 2 23-9*

active.n.odm file
 dynamic reconfiguration backup file *Vol 2 24-15*

adapter function
 defining adapters *Vol 1 18-4*

adapter membership
 resource monitor *Vol 2 31-4*

adapters
 assigning IP addresses *Vol 1 4-24*
 defining to cluster *Vol 1 18-2, 18-5, Vol 2 24-11*
 SSA adapter configuration *Vol 1 5-13*
 SSA disk subsystem *Vol 1 5-13*
 swapping dynamically *Vol 2 24-8*

adding
 cluster nodes *Vol 2 24-4*
 concurrent logical volume
 using C-SPOC *Vol 2 23-24*
 copy to concurrent logical volume
 C-SPOC *Vol 2 23-26*
 disk definition to cluster nodes *Vol 2 22-42*
 HACMP/ES network to global network
 Vol 2 24-11
 network adapters *Vol 2 24-6*
 physical volume to concurrent volume group
 using C-SPOC *Vol 2 23-18*
 resource groups *Vol 2 24-23*
 shared logical volume
 using C-SPOC *Vol 2 22-31*
 user accounts *Vol 2 27-2*

adding JFS
 using C-SPOC *Vol 2 22-21*

- Address Resolution Protocol *Vol 1* 16-2, 17-3
 - SP Switch *Vol 1* 4-13
 - AIX
 - error notification *Vol 1* 13-5
 - files modified by HACMP/ES *Vol 2* 19-4
 - I/O pacing *Vol 1* 13-1
 - setting I/O pacing *Vol 1* 4-18
 - setting syncd frequency *Vol 1* 4-18
 - tailoring for HACMP/ES *Vol 1* 13-1
 - user and group IDs *Vol 1* 13-2
 - AIX Connections
 - adapter failure considerations *Vol 1* 18-13
 - and adapter failure *Vol 1* 3-15
 - and AppleTalk clients *Vol 1* 3-14
 - and NetBIOS clients *Vol 1* 3-14
 - and NetWare clients *Vol 1* 3-14
 - and node failure *Vol 1* 3-15
 - configuring *Vol 1* 18-13
 - defined *Vol 1* 3-11, 3-14, 18-13
 - handling adapter failure *Vol 1* 3-15
 - handling node failure *Vol 1* 3-15
 - planning *Vol 1* 3-3, 3-14
 - realms and services available *Vol 1* 3-14
 - reconfiguring in SMIT *Vol 2* 24-28
 - verification by clverify *Vol 1* 18-14
 - AIX error log
 - Group Services entries *Vol 2* 30-8
 - use by Event Manager daemon *Vol 2* 31-6
 - AIX Error Notification *Vol 1* 13-5
 - automatic *Vol 1* 13-6
 - AIX Fast Connect
 - and adapter failure *Vol 1* 3-14
 - and node failure *Vol 1* 3-13
 - configuring *Vol 1* 18-11
 - converting from AIX Connections *Vol 1* 3-12, 18-11 *Vol 2* 24-26
 - defined *Vol 1* 3-12, *Vol 1* 18-11
 - planning *Vol 1* 3-12
 - reconfiguring in SMIT *Vol 2* 24-29
 - verification *Vol 1* 18-12
 - worksheet *Vol 1* A-31
 - AIX operating system statistics
 - obtained from SPMI *Vol 2* 31-7
 - Application Monitor Worksheets
 - process monitor *Vol 1* A-39
 - user-defined monitor *Vol 1* A-41
 - application monitoring *Vol 2* 21-29
 - changing *Vol 2* 24-22
 - configuring *Vol 1* 18-19
 - overview *Vol 1* 18-17
 - prerequisites *Vol 1* 18-18
 - process monitoring *Vol 1* 18-19
 - removing a monitor *Vol 2* 24-22
 - suspending and resuming *Vol 2* 24-21
 - troubleshooting *Vol 2* 29-24
 - user defined monitoring *Vol 1* 18-21
 - Application Server Worksheet *Vol 1* A-37
 - application servers
 - changing *Vol 2* 24-20
 - changing scripts *Vol 2* 24-20
 - configuring *Vol 1* 18-10
 - defining *Vol 1* 18-10, *Vol 2* 24-19
 - removing *Vol 2* 24-17, 24-19
 - start script *Vol 1* 3-8
 - stop script *Vol 1* 3-8
 - Application Worksheet *Vol 1* A-27
 - applications
 - automating with start/stop scripts *Vol 1* C-2
 - customizing scripts *Vol 2* H-5
 - implementation strategies *Vol 1* C-6
 - integrated with HACMP/ES *Vol 1* 3-3
 - planning *Vol 1* 3-11
 - licensing *Vol 1* 5-7
 - multi-tiered application issues *Vol 1* C-3
 - planning *Vol 1* 3-2, 3-4
 - planning for high availability *Vol 1* 3-1
 - reducing protocol conflict *Vol 1* C-5
 - applying
 - saved cluster configurations dynamically *Vol 2* 26-1
 - ARP
 - enabling on SP switch *Vol 1* 13-4
 - ARP cache
 - clinfo.rc script *Vol 1* 16-2, 17-3
 - updating for clients *Vol 1* 9-3
 - assigning
 - cluster name and ID *Vol 1* 3-5
 - node names *Vol 1* 3-7
 - Asynchronous Transfer Mode *Vol 1* 11-3
 - ATM
 - configuring connections *Vol 1* 11-3
 - hardware address swapping
 - configuration requirements *Vol 1* 4-28
 - specifying an alternate address *Vol 1* 4-28
 - LAN emulation *Vol 1* 11-8
 - ATM adapters
 - specifying alternate HW address *Vol 1* 4-28
 - automatic error notification *Vol 1* 13-6
 - deleting methods assigned *Vol 1* 13-8
 - automounter daemon
 - enabling *Vol 1* 13-5
 - availability
 - definition *Vol 1* 1-1
- ## B
- backing up
 - HACMP/ES system *Vol 2* 19-3
 - backout procedure
 - reversing migration from HACMP to HACMP/ES *Vol 1* 14-18
 - backups *Vol 2* H-19
 - barrier commands
 - user-defined events *Vol 1* 8-4

- boot adapter
 - definition *Vol 1* 4-9
- boot addresses
 - configuring adapter *Vol 1* 11-1
 - in /etc/hosts file *Vol 1* 13-2
 - in clhosts file *Vol 1* 16-1
 - in nameserver configuration *Vol 1* 13-2
- building
 - HACMP/ES clusters *Vol 1* 1-1
- bypass cards
 - SSA disk subsystems *Vol 1* 5-14
- bypass mode
 - SSA disk subsystem *Vol 1* 5-14

C

- cascading
 - definition *Vol 1* 3-6
- cascading resource groups
 - and cldare stop command *Vol 2* 24-35
 - cascading without fallback *Vol 1* 7-4
 - behavior with inactive takeover *Vol 1* 7-4
 - DARE migration issues *Vol 1* 7-2
 - changing priority of nodes dynamically
 - Vol 2* 24-5
 - NFS cross mounting issues *Vol 1* 6-11
 - Vol 2* 28-2
- changing
 - application servers *Vol 2* 24-20
 - characteristics of shared logical volume
 - Vol 2* 22-32
 - cluster environment properly *Vol 2* H-12
 - cluster name or ID *Vol 2* 24-3
 - cluster nodes configuration *Vol 2* 24-4
 - global networks *Vol 2* 24-11
 - IP address properly *Vol 2* H-11
 - name of cluster node *Vol 2* 24-6
 - network adapter attributes *Vol 2* 24-10
 - network adapters configuration *Vol 2* 24-6
 - network modules *Vol 1* 18-7 *Vol 2* 24-12
 - resource group definition *Vol 2* 24-24
 - shared file system *Vol 2* 22-23
 - user accounts *Vol 2* 27-4
- changing cluster configuration
 - effects on components *Vol 2* H-10
- checking
 - installed hardware *Vol 1* 11-1
 - SCSI devices installation *Vol 1* 11-9
- cl_activate_fs utility *Vol 2* E-9
- cl_activate_nfs utility *Vol 2* E-9
- cl_activate_vgs utility *Vol 2* E-10
- cl_convert utility *Vol 1* 15-5
- cl_deactivate_fs utility *Vol 2* E-10
- cl_deactivate_nfs utility *Vol 2* E-10
- cl_deactivate_vgs utility *Vol 2* E-11
- cl_disk_available utility *Vol 2* E-1
- cl_echo utility *Vol 2* E-13
- cl_export_fs utility *Vol 2* E-11
- cl_fs2disk utility *Vol 2* E-2
- cl_get_disk_vg_fs_pvids utility *Vol 2* E-2
- cl_is_array utility *Vol 2* E-3
- cl_is_scsidisk utility *Vol 2* E-3
- cl_log utility *Vol 2* E-13
- cl_lsuser command
 - using *Vol 2* 27-1
- cl_mkuser command
 - using *Vol 2* 27-2
- cl_mkvg command
 - creating concurrent volume group *Vol 2* 23-17
- cl_nfskill command *Vol 2* E-12
- cl_nm_nis_off utility *Vol 2* E-14
- cl_nm_nis_on utility *Vol 2* E-14
- cl_opsconfig *Vol 1* B-10
- cl_raid_vg utility *Vol 2* E-3
- cl_scsidiskreset utility *Vol 2* E-4
- cl_scsidiskrsrv utility *Vol 2* E-4
- cl_swap_HPS_IP_address utility *Vol 2* E-8
- cl_swap_HW_address utility *Vol 2* E-14
- cl_swap_IP_address utility *Vol 2* E-15
- cl_sync_vgs utility *Vol 2* E-5
- cl_unswap_HW_address *Vol 2* E-16
- cl_updatevg command
 - updating ODM data on remote nodes *Vol 2* 22-4
- clconvert_snapshot utility *Vol 1* 14-13, 15-5, 15-7
 - Vol 2* 26-4
- cldiag utility
 - customizing /tmp/hacmp.out file output
 - Vol 2* 29-11
 - customizing output *Vol 2* 29-8
 - initiating a trace session *Vol 2* 29-18
 - obtaining trace information *Vol 2* 29-22
 - options and flags *Vol 2* 29-7
 - viewing the /tmp/hacmp.out file *Vol 2* 29-10
 - viewing the cluster.log file *Vol 2* 29-6
 - viewing the system error log *Vol 2* 29-14
- cleaning subsystems
 - Event Management (emsvcsctrl) *Vol 2* 31-9
- clfindres command *Vol 2* 21-31
- clhandle
 - HACMP utility for Topology Services *Vol 2* 32-7
- clhosts file
 - editing on cluster nodes *Vol 1* 16-1
- client
 - type of resource monitor *Vol 2* 31-3
- client communication
 - with Event Management subsystem *Vol 2* 31-2
 - with Group Services subsystem *Vol 2* 30-3
- client traffic
 - planning considerations *Vol 1* 4-5
- client, Group Services
 - definition *Vol 2* 30-1

- clients
 - not running Clinfo *Vol 1* 9-3
 - planning *Vol 1* 9-1
 - running Clinfo *Vol 1* 9-2
 - starting and stopping cluster services *Vol 2* 20-14
- Clinfo
 - enabling traps *Vol 2* 20-15
 - running on clients *Vol 1* 9-2
 - setting up files and scripts *Vol 1* 16-1
 - starting on clients *Vol 2* 20-14
 - stopping on clients *Vol 2* 20-14
 - trace ID *Vol 2* 29-20
- clinfo daemon *Vol 1* 16-1 *Vol 2* 20-2
 - starting *Vol 2* 20-9
- clinfo.rc script *Vol 2* 19-8
 - editing *Vol 1* 16-2
 - planning *Vol 1* 9-2
 - tailoring *Vol 1* 17-3
- clockd daemon *Vol 2* 20-2
- clRGinfo command *Vol 2* 21-30
 - reference page *Vol 2* E-20
- clRMupdate command *Vol 2* 21-31
 - reference page *Vol 2* E-19
- clsmuxpd daemon *Vol 2* 20-2
- clstat utility
 - monitoring cluster *Vol 2* 21-24
 - multi-cluster mode *Vol 2* 21-26
 - single-cluster mode *Vol 2* 21-25
 - X Window display *Vol 2* 21-27
- clstrmgr daemon *Vol 2* 20-2
 - starting *Vol 2* 20-8, 20-10
- cluster
 - configuring resources *Vol 1* 18-9
 - initial diagram *Vol 1* 3-17
 - initial planning steps *Vol 1* 3-1
 - managing resources *Vol 2* 24-23
 - monitoring
 - overview *Vol 2* 21-1
 - monitoring tools *Vol 2* 21-2
 - naming *Vol 1* 3-5
 - planning
 - list of steps *Vol 1* 2-3
 - planning clients *Vol 1* 9-1
 - planning cluster events *Vol 1* 8-1
 - planning disks *Vol 1* 5-1
 - planning for performance *Vol 1* 4-17
 - planning number of nodes *Vol 1* 3-2
 - planning resource groups *Vol 1* 7-1
 - planning resources *Vol 1* 3-4
 - planning shared LVM components *Vol 1* 6-1
 - topology information *Vol 2* 24-3
 - tuning I/O pacing *Vol 1* 13-1
 - tuning performance parameters *Vol 1* 18-6
 - upgrading *Vol 1* 15-1
 - verifying configuration *Vol 2* 25-1
- cluster configuration
 - adding custom-defined verification methods
 - Vol 2* 25-7
 - changing/showing custom-defined verification methods *Vol 2* 25-8
 - saving *Vol 2* 26-1
- cluster environment
 - synchronizing on all nodes *Vol 1* 18-8
 - verifying *Vol 1* 18-37
- Cluster Event Worksheet *Vol 1* A-45
- cluster events
 - configuring *Vol 1* 18-42
 - customizing *Vol 1* 18-41
 - emulating *Vol 2* 21-32
 - event customization facility *Vol 1* 8-1
 - notification *Vol 1* 8-2
 - post-processing *Vol 1* 8-2
 - pre-processing *Vol 1* 8-2
 - recovery *Vol 1* 8-3
 - retry *Vol 1* 8-3
- cluster history log file
 - message format and content *Vol 2* 29-15
- cluster ID
 - assigning *Vol 1* 3-5
- cluster log files
 - redirecting *Vol 1* 18-40
- Cluster Manager
 - trace ID *Vol 2* 29-20
- cluster monitoring
 - overview of monitoring methods *Vol 2* 21-1
 - with HAView *Vol 2* 21-2
 - with Tivoli
 - defining managed nodes *Vol 1* D-4
 - deinstalling *Vol 1* D-8 *Vol 2* 21-23
 - installation steps and prerequisites *Vol 1* D-1
 - IPAT considerations *Vol 1* D-6
 - polling intervals *Vol 2* 21-23
 - prerequisites *Vol 2* 21-15
 - required Tivoli software *Vol 1* 14-2, D-4
 - subnet requirements *Vol 1* D-2
 - using *Vol 2* 21-14
- cluster name
 - changing *Vol 2* 24-3
- cluster resources
 - reconfiguring *Vol 2* 24-23
 - synchronizing
 - skipping cluster verification during *Vol 2* 24-40
- cluster security
 - DCE authentication *Vol 1* 18-37
 - Kerberos *Vol 1* 18-31
 - PSSP enhanced security *Vol 1* 18-36
- Cluster Security Mode
 - setting in SMIT screen *Vol 1* 18-32

- cluster services
 - daemons *Vol 2* 20-2
 - starting
 - on clients *Vol 2* 20-14
 - starting on a single node *Vol 2* 20-8
 - stopping
 - on clients *Vol 2* 20-14
 - stopping on nodes *Vol 2* 20-11
- Cluster SMUX Peer
 - trace ID *Vol 2* 29-20
- cluster snapshot
 - applying *Vol 2* 26-6
 - backup files *Vol 2* 26-7
 - changing *Vol 2* 26-8
 - changing or removing custom method *Vol 2* 26-5
 - contents *Vol 2* 26-1
 - converting to current version *Vol 1* 15-7
 - converting to Version 4.3 *Vol 1* 15-7
 - creating *Vol 2* 26-5
 - defining a custom method *Vol 2* 26-4
 - files *Vol 2* 26-3
 - naming *Vol 2* 26-6
 - removing *Vol 2* 26-8
 - reverting to previous configuration *Vol 2* 26-7
 - saving and restoring cluster configurations
 - Vol 2* 26-1
 - using *Vol 2* 26-1
- cluster snapshots
 - cron jobs *Vol 2* H-19
 - saved during migration *Vol 1* 14-17
- cluster topology
 - synchronizing *Vol 2* 24-13
 - skipping clverify during *Vol 2* 24-14
 - verifying *Vol 1* 18-39, *Vol 2* 25-4
 - viewing *Vol 2* 24-3
- cluster verification
 - ignoring during synchronization *Vol 2* 24-41
 - skipping during synchronization *Vol 2* 24-14
 - ways to run clverify *Vol 2* 25-1
- cluster.log file *Vol 2* 21-38
 - customizing output *Vol 2* 29-8
 - message format *Vol 2* 29-5
 - recommended use *Vol 2* 29-3
 - viewing its contents *Vol 2* 29-6
- cluster.mmdd file *Vol 2* 21-38
 - cluster history log *Vol 2* 29-15
 - recommended use *Vol 2* 29-4
- clverify
 - run as cron job *Vol 2* H-21
- clverify utility
 - checking cluster topology *Vol 1* 18-39
 - command line mode *Vol 2* 25-5
 - flags *Vol 2* 25-3
 - help option *Vol 2* 25-3
 - interactive mode *Vol 2* 25-3
 - quitting *Vol 2* 25-3
 - running with SMIT *Vol 2* 25-6
 - ways to run *Vol 2* 25-1
- clvmd daemon *Vol 2* 23-16
- commands
 - Cluster Resource Manager *Vol 2* E-19
 - emsvcsctrl script *Vol 2* F-1
 - grpsvcsctrl script *Vol 2* F-6
 - haemd_HACMP *Vol 2* F-11
 - haemqvar *Vol 2* F-12
 - haemtrcoff *Vol 2* F-16
 - haemtrcon *Vol 2* F-18
 - haemunlkrm *Vol 2* F-20
 - topsvcs *Vol 2* F-23
 - topsvcsctrl script *Vol 2* F-24
- communication system *Vol 2* H-10
- concurrent access mode
 - disk fencing *Vol 1* 5-17
 - HACMP/ES scripts *Vol 2* 23-1
 - maintaining RAID devices *Vol 2* 23-1
 - maintaining shared LVM components *Vol 2* 23-1
 - shared LVM components *Vol 1* 12-7
 - varyonvg command *Vol 2* 23-2
- concurrent logical volumes
 - maintaining with C-SPOC *Vol 2* 23-24
- Concurrent Resource Manager
 - installing *Vol 1* 14-10, 15-5
- concurrent volume groups
 - creating *Vol 2* 23-3
 - creating with C-SPOC utility *Vol 2* 23-17
 - maintaining with C-SPOC *Vol 2* 23-18
 - SSA *Vol 1* 5-17, 6-17
- config_too_long message
 - handling during migration *Vol 1* 14-16

- configuring
 - application servers *Vol 1* 18-10
 - ATM networks *Vol 1* 11-3
 - automatic error notification *Vol 1* 13-7
 - cluster environment
 - overview *Vol 1* 18-1
 - cluster events *Vol 1* 18-42
 - cluster resources *Vol 1* 18-9
 - HACMP/ES clients *Vol 1* 17-1
 - IP address takeover
 - example *Vol 1* 4-9
 - multiple networks *Vol 1* 4-3
 - network adapters in AIX *Vol 1* 11-1
 - network modules *Vol 1* 18-6, *Vol 2* 24-11
 - resources for a resource group *Vol 2* 24-26
 - run-time parameters *Vol 1* 18-29, *Vol 2* 28-5, 29-12
 - SSA for optimal performance *Vol 1* 5-16
 - target mode SCSI *Vol 1* 11-15
 - target mode SSA serial network *Vol 1* 11-18
 - TMSSA devices *Vol 1* 11-18
 - undoing a dynamic reconfiguration *Vol 2* 24-15
- configuring clusters
 - using xhacmpm application *Vol 2* 1-1
- connecting
 - SCSI bus configuration *Vol 1* 5-9
- control script
 - component of Event Management *Vol 2* 31-5
 - component of Group Services *Vol 2* 30-4
- conversion
 - from HACMP for AIX *Vol 1* 14-11
- core file
 - Event Manager daemon *Vol 2* 31-6
 - Group Services daemon *Vol 2* 30-5
- creating
 - concurrent capable volume groups *Vol 2* 23-3
 - concurrent volume group *Vol 2* 23-3
 - concurrent volume group using C-SPOC *Vol 2* 23-17
 - resource groups *Vol 1* 18-9
 - shared filesystems *Vol 1* 12-4 *Vol 2* 22-18
 - shared volume groups
 - concurrent access *Vol 1* 12-7
 - NFS issues *Vol 1* 6-10, *Vol 2* 28-1
 - using the TaskGuide *Vol 1* 12-1 *Vol 2* 23-2
 - with AIX commands *Vol 2* 22-5
- CRM
 - installation images *Vol 1* 14-6
 - installing *Vol 1* 14-10, 15-5
- cron
 - taking cluster snapshots *Vol 2* H-19
- cron and NIS *Vol 1* 13-2
- cron jobs
 - making highly available *Vol 1* 18-44
- cron utility *Vol 2* H-20
- cross mounting
 - NFS filesystems *Vol 1* 6-11 *Vol 2* 28-2
- CS/AIX
 - configuring links as a resource *Vol 1* 18-14
 - handling by HACMP/ES *Vol 1* 18-15
 - planning *Vol 1* 3-16
 - reconfiguring links *Vol 2* 24-16
- CS/AIX Communications Links Worksheet *Vol 1* A-35
- CS/AIX Data Link Control profiles
 - as HACMP/ES resources *Vol 1* 3-16
- C-SPOC
 - adding disk definition to the cluster *Vol 2* 22-42
 - commands *Vol 2* 22-3
 - create shared filesystem *Vol 2* 22-21
 - creating concurrent volume group *Vol 2* 23-17
 - maintaining concurrent logical volumes *Vol 2* 23-24
 - maintaining concurrent LVM components *Vol 2* 23-16
 - maintaining shared LVM components *Vol 2* 22-2
 - maintaining shared volume groups *Vol 2* 22-12
 - managing user/group accounts *Vol 2* 27-1
 - operations on shared logical volumes *Vol 2* 22-39
 - removing disk definition from cluster *Vol 2* 22-44
 - starting cluster services *Vol 2* 20-4, 20-10
 - stopping cluster services *Vol 2* 20-6
- cspoc.log file
 - message format *Vol 2* 29-16
 - viewing its contents *Vol 2* 29-17
- cssMembership group *Vol 2* 31-4
 - use by Event Management *Vol 2* 31-8
- custom scripts
 - print queues *Vol 1* 18-45
 - samples *Vol 1* 18-44
- custom verification methods
 - adding *Vol 2* 25-7
 - removing *Vol 2* 25-8
- customizing
 - 7 X 24 maintenance *Vol 2* H-2
 - cluster log files *Vol 1* 18-40
 - events *Vol 1* 18-41 *Vol 2* 24-42
- cycles to fail
 - definition *Vol 1* 4-19

D

- daemons
 - abnormal termination *Vol 2* 20-7
 - checking status *Vol 2* 20-10
 - clinfo *Vol 1* 16-1
 - cluster messages *Vol 2* 29-2
 - godm *Vol 1* 13-3
 - grpqlsmd *Vol 2* 20-3
 - grpsvcsd *Vol 2* 20-3
 - HACMP/ES cluster services *Vol 2* 20-2
 - haemd *Vol 2* F-10
 - hagsd *Vol 2* F-21
 - hagsqlsmd *Vol 2* F-22
 - handling properly under HACMP *Vol 2* H-10
 - monitoring on clients *Vol 2* 21-38
 - monitoring on cluster nodes *Vol 2* 21-37
 - topsvcsd *Vol 2* 20-2
 - trace IDs *Vol 2* 29-20
 - type of resource monitor *Vol 2* 31-3
- DARE
 - Resource Migration utility
 - and cascading without fallback *Vol 1* 7-2
- DARE Resource Migration
 - example *Vol 2* 24-35
 - overview *Vol 2* 24-31
 - using SMIT *Vol 2* 24-36
- DARE utility
 - use for hardware maintenance *Vol 2* H-14
- DCD
 - restoring from ACD *Vol 2* 24-15
- deadman switch *Vol 1* 4-17
 - and /tmp/dms_loads.out *Vol 2* 29-4
 - cluster performance tuning *Vol 1* 18-6
 - formula *Vol 1* 4-19
 - timeouts per network *Vol 1* 4-19
 - turning DMS off *Vol 2* 20-5
- debug levels
 - setting *Vol 1* 18-30
 - setting on a node *Vol 2* 28-5
- default location keyword *Vol 2* 24-33
- defining
 - adapters to cluster *Vol 1* 18-2
 - cluster ID *Vol 1* 18-1
 - cluster name *Vol 1* 18-1
 - cluster nodes *Vol 1* 18-2
 - cluster topology *Vol 1* 18-1
 - custom snapshot method *Vol 2* 26-4
 - global networks *Vol 1* 18-5
 - hardware addresses *Vol 1* 4-26
 - resource groups *Vol 1* 3-6
 - shared LVM components *Vol 1* 12-1
 - concurrent access *Vol 1* 12-7
 - tty device *Vol 1* 11-16
- deinstalling
 - HACMP/ES after migration *Vol 1* 14-18
- deleting
 - application server *Vol 2* 24-17
 - automatic error methods *Vol 1* 13-8
 - cluster nodes *Vol 2* 24-5
 - resource groups *Vol 2* 24-24
 - shared volume group *Vol 2* 22-11
- deleting subsystems
 - Event Management (emsvcsctrl) *Vol 2* 31-9
- destination nodes
 - importing volume group *Vol 1* 12-6
- DGSP message
 - avoiding *Vol 2* H-6
- DHCP
 - allocating IP addresses *Vol 2* H-7
- diagram
 - initial cluster drawing *Vol 1* 3-17
- disk adapters
 - checking SCSI installation *Vol 1* 11-9
 - installing SCSI *Vol 1* 11-9
- disk drive
 - replacing *Vol 2* H-16
- disk failures
 - handling *Vol 2* H-16
- disk fencing
 - and dynamic reconfiguration *Vol 1* 5-18
 - benefits *Vol 1* 5-19
 - enabling *Vol 1* 5-18
 - in dynamic reconfiguration *Vol 2* 24-5
 - SSA subsystem *Vol 1* 5-17
- disk layout
 - planning issues *Vol 2* H-8
- disk subsystems
 - supported for HACMP/ES *Vol 1* 1-3
- disks
 - 2105 Versatile Storage Server *Vol 1* 1-5
 - 7135 RAIDiant Disk Array *Vol 1* 1-3, 5-3
 - 7137 Disk Arrays *Vol 1* 1-4, 5-5
 - concurrent access supported *Vol 2* 23-1
 - configuring a quorum buster *Vol 1* 6-8
 - defining to cluster using C-SPOC *Vol 2* 22-42
 - defining to the cluster using C-SPOC *Vol 2* 22-42
 - IBM 2105 Versatile Storage Server *Vol 1* 5-5
 - planning issues *Vol 2* H-8
 - planning shared disk devices *Vol 1* 5-1
 - RAID devices *Vol 2* 23-1
 - SSA subsystem *Vol 1* 1-4, 5-5
- DNS
 - integrating with HACMP *Vol 2* H-7
 - with HACMP/ES *Vol 1* 4-14
- documentation
 - SSA installation and maintenance *Vol 1* 5-13
- domain nameserving
 - and HACMP/ES usage *Vol 1* 4-15
- domain, operational
 - for Group Services *Vol 2* 30-3
- dynamic
 - adapter swap *Vol 2* 24-8

dynamic reconfiguration
 and disk fencing *Vol 1* 5-18
 DARE Resource Migration *Vol 2* 24-31
 effect of disk fencing *Vol 2* 24-5
 effect on resources *Vol 2* 24-41, 26-7
 emulating *Vol 2* 21-36
 of cluster topology *Vol 2* 24-1
 releasing the SCD lock *Vol 2* 24-14
 restoring the DCD from the ACD *Vol 2* 24-15
 scripts *Vol 2* 19-8
 triggered by applied snapshot *Vol 2* 26-6
 undoing *Vol 2* 24-15

E

Eclock command *Vol 1* 13-6
 editing
 ./rhosts file *Vol 1* 13-3
 /etc/hosts file *Vol 1* 13-2
 /etc/rc.net *Vol 1* 13-4
 cldhosts file *Vol 1* 16-1
 on client *Vol 1* 17-2
 clinfo.rc on client *Vol 1* 17-3
 clinfo.rc script *Vol 1* 16-2
 rules file *Vol 1* 8-4
 snmpd.conf file *Vol 2* 20-15
 EM client (Event Management client)
 definition *Vol 2* 31-1
 registration request cache *Vol 2* 31-6
 restrictions on number *Vol 2* 31-11
 EMAPI (Event Management Application Programming
 Interface)
 component of Event Management *Vol 2* 31-2
 EMAPI libraries
 location *Vol 2* 31-2
 embedded resource monitor
 type of resource monitor *Vol 2* 31-3
 EMCDB (Event Management Configuration Database)
 and Event Manager daemon initialization *Vol 2*
 31-11
 run-time directory *Vol 2* 31-7
 run-time file *Vol 2* 31-5
 version string and joining peer group *Vol 2* 31-10
 version string and peer group state *Vol 2* 31-7
 version string directory *Vol 2* 31-7
 emsvcs daemon *Vol 2* 20-2
 emsvcsctrl command
 cleaning the Event Management subsystem
Vol 2 31-9
 control script for Event Management *Vol 2* 31-5
 deleting the Event Management subsystem
Vol 2 31-9
 starting the Event Management subsystem
Vol 2 31-9
 stopping the Event Management subsystem
Vol 2 31-9
 summary of functions *Vol 2* 31-8
 tracing the Event Management subsystem
Vol 2 31-9
 emsvcsctrl script *Vol 2* F-1
 emulating
 cluster events *Vol 2* 21-32
 error log entries *Vol 1* 13-8
 enabling
 asynchronous event notification *Vol 2* 20-15
 SSA disk fencing *Vol 1* 5-18
 target mode SSA interface *Vol 1* 11-18
 enhanced security
 Kerberos *Vol 1* 18-32
 enMembership group *Vol 2* 31-4
 use by Event Management *Vol 2* 31-8
 Eprimary management
 SP Switch *Vol 1* 4-11
 error emulation *Vol 1* 13-8
 error message output log
 Event Management *Vol 2* 31-6
 error notification *Vol 1* 13-5
 automatic *Vol 1* 13-6
 customizing *Vol 2* H-3
 errpt command *Vol 2* 29-13
 Estart command *Vol 1* 13-6
 event customization facility *Vol 1* 8-1
 Event Management
 disk space and tracing *Vol 2* 31-9
 performance and tracing *Vol 2* 31-9
 shared variables with performance monitoring
Vol 2 31-7
 Event Management API (EMAPI)
 component of Event Management *Vol 2* 31-2
 Event Management client (EM client)
 definition *Vol 2* 31-1
 Event Management communications
 between Event Manager daemons *Vol 2* 31-4
 local EM clients *Vol 2* 31-4
 resource monitors *Vol 2* 31-4
 Event Management Configuration Database (EMCDB)
 and Event Manager daemon initialization
Vol 2 31-11
 run-time directory *Vol 2* 31-7
 version string and joining peer group *Vol 2* 31-10
 version string and peer group state *Vol 2* 31-7
 version string directory *Vol 2* 31-7

- Event Management directories
 - /etc/ha/cfg *Vol 2* 31-7
 - /var/ha/lck/haem *Vol 2* 31-5
 - /var/ha/log *Vol 2* 31-6
 - /var/ha/run *Vol 2* 31-6
 - /var/ha/soc *Vol 2* 31-5
 - Event Management subsystem
 - cleaning with emsvcsctrl command *Vol 2* 31-9
 - client communication *Vol 2* 31-2
 - components *Vol 2* 31-7
 - configuration *Vol 2* 31-8 *Vol 2* 31-9
 - configuring and operating *Vol 2* 31-1
 - control script *Vol 2* F-1
 - deleting with emsvcsctrl command *Vol 2* 31-9
 - dependencies *Vol 2* 31-7
 - Event Manager daemon initialization *Vol 2* 31-9
 - Event Manager daemon operation *Vol 2* 31-11, *Vol 2* 31-12
 - getting subsystem status *Vol 2* 31-12, 31-13
 - installation *Vol 2* 31-8
 - introducing *Vol 2* 31-1
 - recovery from failure (automatic) *Vol 2* 31-11
 - starting with emsvcsctrl command *Vol 2* 31-9
 - stopping with emsvcsctrl command *Vol 2* 31-9
 - tracing with emsvcsctrl command *Vol 2* 31-9
 - Event Manager daemon
 - abnormal termination core file *Vol 2* 31-6
 - communications *Vol 2* 31-4
 - component of Event Management *Vol 2* 31-2
 - current working directory *Vol 2* 31-6
 - definition *Vol 2* 31-1
 - emsvcsctrl control script *Vol 2* 31-5
 - error message log file *Vol 2* 31-6
 - getting status *Vol 2* 31-12, 31-13
 - group membership *Vol 2* 31-7
 - group subscriptions *Vol 2* 31-8
 - initialization *Vol 2* 31-9
 - message trace output log file *Vol 2* 31-6
 - operation *Vol 2* 31-11 *Vol 2* 31-12
 - recovery from failure (automatic) *Vol 2* 31-11
 - trace output log file *Vol 2* 31-6
 - use of AIX error log *Vol 2* 31-6
 - event roll-up *Vol 1* 8-6
 - events
 - changing custom events processing *Vol 2* 24-42, 24-44
 - emulating *Vol 2* 21-33
 - emulating dynamic reconfiguration *Vol 2* 21-36
 - emulating events *Vol 2* 21-32
 - messages relating to *Vol 2* 29-1
 - notification *Vol 1* 8-2
 - planning *Vol 1* 8-1
 - priorities *Vol 1* 8-6
 - recovery *Vol 1* 8-3
 - retry *Vol 1* 8-3
 - roll-up *Vol 1* 8-6
 - user-defined *Vol 1* 8-3
 - extending
 - shared volume groups *Vol 2* 22-7
- ## F
- failed disk drive
 - replacing *Vol 2* 23-11
 - failure detection rate
 - changing *Vol 1* 4-19, 4-20
 - network module *Vol 2* 24-12
 - setting for network modules *Vol 1* 18-7
 - tuning network modules *Vol 2* 24-11
 - fallover
 - intentional *Vol 2* 20-7
 - Fast Connect
 - configuring *Vol 1* 18-11
 - converting from AIX Connections *Vol 1* 18-11 *Vol 2* 24-26
 - defined *Vol 1* 3-12, 18-11
 - planning *Vol 1* 3-12
 - reconfiguring in SMIT *Vol 2* 24-29
 - verification *Vol 1* 18-12
 - worksheet *Vol 1* A-31
 - FDDI hardware address
 - specifying *Vol 1* 4-28
 - file systems
 - as shared LVM component *Vol 1* 6-3
 - mount failures *Vol 2* H-12
 - shared
 - changing *Vol 2* 22-23
 - maintaining *Vol 2* 22-18
 - removing *Vol 2* 22-26
 - files and directories
 - component of Event Management *Vol 2* 31-5
 - component of Group Services *Vol 2* 30-4
 - filesets, Event Management
 - rsct.basic.rte *Vol 2* 31-8
 - rsct.clients.sp *Vol 2* 31-8
 - firstboot file
 - use in node by node migration *Vol 1* 14-16
 - forced inline mode
 - SSA disk subsystem configuration *Vol 1* 5-15
 - fuser command
 - using in scripts *Vol 2* H-12
- ## G
- generating
 - trace report *Vol 2* 29-21
 - global networks
 - changing configuration *Vol 2* 24-11
 - defining *Vol 1* 18-5
 - graceful stops
 - of cluster services on node *Vol 2* 20-7
 - of cluster services on nodes
 - with takeover *Vol 2* 20-7

- group accounts
 - listing *Vol 2 27-6*
 - managing *Vol 2 27-5*
 - Group Leader
 - Topology Services daemon *Vol 2 32-1*
 - group membership list
 - definition *Vol 2 30-1*
 - Group Services
 - definition of client *Vol 2 30-1*
 - definition of group *Vol 2 30-1*
 - disk space and tracing *Vol 2 30-6*
 - performance and tracing *Vol 2 30-6*
 - protocol *Vol 2 30-1*
 - viewing configuration *Vol 1 18-5*
 - Group Services API (GSAPI)
 - component of Group Services *Vol 2 30-3*
 - Group Services communications
 - between Group Services daemons *Vol 2 30-4*
 - local GS clients *Vol 2 30-4*
 - Group Services daemon
 - abnormal termination core file *Vol 2 30-5*
 - communications *Vol 2 30-4*
 - component of Group Services *Vol 2 30-3*
 - current working directory *Vol 2 30-5*
 - getting status *Vol 2 30-8*
 - grpsvcctrl control script *Vol 2 30-4*
 - initialization *Vol 2 30-6*
 - operation *Vol 2 30-8*
 - trace output log file *Vol 2 30-5*
 - Group Services directories
 - /var/ha/lck* *Vol 2 30-4*
 - /var/ha/log* *Vol 2 30-5*
 - /var/ha/run* *Vol 2 30-5*
 - /var/ha/soc* *Vol 2 30-4*
 - Group Services messages *Vol 2 G-1*
 - Group Services subsystem
 - and Event Manager daemon initialization *Vol 2 31-10*
 - client communication *Vol 2 30-3*
 - component summary *Vol 2 30-2*
 - components *Vol 2 30-2*
 - configuration *Vol 2 30-6*
 - configuring and operating *Vol 2 30-1*
 - control script *Vol 2 F-6*
 - dependencies *Vol 2 30-5*
 - dependency by Event Management *Vol 2 31-7*
 - getting subsystem status *Vol 2 30-8*
 - Group Services daemon initialization *Vol 2 30-6*
 - Group Services daemon operation *Vol 2 30-8*
 - initialization errors *Vol 2 30-8*
 - installation *Vol 2 30-6*
 - introducing *Vol 2 30-1*
 - operational domain *Vol 2 30-3*
 - recovery from failure (automatic) *Vol 2 30-8*
 - system groups *Vol 2 31-4*
 - tracing with grpsvcctrl command *Vol 2 30-6*
 - group state
 - and joining the ha_em_peers group *Vol 2 31-10*
 - EMCDB version string *Vol 2 31-7*
 - group state value
 - definition *Vol 2 30-1*
 - groups
 - adding *Vol 2 27-7*
 - changing *Vol 2 27-7*
 - Group Services
 - restrictions on number per client *Vol 2 30-8*
 - restrictions on number per domain *Vol 2 30-8*
 - removing *Vol 2 27-8*
 - grpplsmd daemon *Vol 2 20-3*
 - grpsvcctrl
 - Group Services control script *Vol 2 30-4*
 - grpsvcctrl command
 - control script for Group Services *Vol 2 30-4*
 - summary of functions *Vol 2 30-6*
 - tracing the Group Services subsystem *Vol 2 30-6*
 - grpsvcctrl script *Vol 2 F-6*
 - grpsvcsd daemon *Vol 2 20-3*
 - GS client (Group Services client)
 - restrictions on number *Vol 2 30-8*
 - GS nameserver
 - establishing *Vol 2 30-7*
 - GSAPI (Group Services Application Programming Interface)
 - component of Group Services *Vol 2 30-3*
 - GSAPI libraries
 - location *Vol 2 30-3*
 - guidelines
 - for planning a cluster *Vol 1 2-1*
- ## H
- HA_DOMAIN_NAME environment variable *Vol 2 30-3*
 - ha_em_peers group
 - joining *Vol 2 31-10*
 - use by Event Management *Vol 2 31-7*
 - HACMP
 - specified operating environment *Vol 1 10-3*
 - supported hardware *Vol 1 10-3*
 - HACMP for AIX
 - converting to HACMP/ES *Vol 1 14-11*
 - converting with snapshot *Vol 1 14-12*
 - migration to HACMP/ES *Vol 1 14-13*
 - HACMP log files
 - using cron to maintain *Vol 2 H-20*

HACMP/ES

- AIX tasks *Vol 1* 13-1
- Concurrent Resource Manager *Vol 1* 14-10, 15-5
- event scripts *Vol 2* 19-7
- installation server *Vol 1* 14-4
- installing cluster
 - list of steps *Vol 1* 10-1
- new features *Vol 1* 1-12
- restrictions *Vol 1* 1-12
- scripts *Vol 2* 19-6
- starting and stopping *Vol 2* 20-1
- starting on clients *Vol 2* 20-14
- stopping on clients *Vol 2* 20-14
- system backup *Vol 2* 19-3
- verifying installation of software *Vol 1* 14-11
- HACMP/ES nameserving *Vol 1* 4-14
- HACMP/ES software
 - installing on nodes *Vol 1* 14-1
- haemd daemon *Vol 2* 20-2, F-10
 - location *Vol 2* 31-2
- haemd_HACMP command *Vol 2* F-11
- haemqvar command *Vol 2* F-12
- haemtrcoff command *Vol 2* F-16
- haemtrcon command *Vol 2* F-18
- haemunlkrm command *Vol 2* F-20
- hagsd daemon *Vol 2* F-21
 - location *Vol 2* 30-3
- hagsglsmd daemon *Vol 2* F-22
- HANFS
 - functionality added to HACMP/ES *Vol 1* 1-12, 6-9, 18-25
- hard disk
 - installing HACMP/ES *Vol 1* 14-8
- hardware
 - checking installation *Vol 1* 11-1
 - guidelines for maintenance *Vol 2* H-8
 - required/supported hardware in HACMP *Vol 1* 10-3
- hardware address swapping
 - ATM adapters *Vol 1* 4-28
 - planning *Vol 1* 4-26
- hardware errors
 - list of errors to monitor *Vol 2* H-3
- hardware maintenance
 - proper procedures *Vol 2* H-14

HAView

- and the clhosts file *Vol 2* 21-3
- and the haview_start file *Vol 2* 21-3
- and the snmpd.conf file *Vol 2* 21-4
- browsers provided *Vol 2* 21-12
- cluster administration utility *Vol 2* 21-12
- cluster topology symbols
 - interpreting colors *Vol 2* 21-6
- deleting objects *Vol 2* 21-11
- individual resource symbols *Vol 2* 21-9
- installation notes *Vol 1* 14-6
- monitoring a cluster *Vol 2* 21-2
- NetView hostname requirements *Vol 2* 21-4
- NetView navigation tree *Vol 2* 21-7
- read-only maps *Vol 2* 21-6
- resource group symbols
 - defined *Vol 2* 21-8
 - interpreting colors *Vol 2* 21-9
 - starting *Vol 2* 21-4
- heartbeat rate
 - definition *Vol 1* 4-19
 - tuning *Vol 2* 24-11
- high availability services
 - Event Management subsystem *Vol 2* 31-1
 - Group Services subsystem *Vol 2* 30-1
 - Topology Services subsystem *Vol 2* 32-1
- high water mark
 - setting *Vol 1* 4-18
- host membership
 - resource monitor *Vol 2* 31-4
- HostMembership group *Vol 2* 31-4
 - use by Event Management *Vol 2* 31-8
- hostname resolution
 - HACMP/ES logic *Vol 1* 4-15
- HPS
 - adapter function *Vol 1* 18-4
- HPS switch
 - upgrading to SP switch *Vol 1* 11-2

I

- I/O
 - setting pacing *Vol 1* 4-18
- I/O pacing *Vol 1* 4-18, 13-1
- IBM disk subsystems and arrays *Vol 1* 5-1
- IBM RS/6000 Cluster Technology *Vol 1* 1-9
- IBM.RSCT.harmpd resource monitor *Vol 2* 31-3
- images
 - installation server *Vol 1* 14-4
- importing
 - concurrent volume groups
 - C-SPOC *Vol 2* 23-19
 - shared volume group
 - using C-SPOC *Vol 2* 22-13
 - volume groups *Vol 1* 12-6
- importvg command
 - with shared LVM components *Vol 1* 12-9

inactive takeover
 and cascading without fallback *Vol 1* 18-28
 cascading resource group *Vol 1* 18-28
 Vol 2 24-29

installation
 checking problems *Vol 1* 14-11, 15-6
 Event Management subsystem *Vol 2* 31-8
 Group Services subsystem *Vol 2* 30-6
 hard disk *Vol 1* 14-8
 images *Vol 1* 14-4
 media *Vol 1* 14-7

installing
 concurrent access system *Vol 1* 14-10, 15-5
 from installation server *Vol 1* 14-4
 HACMP/ES cluster
 list of steps *Vol 1* 10-1
 HACMP/ES on clients *Vol 1* 17-1
 HACMP/ES software on servers *Vol 1* 14-1
 saved snapshot
 converting from HACMP for AIX
 Vol 1 14-13
 SSA serial disk subsystem *Vol 1* 11-20

intentional fallover *Vol 2* 20-7

internal resource monitor
 definition *Vol 2* 31-3
 supplied by RSCT *Vol 2* 31-4

IP address
 changing properly *Vol 2* H-11

IP address aliasing
 for monitoring with Tivoli *Vol 1* D-6
 SP Switch configuration *Vol 1* 4-12

IP address takeover
 /etc/inittab *Vol 2* 19-4
 configuring on SP *Vol 1* 4-9

IP source routing
 setting required by Topology Services *Vol 2* F-23

IPX/SPX protocol interference *Vol 1* C-5

J

jfslog *Vol 1* 6-3
 mirroring *Vol 1* 6-5
 renaming *Vol 1* 12-4 *Vol 2* 22-19

joining a Group Services group
 Event Management subsystem *Vol 2* 31-10

K

keepalives
 tuning *Vol 1* 4-20, *Vol 2* 24-12

Kerberos
 enhanced security *Vol 1* 18-32

kill - 9 command
 warning *Vol 2* H-10

L

LAN adapter
 replacing *Vol 2* H-15

LAN emulation for ATM
 configuring in AIX *Vol 1* 11-8

LANG variable
 setting for F1 help *Vol 1* 14-3

licenses
 software *Vol 1* 5-7

local EM clients
 Event Management communications *Vol 2* 31-4

local GS clients
 Group Services communications *Vol 2* 30-4

lock
 SCD
 releasing *Vol 2* 24-14

lock file
 Event Management *Vol 2* 31-5
 Group Services *Vol 2* 30-4

log files
 /tmp/clstrmgr.debug *Vol 2* 29-4
 /tmp/dms_loads.out *Vol 2* 29-4
 /tmp/hacmp.out file *Vol 2* 29-3, 29-8
 changing default pathnames *Vol 1* 18-40
 cluster.log file *Vol 2* 29-3, 29-5
 cluster.mmdd *Vol 2* 29-4, 29-15
 Event Management *Vol 2* 31-6
 Group Services *Vol 2* 30-5
 monitoring a cluster *Vol 2* 21-38
 recommended use *Vol 2* 29-3
 redirecting *Vol 1* 18-40
 system error log *Vol 2* 29-3, 29-12
 types of *Vol 2* 29-2
 with cluster messages *Vol 2* 29-2

log logical volume
 renaming *Vol 1* 12-4 *Vol 2* 22-19

logical volumes
 adding copies *Vol 1* 12-5 *Vol 2* 22-20
 as shared LVM component *Vol 1* 6-3
 changing
 using AIX commands *Vol 2* 22-33
 using C-SPOC *Vol 2* 22-32
 creating *Vol 2* 22-29
 journal logs *Vol 1* 6-5
 removing
 using AIX commands *Vol 2* 22-35
 using C-SPOC *Vol 2* 22-34, 23-25
 renaming *Vol 1* 12-4 *Vol 2* 22-19
 shared
 adding mirror copies *Vol 2* 22-36
 adding with C-SPOC *Vol 2* 22-31
 maintaining *Vol 2* 22-29
 setting characteristics with C-SPOC
 Vol 2 22-39
 synchronizing shared mirrors
 using C-SPOC *Vol 2* 22-41

loopback address
 cldare file *Vol 1* 16-1

low-water mark
 setting *Vol 1* 4-18

lssrc command
 getting Event Management status *Vol 2* 31-12
 getting Group Services status *Vol 2* 30-8

LVM
 forcing an update of ODM data on remote nodes
Vol 2 22-4
 mirroring *Vol 1* 6-4
 planning guidelines *Vol 1* 6-14
 updating ODM definitions on remote nodes
Vol 2 22-3

M

machines list file
 for topsvcs *Vol 2* 32-2

maintaining
 7 X 24 cluster *Vol 2* H-1
 concurrent access environment *Vol 2* 23-1
 HACMP/ES system *Vol 2* 19-1
 NFS *Vol 2* 28-1
 shared LVM components *Vol 2* 22-1
 using C-SPOC *Vol 2* 22-2

managing
 cluster resources *Vol 2* 24-23

manuals
 SSA installation and maintenance *Vol 1* 5-13

mapping
 between events and recovery programs *Vol 1* 8-3

maximum
 chars in node name *Vol 1* 3-5
 cluster ID number *Vol 1* 3-5

membership resource monitor *Vol 2* 31-4

message catalogs
 setting LANG variable *Vol 1* 14-3

message trace output log
 Event Management *Vol 2* 31-6

messages
 cluster state *Vol 2* 29-2
 event notification *Vol 2* 29-1
 group services *Vol 2* G-1
 in verbose mode *Vol 2* 29-1
 resource monitor *Vol 2* G-28
 RS/6000 Cluster Technology *Vol 2* G-70
 topology *Vol 2* G-48

messaging
 required setting of IP source routing *Vol 2* F-23

migrating data
 on shared physical volumes *Vol 2* 22-45

migrating resources
 default location keyword *Vol 2* 24-33
 example using cldare *Vol 2* 24-35
 non-sticky migration *Vol 2* 24-32
 sticky migration *Vol 2* 24-31
 stop location keyword *Vol 2* 24-33
 using SMIT *Vol 2* 24-36
 using the cldare command *Vol 2* 24-34

migration
 backout procedure *Vol 1* 14-18
 cluster snapshots saved *Vol 1* 14-17
 from HACMP for AIX
 node by node *Vol 1* 14-13
 handling node failure during process *Vol 1* 14-17

migration flag
 used for node by node migration *Vol 1* 14-16

mirroring
 concurrent volume groups
 C-SPOC *Vol 2* 23-21
 jfslog *Vol 1* 6-5, 12-5 *Vol 2* 22-20
 logical partitions *Vol 1* 12-5
 logical volumes *Vol 2* 22-20
 LVM *Vol 1* 6-4
 physical partitions *Vol 1* 6-4
 shared volume group
 C-SPOC *Vol 2* 22-15

mksysb backups *Vol 2* H-20

modifying
 shared LVM components *Vol 2* 22-1

monitoring a cluster
 applications *Vol 2* 21-29
 cluster services *Vol 2* 21-37
 overview *Vol 2* 21-1
 tools for monitoring *Vol 2* 21-2
 using clRGinfo and clfindres *Vol 2* 21-30

monitoring applications
 changing the monitor configuration *Vol 2* 24-22
 configuring *Vol 1* 18-19
 overview *Vol 1* 18-17
 prerequisites *Vol 1* 18-18
 removing a monitor *Vol 2* 24-22
 suspending and resuming monitoring *Vol 2* 24-21
 troubleshooting *Vol 2* 29-24

mounting
 NFS *Vol 1* 6-11 *Vol 2* 28-2

Multi-Initiator RAID adapters
 connection for TMSSA *Vol 1* 11-18

N

name resolution
 integrating with HACMP *Vol 2* H-6

named daemon *Vol 1* 4-17

nameserver
 establishing for Group Services *Vol 2* 30-7

nameserver configuration
 and boot address *Vol 1* 13-2

- nameserving
 - configuring run-time parameters *Vol 2* 28-6
 - enabling and disabling under HACMP/ES
 - Vol 1* 4-15
 - under HACMP/ES *Vol 1* 4-14
- naming
 - resource groups *Vol 1* 18-9
- NetView
 - dialog boxes *Vol 2* 21-10
 - traps *Vol 2* 20-15
 - using HAView *Vol 2* 21-2
- network adapters
 - configuring for boot address *Vol 1* 11-1
 - configuring in AIX *Vol 1* 11-1
 - defining to cluster *Vol 1* 18-2, 18-5
 - Vol 2* 24-11
 - functions *Vol 1* 4-9
 - monitoring with clstat *Vol 2* 21-28
 - removing from cluster definition *Vol 2* 24-10
- network attribute
 - supported types *Vol 1* 18-3
- network daemons
 - starting *Vol 2* 20-10
- network interfaces
 - making changes *Vol 2* H-11
- network loads
 - handling problems *Vol 2* H-11
- network modules
 - changing or showing *Vol 1* 18-7
 - changing or showing in SMIT *Vol 2* 24-12
 - configuring *Vol 1* 18-6, *Vol 2* 24-11
 - failure detection parameters *Vol 1* 4-18
- network options
 - tailoring AIX *Vol 1* 13-2
- network services
 - integrating with HACMP *Vol 2* H-6
- networks
 - ATM *Vol 1* 11-3
 - changing configuration of global network
 - Vol 2* 24-11
 - defining global networks *Vol 1* 18-5
 - defining in SMIT *Vol 1* 18-3
 - multiple paths *Vol 1* 4-3
 - NFS mounting filesystems and directories
 - Vol 1* 18-26 *Vol 2* 24-28
 - planning TCP/IP connections *Vol 1* 4-1
 - replacing hardware *Vol 2* H-15
 - strategies for handling failures *Vol 1* 4-5
 - supported by HACMP/ES *Vol 1* 4-4
- NFS
 - caveats about node names *Vol 1* 6-14
 - creating shared volume groups *Vol 1* 6-10
 - Vol 2* 28-1
 - cross mounting filesystems *Vol 1* 6-11
 - exporting filesystems and directories *Vol 1* 6-10
 - maintaining *Vol 2* 28-1
 - mount issues *Vol 1* 6-11 *Vol 2* 28-2
 - mounting filesystems *Vol 1* 6-11 *Vol 2* 28-2
 - mounting filesystems and directories *Vol 2* 24-27
 - nested mount points *Vol 1* 6-12
 - node name issues *Vol 2* 28-5
 - planning *Vol 1* 6-10
 - reliable server functionality *Vol 1* 6-10,
 - Vol 2* 28-5
 - setting up mount points for cascading groups
 - Vol 1* 6-11
 - takeover issues *Vol 1* 6-11 *Vol 2* 28-2
- NFS-Exported File System Worksheet *Vol 1* 6-16, A-21
- NIS
 - configuring run-time parameters *Vol 2* 28-6
 - integrating with HACMP *Vol 2* H-7
 - with HACMP/ES *Vol 1* 4-14
- NIS or name server
 - setting on node *Vol 1* 18-30
- NIS services and cron *Vol 1* 13-2
- node environment
 - verifying *Vol 1* 18-37
- node isolation *Vol 1* 4-6
- node names
 - issues with NFS *Vol 1* 6-14
- node_down_local
 - concurrent access environment *Vol 2* 23-2
- node_up_local event
 - concurrent access environment *Vol 2* 23-2
- node-by-node migration
 - from HACMP to HACMP/ES *Vol 1* 14-13
- nodes
 - adding nodes *Vol 2* 24-4
 - cascading resource group priorities *Vol 2* 24-23
 - changing configuration *Vol 2* 24-4
 - changing name *Vol 2* 24-6
 - changing node number for TMSSA *Vol 1* 11-18
 - defining to cluster *Vol 1* 18-2
 - deleting from the cluster *Vol 2* 24-5
 - initial planning *Vol 1* 3-2
 - naming *Vol 1* 3-7 *Vol 2* 24-6
 - procedure for replacing hardware *Vol 2* H-14
 - removing *Vol 2* 24-5
- nonlocsrcroute option of no command
 - setting required for Topology Services *Vol 2* F-23
- Non-Shared Volume Group Worksheet *Vol 1* A-17, A-23
- non-sticky migration *Vol 2* 24-32
- notification methods
 - testing *Vol 1* 13-9

O

- obtaining trace information
 - using cldiag *Vol 2 29-22*
- ODM data
 - processing during reconfiguration *Vol 2 24-15*
 - restoring the DCD *Vol 2 24-15*
 - saved in cluster snapshot *Vol 2 26-3*
- Online backups *Vol 2 H-20*
- online planning worksheets
 - installing *Vol 1 B-2*
 - overview *Vol 1 B-1*
 - using *Vol 1 B-2, B-5*
- Oracle Database and HACMP *Vol 1 C-7*

P

- partitioned cluster *Vol 1 4-6*
- peer group, Event Management
 - joining *Vol 2 31-10*
- performance *Vol 1 4-17*
- Performance Aide for AIX
 - dependency by Event Management *Vol 2 31-7*
- performance monitoring
 - shared variables with Event Management *Vol 2 31-7*
- physical partitions
 - mirroring *Vol 1 6-4*
- physical volume
 - shared
 - removing using C-SPOC *Vol 2 22-44*
 - adding to cluster using C-SPOC *Vol 2 22-42*
 - maintaining *Vol 2 22-42*
- physical volumes
 - as shared LVM components *Vol 1 6-2*
 - shared
 - migrating data *Vol 2 22-45*
- planning
 - application servers *Vol 1 3-2*
 - applications *Vol 1 3-2, 3-4, C-1*
 - CS/AIX communications links *Vol 1 3-16*
 - for 7 X 24 maintenance *Vol 2 H-2*
 - for software maintenance *Vol 2 H-9*
 - HACMP/ES clients *Vol 1 9-1*
 - HACMP/ES cluster
 - applications and application servers *Vol 1 3-2*
 - list of steps *Vol 1 2-3*
 - resource groups *Vol 1 7-1*
 - resources *Vol 1 3-4*
 - HACMP/ES cluster events *Vol 1 8-1*
 - NFS *Vol 1 6-10*
 - shared disks
 - 7135 RAIDiant Disk Array *Vol 1 1-3*
 - 7137 Disk Array *Vol 1 1-4*
 - SSA disk subsystem *Vol 1 1-4*

- shared LVM components *Vol 1 6-1*
 - file systems *Vol 1 6-3*
 - logical volumes *Vol 1 6-3*
 - physical volumes *Vol 1 6-2*
 - volume groups *Vol 1 6-3*
- shared SCSI disks *Vol 1 5-2*
- shared SSA disk subsystem *Vol 1 5-5*
- TCP/IP networks *Vol 1 4-1*
- polling interval
 - HAView
 - changing *Vol 2 21-10*
 - defined *Vol 2 21-10*
- port numbers
 - component of Event Management *Vol 2 31-4*
 - component of Group Services *Vol 2 30-3*
- post-processing
 - cluster events *Vol 1 8-2*
- power supplies
 - and shared disks *Vol 1 5-6*
- predicate
 - definition *Vol 2 31-1*
- pre-processing
 - cluster events *Vol 1 8-2*
- prerequisites
 - for installing HACMP/ES *Vol 1 14-1*
- preventive maintenance *Vol 2 H-19*
- print jobs
 - making queue highly available *Vol 1 18-45*
- print queues
 - custom script *Vol 1 18-45*
- priorities
 - nodes
 - for cascading resources *Vol 2 24-23*
- procedure
 - replace failed SSA disk *Vol 2 H-17*
- process application monitoring *Vol 1 18-19*
- process resource monitor *Vol 2 31-3*
- provider
 - definition *Vol 2 30-1*

Q

- quorum
 - using with HACMP *Vol 1 6-5*
- quorum buster disk *Vol 1 6-8*

R

- RAID devices
 - concurrent access environment *Vol 2 23-1*
 - importing volume groups *Vol 1 12-10*
 - Vol 2 23-8*
 - shared LVM components *Vol 1 12-2*
 - concurrent access *Vol 1 12-7*
- rc.cluster script
 - starting clients *Vol 2 20-14*

- reconfiguring
 - processing ODM data *Vol 2* 24-15
- recovery
 - SP Switch failures *Vol 1* 13-5
- recovery from failure
 - Event Management *Vol 2* 31-11
 - Group Services *Vol 2* 30-8
- redirection
 - of cluster log files *Vol 1* 18-40
- reducing
 - shared volume group *Vol 2* 22-9
- registration cache
 - and Event Manager daemon *Vol 2* 31-9
 - Event Management *Vol 2* 31-6
- reintegration
 - of cluster nodes *Vol 2* 20-9
- Reliable NFS server *Vol 1* 6-9
- reliable NFS server *Vol 1* 6-10
- removing
 - application servers *Vol 2* 24-17, 24-19
 - cluster nodes *Vol 2* 24-5
 - cluster snapshot *Vol 2* 26-8
 - concurrent logical volume
 - C-SPOC *Vol 2* 23-25
 - copy of concurrent logical volume
 - C-SPOC *Vol 2* 23-26
 - filesystems
 - using C-SPOC *Vol 2* 22-3
 - HACMP/ES network from global network
 - Vol 2* 24-11
 - logical volumes
 - using C-SPOC *Vol 2* 22-3
 - network adapters *Vol 2* 24-10
 - physical volume from concurrent volume group
 - C-SPOC *Vol 2* 23-20
 - previous version of HACMP *Vol 1* 14-12
 - resource groups *Vol 2* 24-24
 - shared file system *Vol 2* 22-26
 - shared volume group *Vol 2* 22-11
 - C-SPOC *Vol 2* 22-14
- renaming
 - log logical volume *Vol 1* 12-4
 - logical volumes *Vol 1* 12-4 *Vol 2* 22-19
- replacing
 - cluster node *Vol 2* H-14
 - failed disk drive *Vol 2* 23-11
 - LAN adapter *Vol 2* H-15
 - mirrored disk drive *Vol 2* H-16
 - network hardware *Vol 2* H-15
 - topology hardware *Vol 2* H-14
- resource chains
 - setting up *Vol 1* 18-9
- Resource Group Worksheet *Vol 1* A-43
- resource groups
 - adding *Vol 2* 24-23
 - cascading without fallback *Vol 2* 24-30
 - changing *Vol 2* 24-24
 - changing resources in *Vol 2* 24-25
 - creating *Vol 1* 18-9
 - IP address requirements *Vol 1* 4-10
 - monitoring status and location *Vol 2* 21-30
 - cIRMupdate and cIRGINfo commands
 - Vol 2* E-19
 - planning *Vol 1* 7-1
 - relation to C-SPOC utility *Vol 2* 22-3
 - removing *Vol 2* 24-24
- resource monitor
 - component of Event Management *Vol 2* 31-3
 - connection types *Vol 2* 31-3
 - current working directory *Vol 2* 31-6
 - definition *Vol 2* 31-1
 - Event Management communications *Vol 2* 31-4
- Resource Monitor Application Programming Interface (RMAPI)
 - component of Event Management *Vol 2* 31-3
 - use of EMCDB version string directory *Vol 2* 31-7
 - use of shared memory key *Vol 2* 31-5
- resource monitor messages *Vol 2* G-28
- resource monitors supplied by RSCT *Vol 2* 31-3
 - aixos *Vol 2* 31-3
 - Membership (internal) *Vol 2* 31-4
- resource variable
 - definition *Vol 2* 31-1
 - IBM.PSSP.Membership.LANAdapter.state
 - Vol 2* 31-8
 - IBM.PSSP.Membership.Node.state *Vol 2* 31-8
- resources
 - changing in resource group *Vol 2* 24-26
 - effect of dynamic reconfiguration *Vol 2* 24-41, 26-7
 - managing *Vol 2* 24-23
 - migrating dynamically *Vol 2* 24-34
 - selecting type during planning phase *Vol 1* 3-4
 - synchronizing *Vol 1* 18-30 *Vol 2* 24-41
- restarting
 - concurrent access daemon *Vol 2* 23-16
- restoring
 - saved cluster configurations dynamically
 - Vol 2* 26-1

- restrictions
 - Event Management
 - number of EM clients *Vol 2* 31-11
 - observation of resource variable values
Vol 2 31-11
 - Group Services
 - groups per client *Vol 2* 30-8
 - groups per domain *Vol 2* 30-8
 - number of GS clients *Vol 2* 30-8
 - Topology Services
 - IP source routing setting *Vol 2* F-23
- RMAPI (Resource Monitor Application Programming Interface)
 - use of EMCDB version string directory
Vol 2 31-7
 - use of shared memory key *Vol 2* 31-5
- RMAPI libraries
 - location *Vol 2* 31-3
- rootvg
 - planning volume groups *Vol 2* H-8
- routerevalidate network option
 - changing setting *Vol 1* 13-2
- routing, IP source
 - setting required for Topology Services *Vol 2* F-23
- RS/6000 Cluster Technology
 - common messages *Vol 2* G-70
- RS/6000 SP system
 - See SP *Vol 2* E-8
- RS232
 - checking connections *Vol 1* 11-2
- RSCT
 - installation images *Vol 1* 14-4
 - services for high availability *Vol 1* 1-9
- RSCT error messages
 - Event Management messages *Vol 2* G-2
 - Group Services messages *Vol 2* G-1
 - resource monitor messages *Vol 2* G-28
 - RSCT common messages *Vol 2* G-48
 - Topology Services messages *Vol 2* G-48
- RSCT services
 - common commands and utilities *Vol 2* F-1
 - Event Management *Vol 2* 31-1
 - Group Services *Vol 2* 30-1
 - Topology Services *Vol 2* 32-1
- rsct.basic.rte fileset *Vol 2* 30-6
- rsct.basic.sp fileset *Vol 2* 30-6
- rules file
 - for user-defined events *Vol 1* 8-4
- run-time EMCDB
 - creating *Vol 2* 31-5 *Vol 2* 31-11
- run-time maintenance *Vol 2* H-9
- run-time parameters
 - changing on a node *Vol 2* 28-5
 - configuring *Vol 1* 18-29

S

- SAP R/3 and HACMP *Vol 1* C-7
- saving
 - cluster configurations *Vol 2* 26-1
- SCD
 - removing *Vol 2* 24-14
- scripts
 - activating verbose mode *Vol 2* 29-12
 - application start/stop scripts *Vol 1* 3-8, C-2
 - clinfo.rc *Vol 1* 16-2, 17-3
 - cluster messages *Vol 2* 29-1
 - emsvcsctrl *Vol 2* F-1
 - grpsvcctrl *Vol 2* F-6
 - HACMP/ES *Vol 2* 19-6
 - HACMP/ES shutdown *Vol 2* 19-6
 - HACMP/ES startup *Vol 2* 19-6
 - making print queues highly available *Vol 1* 18-45
 - sample custom scripts *Vol 1* 18-44
 - setting debug levels *Vol 2* 28-5
 - tailoring clinfo.rc *Vol 1* 16-2, 17-3
 - topsvcsctrl *Vol 2* F-24
 - verbose output *Vol 2* 29-1
- SCSI devices
 - installing shared disks *Vol 1* 11-9
 - planning considerations *Vol 1* 5-2
 - planning shared disks *Vol 1* 5-8
- scsidiskutil utility *Vol 2* E-5
- security
 - DCE authentication *Vol 1* 18-37
 - enhanced with Kerberos *Vol 1* 18-32
 - PSSP enhanced security *Vol 1* 18-36
- Serial Network Adapter Worksheet *Vol 1* A-9
- serial networks
 - configuring *Vol 1* 11-2
 - defining TMSSA to HACMP *Vol 1* 11-19
 - planning considerations *Vol 1* 4-8
 - testing target mode SCSI *Vol 1* 11-17
- Serial Networks Worksheet *Vol 1* A-7
- server
 - type of resource monitor *Vol 2* 31-3
- service adapter
 - definition *Vol 1* 4-9
- setting *Vol 1* 4-18
 - I/O Pacing *Vol 1* 4-18
 - syncd frequency rate *Vol 1* 18-7
- setting up
 - Clinfo files and scripts *Vol 1* 16-1
- shared
 - physical volume
 - maintaining *Vol 2* 22-42
 - volume groups
 - NFS issues *Vol 1* 6-10
- shared disk devices
 - versatile storage server (VSS) *Vol 1* 5-5
 - VSS *Vol 1* 1-5

- shared disks
 - 7135 RAIDiant Disk Array *Vol 1* 1-3, 5-3
 - 7137 Disk Arrays *Vol 1* 1-4, 5-5
 - planning *Vol 1* 5-1
 - SCSI-2 *Vol 1* 5-8
 - SSA disk subsystems *Vol 1* 1-4, 5-5
- shared file systems
 - changing
 - using AIX commands *Vol 2* 22-24
 - using C-SPOC *Vol 2* 22-23
 - creating *Vol 1* 12-4, *Vol 2* 22-18
 - maintaining *Vol 2* 22-18
 - removing
 - using AIX commands *Vol 2* 22-26
 - using C-SPOC *Vol 2* 22-26
- Shared IBM SCSI Disk Array Worksheet *Vol 1* A-13
- Shared IBM SSA Disk Subsystem Worksheet *Vol 1* A-15
- shared logical volume
 - adding copy *Vol 2* 22-40
 - creating *Vol 2* 22-29
 - increasing size *Vol 2* 22-39
 - maintaining *Vol 2* 22-29
 - removing copy *Vol 2* 22-40
 - renaming *Vol 2* 22-39
- shared LVM components
 - concurrent access *Vol 1* 12-7
 - defining *Vol 1* 12-1
 - file systems *Vol 1* 6-3
 - logical volumes *Vol 1* 6-3
 - maintaining *Vol 2* 22-1
 - maintaining in concurrent access environment *Vol 2* 23-1
 - maintaining with C-SPOC *Vol 2* 22-2
 - planning *Vol 1* 6-1
- shared memory key
 - use by RMAPI *Vol 2* 31-5
- shared physical volume
 - migrating data *Vol 2* 22-45
- Shared SCSI-2 Differential or Differential Fast/Wide Disks Worksheet *Vol 1* A-11
- Shared Volume Group/File System Worksheet
 - non-concurrent access *Vol 1* A-19, A-25
- shared volume groups
 - creating *Vol 1* 12-4 *Vol 2* 22-5
 - creating with TaskGuide *Vol 1* 12-1 *Vol 2* 23-2
 - deleting *Vol 2* 22-11
 - extending *Vol 2* 22-7
 - maintaining *Vol 2* 22-4
 - NFS issues *Vol 2* 28-1
 - reducing *Vol 2* 22-9
- showing
 - cluster definitions *Vol 2* 24-3
 - current characteristics of concurrent logical volume C-SPOC *Vol 2* 23-27
 - topology and group services configuration *Vol 1* 18-5
- shutdown modes
 - stopping cluster services *Vol 2* 20-12, 20-13
- single points of failure
 - potential cluster components *Vol 1* 2-2
- smit clstop fastpath *Vol 2* 27-4
- smit install_commit utility *Vol 1* 15-1
- smit install_remove utility *Vol 1* 14-12
- SMIT interface
 - for installing HACMP/ES *Vol 1* 10-1
- SNA network *see also* CS/AIX
 - configuring as cluster resource *Vol 1* 18-14
 - planning for CS/AIX communications links *Vol 1* 3-16
- snapshot *see* cluster snapshot
- SNAPSHOTPATH environment variable *Vol 2* 26-8
- snmpd daemon *Vol 2* 20-2
- snmpd.conf file
 - editing *Vol 2* 20-15
- sockets
 - component of Event Management *Vol 2* 31-4
 - component of Group Services *Vol 2* 30-3
 - names used by Event Management *Vol 2* 31-4
- software licenses *Vol 1* 5-7
- software maintenance
 - planning *Vol 2* H-9
- source node
 - creating shared volume groups *Vol 1* 12-4
- SP
 - IP address takeover *Vol 1* 4-9
 - serial networks *Vol 1* 4-5
- SP Ethernet
 - network types *Vol 1* 4-5
- SP Switch *Vol 2* E-8
 - adapters *Vol 1* 4-12
 - base IP address *Vol 1* 4-11
 - checking configuration *Vol 1* 11-2
 - configuring IP address takeover *Vol 1* 4-12
 - enabling ARP *Vol 1* 13-4
 - Eprimary management *Vol 1* 4-11
 - handling outages *Vol 1* 4-5
 - IP address takeover *Vol 1* 4-9
 - network *Vol 1* 4-5
 - specifying AIX error notification *Vol 1* 13-6
- SP Utilities *Vol 2* E-8
- specified operating environment *Vol 1* 10-3
- splitlvcopy utility *Vol 2* H-20
- SPMI (System Performance Monitor Interface)
 - dependency by Event Management *Vol 2* 31-7
- SPMI library
 - shared memory key *Vol 2* 31-5
- SRC (System Resource Controller)
 - and Event Manager daemon *Vol 2* 31-9
 - and Group Services daemon *Vol 2* 30-7
 - dependency by Group Services *Vol 2* 30-5

- SSA
 - adapter configuration *Vol 1* 5-14
 - concurrent volume groups *Vol 1* 5-17, 6-17
 - configuring for HA *Vol 1* 5-15
 - disk fencing *Vol 1* 5-17
 - disk subsystems *Vol 1* 1-4, 5-5
 - IBM documentation *Vol 1* 5-13
- SSA Disk Subsystems
 - in concurrent access environment *Vol 2* 23-1
 - planning shared disks *Vol 1* 5-13
 - verifying installation *Vol 1* 11-20
- SSA disks
 - procedure for replacing *Vol 2* H-17
- SSA domains *Vol 1* 5-16
- SSA Fiber Optic Extenders
 - configuring for HA *Vol 1* 5-15
- SSA Multi-Initiator RAID adapters *Vol 1* 5-14
- ssa_clear utility *Vol 2* E-6
- ssa_clear_all utility *Vol 2* E-7
- ssa_configure utility *Vol 2* E-7
- ssa_fence utility *Vol 2* E-6
- standby adapter
 - definition *Vol 1* 4-9
- start script
 - application servers *Vol 1* 3-8
 - recommendations *Vol 1* C-2
- starting
 - cluster services *Vol 2* 20-3
 - using C-SPOC *Vol 2* 20-4, 20-10
 - cluster services on clients *Vol 2* 20-14
 - NetView/HAVView *Vol 2* 21-4
 - network daemons *Vol 2* 20-10
- starting daemons
 - proper procedure *Vol 2* H-10
- starting subsystems
 - Event Management (emsvcsctrl) *Vol 2* 31-9
- status, Event Management
 - output of lssrc command *Vol 2* 31-12, *Vol 2* 31-13
- status, Group Services
 - output of lssrc command *Vol 2* 30-8
- sticky migration *Vol 2* 24-31
 - removing sticky markers *Vol 2* 24-40
- stop location keyword *Vol 2* 24-33
- stop script
 - application servers *Vol 1* 3-8
 - recommendations *Vol 1* C-2
- stopping
 - cluster service on clients *Vol 2* 20-14
 - cluster services *Vol 2* 20-5
 - on single node *Vol 2* 20-11
 - shutdown modes *Vol 2* 20-12
 - using C-SPOC *Vol 2* 20-6, 20-12
- stopping daemons
 - proper procedure *Vol 2* H-10
- stopping subsystems
 - Event Management (emsvcsctrl) *Vol 2* 31-9
- stopping the cluster
 - tasks requiring *Vol 2* H-9
- stopsrc command *Vol 2* 20-12
 - stopping Clinfo on clients *Vol 2* 20-14
 - stopping cluster services *Vol 2* 20-12
- subnets
 - in Tivoli-monitored clusters *Vol 1* 4-22
- subscriber
 - definition *Vol 2* 30-1
- subsystem
 - Event Management *Vol 2* 31-1
 - Group Services *Vol 2* 30-1
- subsystem control scripts
 - emsvcsctrl *Vol 2* F-1
 - grpsvcsctrl *Vol 2* F-6
 - topsvcsctrl *Vol 2* F-24
- subsystem status
 - for Event Management *Vol 2* 31-12, 31-13
 - for Group Services *Vol 2* 30-8
- supported hardware *Vol 1* 10-3
- symbolic links
 - changes in HACMP/ES *Vol 1* 14-3
- syncd
 - setting frequency for flushing buffers *Vol 1* 4-18
 - setting frequency rate *Vol 1* 18-7
- synchronizing
 - cluster resources *Vol 1* 18-30 *Vol 2* 24-40
 - cluster topology *Vol 1* 18-8 *Vol 2* 24-13
 - cluster topology configuration *Vol 2* 24-13
 - concurrent LVM mirrors
 - C-SPOC *Vol 2* 23-23, 23-27
 - prevented by SCD lock *Vol 2* 24-14
 - shared LVM mirrors
 - using C-SPOC *Vol 2* 22-41
 - shared volume group definition *Vol 2* 22-18
 - shared volume groups
 - C-SPOC *Vol 2* 22-17
 - skipping cluster verification during *Vol 2* 24-42
- sysback utility *Vol 2* H-20
- system error log file *Vol 2* 21-39
 - customizing output *Vol 2* 29-14
 - message formats *Vol 2* 29-12
 - recommended use *Vol 2* 29-3
 - understanding its contents *Vol 2* 29-12
- System Performance Monitor Interface (SPMI)
 - dependency by Event Management *Vol 2* 31-7
- System Resource Controlle(SRC)
 - functions *Vol 2* 20-3
- System Resource Controller (SRC)
 - and Event Manager daemon *Vol 2* 31-9
 - and Group Services daemon *Vol 2* 30-7
 - dependency by Group Services *Vol 2* 30-5
- System Resource Controller(SRC)
 - and clstart script *Vol 2* 19-6

T

- takeover
 - NFS issues *Vol 1* 6-11
- target mode SCSI
 - configuring serial network *Vol 1* 11-15
 - defining adapters to HACMP/ES *Vol 1* 18-2
 - enabling interface *Vol 1* 11-16
 - testing connections *Vol 1* 11-17
- target mode SSA
 - configuring connections *Vol 1* 11-18
 - defining serial network to HACMP/ES *Vol 1* 11-19
 - device files *Vol 1* 11-19
 - enabling interfaces *Vol 1* 11-18
 - testing connections *Vol 1* 11-19
- TaskGuide for creating shared volume groups
 - Vol 1* 12-1 *Vol 2* 23-2
- TCP/IP Network Adapter Worksheet *Vol 1* A-5
- TCP/IP Networks Worksheet *Vol 1* A-3
- TCP/IP services
 - proper procedure for stopping *Vol 2* H-10
- testing
 - SCSI target mode connection *Vol 1* 11-17
 - target mode SSA connection *Vol 1* 11-19
- testing loops
 - SSA disk subsystem *Vol 1* 5-16
- time limits
 - Event Management
 - connection to Group Services *Vol 2* 31-10
 - observation intervals *Vol 2* 31-12
 - peer group joining *Vol 2* 31-10
 - reconnection to resource monitors *Vol 2* 31-11
 - Group Services
 - connection to Topology Services *Vol 2* 30-6
- Tivoli, cluster monitoring
 - defining managed nodes *Vol 1* D-4
 - deinstalling *Vol 1* D-8, *Vol 2* 21-23
 - installation instructions *Vol 1* D-1
 - IPAT considerations *Vol 1* D-6
 - overview of installation process *Vol 1* D-2
 - polling intervals *Vol 2* 21-23
 - prerequisites *Vol 2* 21-15
 - required Tivoli software *Vol 1* 14-2, D-4
 - subnet requirements *Vol 1* 4-22, D-2
 - using *Vol 2* 21-14
- topology
 - clverify cluster option *Vol 2* 25-4
 - synchronizing *Vol 1* 18-8
 - verifying *Vol 1* 18-39
- topology hardware
 - replacing *Vol 2* H-14
- topology messages *Vol 2* G-48
- Topology Services
 - components *Vol 2* 32-2
 - configuring and operating *Vol 2* 32-5
 - daemon *Vol 2* 32-2
 - daemons for different roles *Vol 2* 32-3
 - displaying status *Vol 2* 32-8
 - files and directories *Vol 2* 32-4
 - initializing *Vol 2* 32-7
 - intra-cluster port numbers *Vol 2* 32-4
 - introduction *Vol 2* 32-1
 - port numbers and sockets *Vol 2* 32-3
 - required setting of IP source routing *Vol 2* F-23
 - services on which dependent *Vol 2* 32-5
 - tuning *Vol 2* 24-11
 - viewing configuration *Vol 1* 18-5
- Topology Services subsystem
 - and Group Services daemon initialization *Vol 2* 30-6
 - control script *Vol 2* F-24
 - dependency by Group Services *Vol 2* 30-5
- topsvcs command *Vol 2* F-23
- topsvcs startup script *Vol 2* 32-2
- topsvcsctrl script *Vol 2* F-24
- topsvcsd daemon *Vol 2* 20-2
- trace output log
 - Event Management *Vol 2* 31-6
 - Group Services *Vol 2* 30-5
- trace report
 - generating *Vol 2* 29-21
- tracing HACMP/ES daemons
 - disabling using SMIT *Vol 2* 29-19
 - generating a trace report using SMIT *Vol 2* 29-21
 - sample trace report *Vol 2* 29-23
 - specifying a trace report format *Vol 2* 29-19
 - specifying a trace report output file *Vol 2* 29-21
 - specifying content of trace report *Vol 2* 29-21
 - starting a trace session using SMIT *Vol 2* 29-20
 - stopping a trace session using SMIT *Vol 2* 29-20
 - trace IDs *Vol 2* 29-20
 - using cldiag *Vol 2* 29-22
 - using SMIT *Vol 2* 29-18
- tracing subsystems
 - Event Management (emsvcsctrl) *Vol 2* 31-9
 - Group Services (grpsvcsctrl) *Vol 2* 30-6
- trap
 - Clinfo *Vol 2* 20-15
- troubleshooting
 - Group Services subsystem
 - abnormal termination core file *Vol 2* 30-5
 - solving common problems *Vol 2* 29-24
- tuning
 - failure detection rate *Vol 2* 24-12
 - I/O pacing *Vol 1* 13-1
 - network module *Vol 2* 24-12
- tuning parameters *Vol 1* 4-17
 - cluster performance *Vol 1* 18-6
- tuning the cluster *Vol 1* 4-17

U

- UDP port
 - use by Event Management *Vol 2 31-4*
 - use by Group Services *Vol 2 30-3*
- Unix domain socket
 - Event Management client communication *Vol 2 31-2*
 - Group Services client communication *Vol 2 30-3*
 - resource monitor communication *Vol 2 31-2*
 - Topology Services use *Vol 2 32-4*
 - use by Event Management *Vol 2 31-4*
 - use by Group Services *Vol 2 30-3*
- unmirroring
 - concurrent volume groups
 - C-SPOC *Vol 2 23-22*
 - shared volume group
 - C-SPOC *Vol 2 22-16*
- updating
 - ARP cache
 - non-clinfo clients *Vol 1 17-4*
- updating LVM components *Vol 2 22-3*
- upgrading
 - HACMP/ES cluster *Vol 1 15-1*
- upgrading cluster
 - maintaining concurrent access *Vol 2 23-15*
- user accounts
 - adding *Vol 2 27-2*
 - changing attributes *Vol 2 27-4*
 - listing *Vol 2 27-1*
 - managing with C-SPOC *Vol 2 27-1*
 - removing *Vol 2 27-5*
- user defined application monitoring *Vol 1 18-21*
- usr/sbin/cluster/events/utills/convaryonvg command *Vol 2 23-10*

utilities

- cl_activate_fs *Vol 2 E-9*
- cl_activate_nfs *Vol 2 E-9*
- cl_activate_vgs *Vol 2 E-10*
- cl_deactivate_fs *Vol 2 E-10*
- cl_deactivate_nfs *Vol 2 E-10*
- cl_deactivate_vgs *Vol 2 E-11*
- cl_disk_available *Vol 2 E-1*
- cl_echo *Vol 2 E-13*
- cl_export_fs *Vol 2 E-11*
- cl_fs2disk *Vol 2 E-2*
- cl_getdisk_vg_fs_pvids *Vol 2 E-2*
- cl_is_array *Vol 2 E-3*
- cl_is_scsidisk *Vol 2 E-3*
- cl_log *Vol 2 E-13*
- cl_nfskill *Vol 2 E-12*
- cl_nm_nis_off *Vol 2 E-14*
- cl_nm_nis_on *Vol 2 E-14*
- cl_opsconfig *Vol 1 B-10*
- cl_raid_vg *Vol 2 E-3*
- cl_scdiskreset *Vol 2 E-4*
- cl_scsidiskrsrv *Vol 2 E-4*
- cl_swap_HPS_IP_address *Vol 2 E-8*
- cl_swap_HW_address *Vol 2 E-14*
- cl_swap_IP_address *Vol 2 E-15*
- cl_sync_vgs *Vol 2 E-5*
- cl_unswap_HW_address *Vol 2 E-16*
- cldiag *Vol 2 29-6*
- clverify *Vol 1 18-37*
- event emulation *Vol 2 21-32*
- scripts *Vol 2 E-1*
- scsidiskutil *Vol 2 E-5*
- smit install_commit *Vol 1 15-1*
- smit install_remove *Vol 1 14-12*
- ssa_clear *Vol 2 E-6*
- ssa_clear_all *Vol 2 E-7*
- ssa_configure *Vol 2 E-7*
- ssa_fence *Vol 2 E-6*

V

- var/ha/log/grpqlsm *Vol 2 29-5*
- varyonvg command
 - in concurrent access mode *Vol 2 23-2*
- verbose script output
 - activating *Vol 2 29-12*
- verification
 - errors ignored during synchronization *Vol 1 18-8*
Vol 2 24-13, 24-41, 26-7

verifying
 cluster configuration *Vol 2* 25-1
 cluster configuration using SMIT *Vol 2* 25-6
 cluster environment *Vol 1* 18-37
 cluster topology *Vol 1* 18-39
 install of HACMP/ES software *Vol 1* 14-11
 networks and resources *Vol 2* 25-6
 node environment *Vol 1* 18-37
 owned resources *Vol 2* 25-5

Versatile Storage Server (VSS)
 and HACMP *Vol 1* 5-12
 concurrent access volume group *Vol 2* 23-6
 creating concurrent access volume group
 Vol 1 12-9
 features *Vol 1* 5-5
 overview *Vol 1* 1-5
 planning configuration *Vol 1* 6-1

viewing
 cluster.log file *Vol 2* 29-6
 cspoc.log file *Vol 2* 29-17
 details about cluster
 NetView *Vol 2* 21-10
 topology and group services settings *Vol 1* 18-5

volume groups
 activating
 concurrent access mode *Vol 2* 23-9
 as shared LVM component *Vol 1* 6-3
 changing startup status *Vol 1* 12-6
 checking access mode *Vol 2* 23-10
 concurrent access
 maintaining *Vol 2* 23-2
 creating concurrent capable *Vol 2* 23-3
 creating for concurrent access *Vol 1* 12-7
 dormant at startup *Vol 1* 12-12 *Vol 2* 23-9
 importing *Vol 1* 12-6
 planning issues *Vol 2* H-8
 quorum *Vol 1* 6-5
 shared
 creating *Vol 2* 22-5
 creating with TaskGuide *Vol 1* 12-1,
 Vol 2 23-2
 deleting with AIX commands *Vol 2* 22-11
 extending *Vol 2* 22-7
 extending with C-SPOC *Vol 2* 22-13
 importing with C-SPOC *Vol 2* 22-13
 maintaining *Vol 2* 22-4
 mirroring with C-SPOC *Vol 2* 22-15
 reducing with AIX commands *Vol 2* 22-9
 removing with C-SPOC *Vol 2* 22-14
 unmirroring with C-SPOC *Vol 2* 22-16
 synchronizing definition with C-SPOC
 Vol 2 22-18
 synchronizing mirrors with C-SPOC *Vol 2* 22-17
 vary on in concurrent mode *Vol 2* 23-9

VSM (Visual System Management)
 overview *Vol 2* I-1

VSS *see* Versatile Storage Server

W

worksheets *Vol 1* A-1
 AIX Fast Connect *Vol 1* A-31
 Application Server Worksheet *Vol 1* A-37
 Application Worksheet *Vol 1* A-27
 Cluster Event Worksheet *Vol 1* A-45
 CS/AIX Communications Links *Vol 1* A-35
 NFS-Exported Filesystem Worksheet *Vol 1* 6-16,
 A-21
 Non-Shared Volume Group Worksheet
 Vol 1 A-17, A-23
 online planning worksheet program
 installing *Vol 1* B-2
 overview *Vol 1* B-1
 Resource Group Worksheet *Vol 1* A-43
 Serial Network Adapter Worksheet *Vol 1* A-9
 Serial Networks Worksheet *Vol 1* A-7
 Shared IBM SCSI Disk Array Worksheet
 Vol 1 A-13
 Shared IBM SSA Disk Subsystem Worksheet
 Vol 1 A-15
 Shared SCSI-2 Differential or Differential
 Fast/Wide Disks *Vol 1* A-11
 Shared Volume Group/File System Worksheet
 concurrent access *Vol 1* A-25
 non-concurrent access *Vol 1* A-19
 TCP/IP Network Adapter Worksheet *Vol 1* A-5
 TCP/IP Networks Worksheet *Vol 1* A-3

XYZ

X Window System display
 using clstat *Vol 2* 21-27

xhacmpm application
 overview *Vol 2* I-1

ypbind *Vol 1* 4-16

ypserv daemon *Vol 1* 4-16

Vos remarques sur ce document / Technical publication remark form

Titre / Title : Bull HACMP 4.4 Enhanced Scalability Installation and Administration Guide
Volume 1/2

N° Référence / Reference N° : 86 A2 62KX 02

Daté / Dated : August 2000

ERREURS DETECTEES / ERRORS IN PUBLICATION

AMELIORATIONS SUGGEREES / SUGGESTIONS FOR IMPROVEMENT TO PUBLICATION

Vos remarques et suggestions seront examinées attentivement.

Si vous désirez une réponse écrite, veuillez indiquer ci-après votre adresse postale complète.

Your comments will be promptly investigated by qualified technical personnel and action will be taken as required.

If you require a written reply, please furnish your complete mailing address below.

NOM / NAME : _____ Date : _____

SOCIETE / COMPANY : _____

ADRESSE / ADDRESS : _____

Remettez cet imprimé à un responsable BULL ou envoyez-le directement à :

Please give this technical publication remark form to your BULL representative or mail to:

**BULL CEDOC
357 AVENUE PATTON
B.P.20845
49008 ANGERS CEDEX 01
FRANCE**

Technical Publications Ordering Form

Bon de Commande de Documents Techniques

To order additional publications, please fill up a copy of this form and send it via mail to:

Pour commander des documents techniques, remplissez une copie de ce formulaire et envoyez-la à :

BULL CEDOC
ATTN / MME DUMOULIN
357 AVENUE PATTON
B.P.20845
49008 ANGERS CEDEX 01
FRANCE

Managers / Gestionnaires :
Mrs. / Mme : C. DUMOULIN +33 (0) 2 41 73 76 65
Mr. / M : L. CHERUBIN +33 (0) 2 41 73 63 96
FAX : +33 (0) 2 41 73 60 19
E-Mail / Courrier Electronique : srv.Cedoc@franp.bull.fr

Or visit our web site at: / Ou visitez notre site web à:

<http://www-frec.bull.com> (PUBLICATIONS, Technical Literature, Ordering Form)

CEDOC Reference # N° Référence CEDOC	Qty Qté	CEDOC Reference # N° Référence CEDOC	Qty Qté	CEDOC Reference # N° Référence CEDOC	Qty Qté
___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]	
___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]	
___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]	
___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]	
___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]	
___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]	
___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]		___ - ___ - ___ - ___ [___]	

[___] : no revision number means latest revision / pas de numéro de révision signifie révision la plus récente

NOM / NAME : _____ Date : _____

SOCIETE / COMPANY : _____

ADRESSE / ADDRESS : _____

PHONE / TELEPHONE : _____ FAX : _____

E-MAIL : _____

For Bull Subsidiaries / Pour les Filiales Bull :

Identification: _____

For Bull Affiliated Customers / Pour les Clients Affiliés Bull :

Customer Code / Code Client : _____

For Bull Internal Customers / Pour les Clients Internes Bull :

Budgetary Section / Section Budgétaire : _____

For Others / Pour les Autres :

Please ask your Bull representative. / Merci de demander à votre contact Bull.

**BULL CEDOC
357 AVENUE PATTON
B.P.20845
49008 ANGERS CEDEX 01
FRANCE**

**ORDER REFERENCE
86 A2 62KX 02**

PLACE BAR CODE IN LOWER
LEFT CORNER



Utiliser les marques de découpe pour obtenir les étiquettes.
Use the cut marks to get the labels.

AIX
HACMP 4.4
Enhanced Scalability
Install. & Admin.
Guide
Volume 1/2
86 A2 62KX 02

AIX
HACMP 4.4
Enhanced Scalability
Install. & Admin.
Guide
Volume 1/2
86 A2 62KX 02

AIX
HACMP 4.4
Enhanced Scalability
Install. & Admin.
Guide
Volume 1/2
86 A2 62KX 02