Bull

HACMP 4.4
Enhanced Scalability Installation and Administration Guide

Volume 2/2

AIX

ORDER REFERENCE
86 A2 89KX 01

Bull

HACMP 4.4
Enhanced Scalability Installation and Administration Guide

Volume 2/2

AIX

Software

August 2000

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

ORDER REFERENCE
86 A2 89KX 01

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.

The information in this document is subject to change without notice. Groupe Bull will not be liable for errors contained herein, or for incidental or consequential damages in connection with the use of this material.

Contents

Chapter 21 Monitoring an HACMP/ES Cluster 21-1

Chapter 22 Maintaining Shared LVM Components 22-1

Chapter 23 Maintaining Shared LVM Components in a Concurrent Access Environment 23-1

Chapter 27 Managing Users and Groups in a Cluster 27-1

Chapter 28 Additional Tasks: NFS and Run-Time Parameters 28-1

Chapter 29 Troubleshooting HACMP/ES Clusters 29-1

Part 5 Volume 2 Appendixes

About This Guide

This is Volume 2 of the *High Availability Cluster Multi-Processing for AIX*: *Enhanced Scalability Installation and Administration Guide*. This volume provides cluster administration and troubleshooting information, as well as RSCT-specific information.

A master index covering topics in Volume 1 and Volume 2 is included at the end of each volume.

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.

How to Use This Guide

Overview of Contents

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.
- 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, discusses important points to keep in mind when maintaining a cluster on a 7x24 basis.
- Appendix I, VSM Graphical Configuration Application, describes the AIX Visual System Management program.

Highlighting

The following highlighting conventions are used in this guide:

Italic	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/es/lpp/cluster/doc/release_notes** describe hardware and software requirements
- *HACMP for AIX, Version 4.4: Concepts and Facilities,* order number 86 A2 54KX 02
- *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

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.

Ordering Publications

To order additional copies of this guide, use Order Number 86 A2 89KX 01.

Part 4 Administering an HACMP/ES Cluster

This part of the book contains instructions for general administrative tasks involved in maintaining a running cluster.

Chapter 19 Maintaining an HACMP/ES Cluster

This chapter 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.

Note: The directory **/usr/sbin/cluster** and subdirectories have symbolic links to the **/usr/es/sbin /cluster** directory and subdirectories. However, files in these directories are *not* linked as they were in releases prior to 4.3.1.

Maintenance Tasks

The following maintenance tasks for an HACMP/ES system are described in detail in subsequent chapters:

- Starting and stopping cluster services
- Monitoring a cluster
- Maintaining shared LVM components
- Changing the cluster configuration
- Verifying the cluster configuration
- Saving and restoring a cluster configuration
- Managing users and groups in a cluster
- Additional tasks related to NFS and run-time parameters

Starting and Stopping Cluster Services

Various methods for starting and stopping cluster services are available. Chapter 20, Starting and Stopping Cluster Services, describes how to start and stop HACMP/ES on server nodes and client nodes with and without using the C-SPOC utility.

Monitoring the Cluster

By design, failures of components in the cluster are handled automatically. But you need to be aware of all such events. Chapter 21, Monitoring an HACMP/ES Cluster, describes various tools you can use to check the status of an HACMP/ES cluster, the nodes, networks, and resource groups within that cluster, and the daemons that run on the nodes.

Although the combination of HACMP/ES and the inherent high availability features built into the AIX system keeps single points of failure to a minimum, there are still failures that, although detected, can cause other problems. See Chapter 13, Tailoring AIX for HACMP/ES, for suggestions on customizing error notification.

Maintaining Shared Logical Volume Manager Components

Any changes to logical volume components must be synchronized across all nodes in the cluster. Chapter 22, Maintaining Shared LVM Components, and Chapter 23, Maintaining Shared LVM Components in a Concurrent Access Environment, describe a set of procedures for maintaining cluster LVM components with or without using the C-SPOC utility.

Changing the Cluster Topology, Security Mode, or Resources

Any changes to cluster topology, security mode, or resources require updating the cluster across all nodes. Chapter 24, Changing the Cluster Configuration, describes how to modify cluster topology, cluster security mode, or cluster resources after the initial configuration.

To change the cluster topology, you must stop and restart cluster services to make the changed configuration the currently active configuration.

You can change the resource configuration or migrate resources to other nodes dynamically, using the **cldare** command from the command line or through SMIT.

Verifying the Cluster Configuration

Verifying the cluster configuration assures you 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. You should verify the cluster configuration after making changes to a cluster or node. Chapter 25, Verifying a Cluster Configuration, describes how to use the **clverify** utility.

Saving and Restoring Cluster Configurations

After you configure the topology and resources of a cluster, you can save the cluster configuration. This saved configuration can later be used to restore the configuration. A cluster snapshot can also be applied to an active cluster to dynamically reconfigure the cluster. Chapter 26, Saving and Restoring Cluster Configurations, describes how to use the cluster snapshot utility.

Managing Users and Groups in a Cluster

Using C-SPOC, you can create, change, or remove users and groups from all cluster nodes by executing a C-SPOC command on any single cluster node. Chapter 27, Managing Users and Groups in a Cluster, describes these tasks.

Additional HACMP/ES Maintenance Tasks

Additional tasks that you must perform to maintain an HACMP/ES system include changing the run-time parameters for a node and configuring NFS filesystems. Chapter 28, Additional Tasks: NFS and Run-Time Parameters, describes these tasks.

Related Administrative Tasks

The tasks below, while not specifically discussed in this book, are essential for effective system administration.

Backing Up Your System

The practice of allocating multiple copies of a logical volume can enhance high availability in a cluster environment, but it should not be considered a replacement for regular system backups. Although HACMP/ES is designed to survive failures within the cluster, it cannot survive a catastrophic failure where multiple points of failure leave data on disks unavailable. Therefore, it is imperative that you back up your system on a regular basis. You must have a backup procedure in place to ensure data reliability and to protect against catastrophic physical volume failure.

To maintain your HACMP/ES environment, you must back up the root volume group (which contains the HACMP/ES software) and the shared volume groups (which contain the data for highly available applications) regularly. HACMP/ES is like other AIX environments from this perspective. Back up all nodes.

Documenting Your System

As your HACMP/ES system grows and changes, it differs from its initial cluster configuration. It is your responsibility as system administrator to document all aspects of the HACMP/ES system unique to your environment. This responsibility includes documenting procedures concerning the highly available applications, recording changes that you make to the configuration scripts distributed with HACMP/ES, documenting any custom scripts you write, recording the status of backups, maintaining a log of user problems, and maintaining records of all hardware.

Maintaining Highly Available Applications

As system administrator, you should understand the relationship between your applications and HACMP/ES. Highly available applications are started and stopped by HACMP/ES in response to cluster events. Understanding when, how, and why this happens is critical to keeping a cluster highly available, as problems can occur that require corrective actions.

For a discussion of strategies for making your applications highly available, see the planning chapters and the appendix on Applications and HACMP in Volume 1 of this guide.

Helping Users

As the resident HACMP/ES expert, you can expect to receive many questions from end users at your site about HACMP/ES. The more you know about HACMP/ES, the better you are able to answer these questions. If you cannot answer questions about your HACMP/ES cluster environment, contact your IBM support representative.

AIX Files Modified by HACMP/ES

The following AIX files are modified to support HACMP/ES. They are not distributed with HACMP/ES.

/.rhosts

If you are using Standard security mode: On each node in the cluster, list the service and boot adapters for each node in the cluster. This allows the Global ODM (**godm**) daemon to run so you can configure the cluster nodes from a central location. It also allows the **clruncmd** to run.

If you are using Enhanced security (Kerberos): Do *not* make these **/.rhosts** entries; remove them if you change from Standard to Enhanced security mode.

/etc/hosts

The cluster event scripts use the **/etc/hosts** file for name resolution. All cluster node IP interfaces must be added to this file on each node. Note that DNS and NIS are disabled during HACMP/ES-related name resolution. This is why HACMP/ES IP addresses must be maintained locally.

/etc/inetd.conf

The **/etc/inetd.conf** file is the Internet server configuration file. The install process adds the following entry for the HACMP/ES Global ODM socket (**godm**) to this file:

```
godm  stream  tcp  nowait  root  /usr/es/sbin/cluster/godmd
```

/etc/inittab

The **/etc/inittab** file is modified when HACMP/ES is configured for IP address takeover or the **Start at system restart** option is chosen on the SMIT Start Cluster Services screen.

IP Address Takeover

The following entry is added to the **/etc/inittab** file for HACMP/ES network startup with IP address takeover.

```
harc:2:wait:/usr/es/sbin/cluster/etc/harc.net # HACMP network startup
```

When IP address takeover is enabled, the system edits **/etc/inittab** to change the **rc.tcpip** and **inet**-dependent entries from run level “2” (the default multi-user level) to run level “a”. Entries that have run level “a” are processed only when the **telinit** command is executed specifying that specific run level.

System Boot

The **/etc/inittab** file is used by the **init** process to control the startup of processes at boot time. The following entry is added to the **/etc/inittab** file if the **Start at system restart** option is chosen on the SMIT Start Cluster Services screen:

```
hacmp:2:wait:/usr/sbin/etc/rc.cluster -boot> /dev/console 2>&1 # Bring
up Cluster
```

When the system boots, the **/etc/inittab** file calls the **/usr/es/sbin/cluster/etc/rc.cluster** script to start HACMP/ES.

Because the **inet** daemons must not be started until after HACMP/ES-controlled adapters have swapped to their service address, HACMP/ES also adds the following entry to the end of the **/etc/inittab** file to indicate that **/etc/inittab** processing has completed.

```
clinit:a:wait:touch /usr/es/sbin/cluster/.telinit # HACMP This must be
last entry in inittab!
```

See Chapter 20, Starting and Stopping Cluster Services, for more information about the files involved in starting and stopping HACMP/ES.

/etc/rc.net

The **/etc/rc.net** file is called by **cfgmgr** to configure and start TCP/IP during the boot process. It sets hostname, default gateway and static routes. The following entry is added at the beginning of the file for a node on which IP address takeover is enabled:

```
# HACMP for AIX
# HACMP for AIX These lines added by HACMP for AIX software
[ "$1" = "-boot" ] && shift || { ifconfig lo0 127.0.0.1 up; exit 0; } #HACMP for AIX
# HACMP for AIX
```

The entry prevents **cfgmgr** from reconfiguring boot and service addresses while HACMP/ES is running.

/etc/services

The **/etc/services** file defines the sockets and protocols used for network services on a system. The ports and protocols used by the HACMP/ES components are defined here.

```
#clinfo_deadman               6176/tcp
#clm_keepalive                6255/udp
#cllockd                      6100/udp
#clm_pts                      6200/tcp
#clsmuxpd                     6270/tcp
#clm_lkm                      6150/tcp
#clm_smux                     6175/tcp
#godm                         6177/tcp
#topsvcs                      6178/udp
#grpsvcs                      6179/udp
#emsvcs                       6180/udp
```

/etc/snmpd.conf

The SNMP daemon reads the **/etc/snmpd.conf** configuration file when it starts up and when a refresh or **kill -1** signal is issued. This file specifies the community names and associated access privileges and views, hosts for trap notification, logging attributes, **snmpd**-specific parameter configurations, and SMUX configurations for the **snmpd**. The HACMP/ES installation process adds the **clsmuxpd** password to this file. The following entry is added to the end of the file, to include the HACMP/ES MIB managed by the **clsmuxpd**.

```
smux   1.3.6.1.4.1.2.3.1.2.1.5 "clsmuxpd_password" # HACMP clsmuxpd
```

/etc/snmpd.peers

The **/etc/snmpd.peers** file configures **snmpd** SMUX peers. The HACMP/ES install process adds the following entry to include the **clsmuxpd**.

```
clsmuxpd 1.3.6.1.4.1.2.3.1.2.1.5 "clsmuxpd_password" # HACMP clsmuxpd
```

/etc/syslog.conf

The **/etc/syslog.conf** file is used to control output of the **syslogd** daemon, which logs system messages. During the install, process HACMP/ES adds entries to this file that direct the output from HACMP/ES-related problems to certain files.

```
# example:
# "mail messages, at debug or higher, go to Log file. File must exist."
# "all facilities, at debug and higher, go to console"
# "all facilities, at crit or higher, go to all users"
#  mail.debug           /usr/spool/mqueue/syslog
#  *.debug              /dev/console
#  *.crit                       *
# HACMP/ES Critical Messages from HACMP/ES
local0.crit /dev/console
# HACMP/ES Informational Messages from HACMP/ES
local0.info /usr/es/adm/cluster.log
# HACMP/ES Messages from Cluster Scripts
user.notice /usr/es/adm/cluster.log
```

The **/etc/syslog.conf** file should be identical on all cluster nodes.

/etc/tcp.clean

The **/etc/tcp.clean** file stops TCP/IP daemons, brings down network interfaces, and removes lock files. During the install, HACMP/ES adds the **clstop** command to this file. The **clstop** command is run when the AIX shutdown command is issued to shut the AIX operating system down.

/etc/trcfmt

The **/etc/trcfmt** file is the template file for the system trace logging and report utility, **trcrpt**. The install process adds HACMP/ES tracing to the trace format file. HACMP/ES tracing applies to the following daemons: **clstrmgr**, **clinfo**, and **clsmuxpd**.

/var/spool/cron/crontab/root

The **/var/spool/cron/crontab/root** file contains commands needed for basic system control. The install process adds HACMP/ES logfile rotation to the file.

Scripts

The following scripts are supplied with the HACMP/ES software.

Startup and Shutdown Scripts

Each of the following scripts is involved in starting and stopping the HACMP/ES software. See Chapter 20, Starting and Stopping Cluster Services, for more information about these scripts.

/usr/es/sbin/cluster/utilities/clstart

The **/usr/es/sbin/cluster/utilities/clstart** script, which is called by the **/usr/es/sbin/cluster/utilities/rc.cluster** script, invokes the AIX System Resource Controller (SRC) facility to start the cluster daemons. The **clstart** script starts HACMP/ES with the options currently specified on the Start Cluster Services screen. See the **clstart** man page for additional information.

There is a corresponding C-SPOC version of this script that starts cluster services on each cluster node. The **/usr/es/sbin/cluster/sbin/cl_clstart** script calls the HACMP/ES **clstart** script. See the **cl_clstart** man page for additional information.

/usr/es/sbin/cluster/utilities/clstop

The **/usr/es/sbin/cluster/utilities/clstop** script, which is called by the SMIT Stop Cluster Services screen, invokes the SRC facility to stop the cluster daemons with the options specified on the Stop Cluster Services screen. See the **clstop** man page for additional information.

There is a corresponding C-SPOC version of this script that stops cluster services on each cluster node. The **/usr/es/sbin/cluster/sbin/cl_clstop** script calls the HACMP/ES **clstop** script. See the **cl_clstop** man page for additional information.

/usr/es/sbin/cluster/utilities/clexit.rc

If the SRC detects that the **clstrmgr** daemon has exited abnormally, it executes the **/usr/es/sbin/cluster/utilities/clexit.rc** script to halt the system. If the SRC detects that any other HACMP/ES daemon has exited abnormally, it executes the **clexit.rc** script to stop these processes, but does not halt the node. See the **clexit.rc** man page for additional information.

/usr/es/sbin/cluster/etc/rc.cluster

If the **Start at system restart** option is chosen on the Start Cluster Services screen, the **/usr/es/sbin/cluster/etc/rc.cluster** script is called by the **/etc/inittab** file to start HACMP/ES. The **/usr/es/sbin/cluster/etc/rc.cluster** script does some necessary initialization and then calls the **usr/es/sbin/cluster/utilities/clstart** script to start HACMP/ES.

The **/usr/es/sbin/cluster/etc/rc.cluster** script is also used to start the **clinfo** daemon on a client. See the **rc.cluster** man page for additional information.

There is a corresponding C-SPOC version of this script that starts cluster services on each cluster node. The **/usr/es/sbin/cluster/sbin/cl_rc.cluster** script calls the HACMP/ES **rc.cluster** script. See the **cl_rc.cluster** man page for additional information.

/etc/rc.net

The **/etc/rc.net** script is called by the **/usr/es/sbin/cluster/etc/rc.cluster** script to configure and start the TCP/IP interfaces and to set the required network options. The **/etc/rc.net** script is used in the boot process to retrieve interface information from the ODM and to configure all defined interfaces. If IP address takeover is configured, the **/etc/rc.net** script is called from the **/usr/es/sbin/cluster/etc/rc.cluster** script at cluster startup instead of during the boot process.

Event Scripts

The event scripts described below are called by the cluster daemons to respond to cluster events. See Chapter 8, Cluster Events: Tailoring and Creating, and Chapter 24, Changing the Cluster Configuration, for more information about these scripts. See the man pages for additional information.

The /usr/es/sbin/cluster/events/node_up Scripts

The **/usr/es/sbin/cluster/events/node_up** scripts handle node attachment.

- **/usr/es/sbin/cluster/events/node_up_local**
- **/usr/es/sbin/cluster/events/node_up_remote**
- **/usr/es/sbin/cluster/events/node_up_local_complete**
- **/usr/es/sbin/cluster/events/node_up_remote_complete**

The /usr/es/sbin/cluster/events/node_down Scripts

The **/usr/es/sbin/cluster/events/node_down** scripts handle node detachment.

- **/usr/es/sbin/cluster/events/node_down_local**
- **/usr/es/sbin/cluster/events/node_down_remote**
- **/usr/es/sbin/cluster/events/node_down_local_complete**
- **/usr/es/sbin/cluster/events/node_down_remote_complete**

The /usr/es/sbin/cluster/events/swap_adapter Script

The **/usr/es/sbin/cluster/events/swap_adapter** script swaps the service and standby adapters.

The /usr/es/sbin/cluster/events/network Scripts

The **/usr/es/sbin/cluster/events/network_up** and **/usr/es/sbin/cluster/events/network_down** scripts handle network failure and recovery. These scripts take no default actions; they must be customized for the site.

The /usr/es/sbin/cluster/events/reconfig Scripts

These scripts handle dynamic reconfiguration activities.

- **/usr/es/sbin/cluster/events/reconfig_resource_acquire**
- **/usr/es/sbin/cluster/events/reconfig_resource_release**
- **/usr/es/sbin/cluster/events/reconfig_resource_complete**

The /usr/es/sbin/cluster/events/server Scripts

These scripts handle application server start and stop activities and application failure notification:

- **/usr/es/sbin/cluster/events/server_down**
- **/usr/es/sbin/cluster/events/server_down_complete**
- **/usr/es/sbin/cluster/events/server_restart**
- **/usr/es/sbin/cluster/events/server_restart_complete**

Two other scripts in the same category of “application server events” are named with a slightly different convention:

- **/usr/es/sbin/cluster/events/start_server**
- **/usr/es/sbin/cluster/events/stop_server**

The /usr/es/sbin/cluster/events/rg Scripts

These scripts handle resource group movement to support application monitor activity:

- **/usr/es/sbin/cluster/events/rg_move**
- **/usr/es/sbin/cluster/events/rg_move_complete**

The /usr/es/sbin/cluster/etc/clinfo.rc Script

The **/usr/es/sbin/cluster/etc/clinfo.rc** script, which is invoked by the **clinfo** daemon whenever a network or node event occurs, updates the system’s ARP cache. You can customize this script for additional processing. There must be a copy of the **/usr/es/sbin/cluster/etc/clinfo.rc** script on each node in the cluster. See the **clinfo.rc** man page for additional information.

Chapter 20 Starting and Stopping Cluster Services

This chapter explains how to start and stop cluster services on cluster nodes and clients. The chapter also describes how to use the Cluster-Single Point of Control (C-SPOC) utility to start and stop cluster services in cluster environments.

Overview

Starting cluster services refers to the process of starting the HACMP/ES daemons that enable the coordination required between nodes in a cluster. Starting cluster services on a node also triggers the execution of certain HACMP/ES scripts that create the cluster. Stopping cluster services refers to stopping these same daemons on a node and may or may not cause the execution of additional HACMP/ES scripts, depending on the type of shutdown you perform.

Before starting or stopping cluster services, you must have a thorough understanding of the node interactions it causes and the impact on your system's availability. The following sections define the HACMP/ES cluster services and briefly describe the processing involved in startup or shutdown of these services. Later sections in this chapter describe the step-by-step procedures you must follow to start or stop cluster services on a node. These sections include information about starting or stopping cluster services on multiple nodes in clusters using the C-SPOC utility. The final section of this chapter describes how to start and stop cluster services on HACMP/ES clients.

Note: The directory **/usr/es/sbin/cluster** and subdirectories have symbolic links to the **/usr/sbin /cluster** directory and subdirectories. However, files in these directories are *not* linked as they were in releases prior to 4.3.1.

Cluster Services

Note: If you list the daemons in the AIX System Resource Controller (SRC), you will see ES appended to their names. The actual executables do not have the ES appended; the process table shows the executable by path (**/usr/es/sbin/cluster...**).

The following lists the required and optional HACMP/ES daemons:

Cluster Manager daemon (clstrmgr)

This daemon monitors the status of the nodes and their interfaces, and invokes the appropriate scripts in response to node or network events. It also centralizes the storage of and publishes updated information about HACMP-defined resource groups. The Cluster Manager on each node coordinates information gathered from the HACMP global ODM, and other Cluster Managers in the cluster to maintain updated information about the content, location, and status of all HACMP resource groups. This information is updated and synchronized among all nodes whenever an event occurs that affects resource group configuration, status, or location.

All cluster nodes must run the **clstrmgr** daemon.

Cluster SMUX Peer daemon (clsmuxpd)

This daemon maintains status information about cluster objects. This daemon works in conjunction with the Simple Network Management Protocol (**snmpd**) daemon. All cluster nodes must run the **clsmuxpd** daemon.

Note: The **clsmuxpd** daemon cannot be started unless the **snmpd** daemon is running.

Cluster Information Program daemon (clinfo)

This daemon provides status information about the cluster to cluster nodes and clients and invokes the **/usr/es/sbin/cluster/etc/clinfo.rc** script in response to a cluster event. The **clinfo** daemon is optional on cluster nodes and clients.

Cluster Lock Manager daemon (cllockd)

This daemon provides advisory locking services. The **cllockd** daemon is required on cluster nodes only if those nodes are part of a concurrent access configuration.

Cluster Topology Services daemon (topsvcsd)

This daemon monitors the status of network adapters in the cluster. All cluster nodes must run the **topsvcsd** daemon.

Cluster Event Management daemon (emsvcsd)

This daemon matches information about the state of system resources with information about resource conditions of interest to client programs (applications, subsystems, and other programs).The **emsvcsd** daemon runs on each node of a domain.

Event Management AIX Operating System Resource Monitor (emaixos)

This daemon acts as a resource monitor for the event management subsystem and provides information about the operating system characteristics and utilization. The **emaixos** daemon is started automatically by Event Management

Cluster Group Services daemon (grpsvcsd)
This daemon manages all of the distributed protocols required for cluster operation. All cluster nodes must run the **grpsvcsd** daemon.

Cluster Globalized Server Daemon daemon (grpglsmd)
This daemon operates as a **grpsvcs** client; its function is to make switch adapter membership global across all cluster nodes. All cluster nodes must run the **grpglsmd** daemon.

The AIX System Resource Controller (SRC) controls the HACMP/ES daemons (except for **cllockd**, which is a kernel extension). It provides a consistent interface for starting, stopping, and monitoring processes by grouping sets of related programs into subsystems and groups. In addition, it provides facilities for logging of abnormal terminations of subsystems or groups and for tracing of one or more subsystems. For information about tracing HACMP/ES subsystems, see the section on tracing in Chapter 29, Troubleshooting HACMP/ES Clusters.

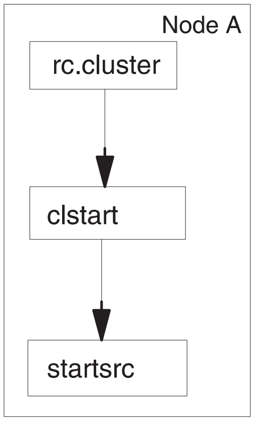
The HACMP/ES daemons are collected into the following SRC subsystems and groups:

Daemon	Subsystem	Group
/usr/es/sbin/cluster/clstrmgr	clstrmgrES	cluster
/usr/es/sbin/cluster/clinfo	clinfoES	cluster
/usr/es/sbin/cluster/clsmuxpd	clsmuxpdES	cluster
/usr/es/sbin/cluster/cllockd	cllockdES	lock
/usr/sbin/rsct/bin/emsvcs	emsvcs	emsvcs
/usr/sbin/rsct/bin/topsvcs	topsvcs	topsvcs
/usr/sbin/rsct/bin/hagsd	grpsvcs	grpsvcs
/usr/sbin/rsct/bin/hagsglsmd	grpglsm	grpsvcs

When using the SRC commands, you can control the **clstrmgr**, **clinfo**, and **clsmuxpd** daemons by specifying the SRC **cluster** group.

Understanding Cluster Service Startup

You start cluster services on a node by executing the HACMP/ES **/usr/es/sbin/cluster/etc/rc.cluster** script. Use the Start Cluster Services SMIT screen, described in the section Starting Cluster Services on page 20-8, to build and execute this command. The **rc.cluster** script initializes the environment required for HACMP/ES by setting environment variables and then calls the **/usr/es/sbin/cluster/utilities/clstart** script to start the HACMP/ES daemons. The **clstart** script is the HACMP/ES script that starts all the cluster services. The **clstart** script calls the SRC **startsrc** command to start the specified subsystem or group. The following figure illustrates the major commands and scripts called at cluster startup.



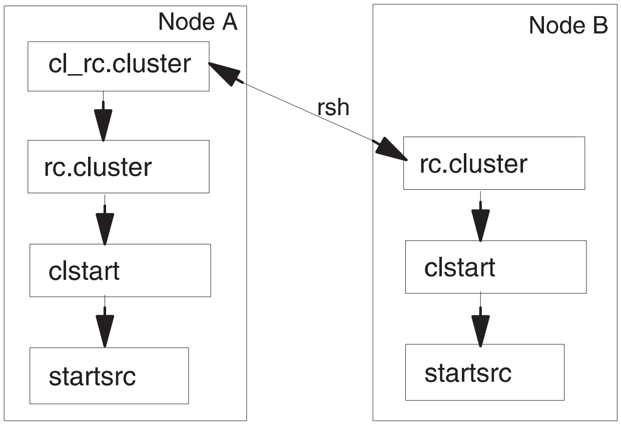
Flow of Commands Used in Cluster Startup

The HACMP/ES daemons are started in the following order:

- RSCT daemons (Group Services, Topology Services, then Event Management)
- Cluster Manager
- Cluster SMUX daemon
- Cluster Information Program daemon (optional)

Using the C-SPOC utility, you can start cluster services on any node (or on all nodes) in a cluster by executing the C-SPOC **/usr/es/sbin/cluster/sbin/cl_rc.cluster** command on a single cluster node. The C-SPOC **cl_rc.cluster** command calls the **rc.cluster** command to start cluster services on the nodes specified from the one node. The nodes are started in sequential order - not in parallel. The output of the command run on the remote node is returned to the originating node. Because the command is executed remotely, there can be a delay before the command output is returned.

The following example shows the major commands and scripts executed on all cluster nodes when cluster services are started in clusters using the C-SPOC utility.



Flow of Commands Used at Cluster Startup by C-SPOC Utility

Automatically Restarting Cluster Services

You can optionally have cluster services start whenever the system is rebooted. If you specify the **-R** flag to the **rc.cluster** command, or specify "restart or both" in the Start Cluster Services SMIT screen, the **rc.cluster** script adds the following line to the **/etc/inittab** file.

```
hacmp:2:wait:/usr/es/sbin/cluster/etc/rc.cluster -boot> /dev/console
2>&1
# Bring up Cluster
```

At system boot, this entry causes AIX to execute the **/usr/es/sbin/cluster/etc/rc.cluster** script to start HACMP/ES.

Note: Be aware that if the cluster services are set to restart automatically at boot time, you may face problems with node integration after a power failure and restoration, or you may want to test a node after doing maintenance work before having it rejoin the cluster.

Starting Cluster Services with IP Address Takeover Enabled

If IP address takeover is enabled, the **/usr/es/sbin/cluster/etc/rc.cluster** script calls the **/etc/rc.net** script to configure and start the TCP/IP interfaces and to set the required network options.

Editing the rc.cluster File to Turn Deadman Switch Off

In HACMP/ES, the Deadman Switch (DMS) is controlled by RSCT Topology Services. If, in a rare case, you want to turn the DMS off, you must edit the **rc.cluster** file as follows:

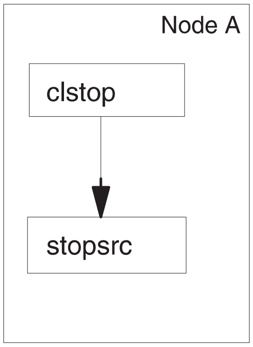
- There is a -D flag in clstart, located in **/usr/es/sbin/cluster/utilities**
- In the **/usr/es/sbin/cluster/etc/rc.cluster** file, find a call to "clstart" at about line #486. Edit this call to include the -D flag.

For more information about the Deadman Switch, see *Chapter 4, Planning Cluster Network Connectivity.*

Understanding Stopping Cluster Services

You stop cluster services on a node by executing the HACMP/ES **/usr/es/sbin/cluster/utilities/clstop** script. Use the HACMP for AIX Stop Cluster Services SMIT screen, described in the section Stopping Cluster Services on page 20-11 to build and execute this command. The **clstop** script stops an HACMP/ES daemon or daemons. The **clstop** script starts all the cluster services or individual cluster services by calling the SRC command **stopsrc**.

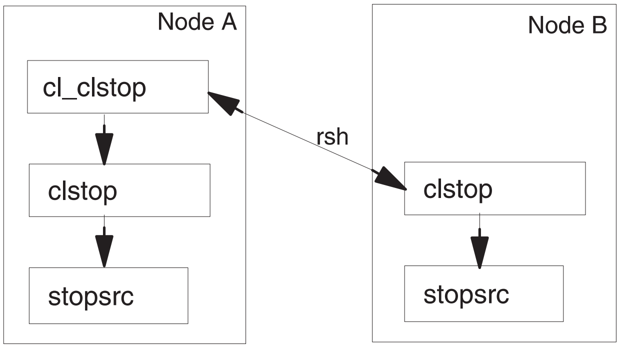
The following figure illustrates the major commands and scripts called at cluster shutdown.



Flow of Commands Used at Cluster Shutdown

Using the C-SPOC utility, you can stop cluster services on a single node or on all nodes in a cluster by executing the C-SPOC **/usr/es/sbin/cluster/sbin/cl_clstop** command on a single node. The C-SPOC **cl_clstop** command performs some cluster-wide verification and then calls the **clstop** command to stop cluster services on the specified nodes. The nodes are stopped in sequential order—not in parallel. The output of the command run on the remote node is returned to the originating node. Because the command is executed remotely, there can be a delay before the command output is returned.

The following figure provides an illustration of how commands and scripts are executed when cluster services are stopped on a single-node in a cluster using the C-SPOC utility. All nodes in the cluster are stopped accordingly, if they are specified from the node originating the command.



Flow of Commands Used at Cluster Shutdown

When to Stop Cluster services

You typically stop cluster services:

- Before making any hardware or software changes or other scheduled node shutdowns or reboots. Failing to do so may cause unintended cluster events to be triggered on other nodes.

- Before certain reconfiguration activity. Some changes to the cluster information stored in the ODM require stopping and restarting the cluster services on *all* nodes for the changes to become active. For example, if you wish to change the name of the cluster, the name of a node, or the name of an adapter, you must stop and restart the cluster.

  For more information about which changes to the cluster require reconfiguration, see Chapter 24, Changing the Cluster Configuration.

Types of Stops

When you stop cluster services, you must also decide how to handle the resources that were owned by the node you are removing from the cluster. You have the following options:

Graceful In a *graceful* stop, the HACMP/ES software shuts down its applications and releases its resources. The other nodes *do not* take over the resources of the stopped node.

Graceful with Takeover In a *graceful with takeover* stop, the HACMP/ES software shuts down its applications and releases its resources. The surviving nodes take over these resources. This is also called *intentional fallover*.

Note: When the AIX operating system is shut down (with the **shutdown** command), HACMP/ES is stopped gracefully, without takeover.

The forced down option is not supported in HACMP/ES.

Abnormal Termination of a Cluster Daemon

If the SRC detects that any HACMP/ES daemon has exited abnormally (without being shut down using the **clstop** command), it executes the **/usr/es/sbin/cluster/utilities/clexit.rc** script to halt the system. This prevents unpredictable behavior from corrupting the data on the shared disks. See the **clexit.rc** man page for additional information.

The **clexit.rc** script creates an AIX error log entry. Here is an example showing the long output:

```
LABEL:          OPMSG
IDENTIFIER:     AA8AB241

Date/Time:       Fri Jan  7 10:44:46
Sequence Number: 626
Machine Id:      000001331000
Node Id:         ppstest8
Class:           O
Type:            TEMP
Resource Name:   OPERATOR

Description
OPERATOR NOTIFICATION

User Causes
ERRLOGGER COMMAND

        Recommended Actions
        REVIEW DETAILED DATA

Detail Data
MESSAGE FROM ERRLOGGER COMMAND
clexit.rc : Unexpected termination of clstrmgrES
```

The **clexit.rc** error message in short form looks like this:

```
AA8AB241   0107104400 T O OPERATOR       OPERATOR NOTIFICATION
```

Warning: Never use the **kill -9** command on the **clstrmgr** daemon. Using the **kill** command causes the **clstrmgr** daemon to exit abnormally. This causes the SRC to run the script **/usr/es/sbin/cluster/utilities/clexit.rc**, which halts the system immediately and causes the surviving nodes to initiate fallover.

You can modify the file **/etc/cluster/hacmp.term** to change the default action after an abnormal exit. The **clexit.rc** script checks for the presence of this file, and, if executable, will call it instead of halting the system automatically.

Starting Cluster Services

The following sections describe how to start cluster services on a single node or on multiple nodes in a cluster by executing a C-SPOC command.

Note: If a node has a **tty** console, it must be powered on for HACMP/ES to be started on that node. If this is not desirable, find the line in the **/usr/es/sbin/cluster/etc/rc.cluster** script that ends with `2>/dev/console` and change this line to reflect whatever behavior is desired. For example, you can redirect output to another **tty**. Be aware that this redirects the startup messages on the node.

Starting Cluster Services on a Single Node

The steps below describe the procedure for starting cluster services on a cluster node. If you have defined cascading or rotating resources, start the nodes in the logical order for the correct distribution of those resources.

Note: All cluster nodes should be running. For cold boots, let one node fully boot before powering up the other nodes.

To start cluster services on a single node, perform the following procedure:

1. Type `smit clstart`.

   The Start Cluster Services screen appears.

2. Enter field values as follows:

   Start now, on system restart or both

   Indicate whether you want the **clstrmgr** and **clsmuxpd** daemons to start when you commit the values on this screen (now), when the operating system reboots (restart), or on both occasions. Choosing on system restart or both means that the cluster services are always brought up automatically after a system reboot.

BROADCAST message at startup?
Indicate whether you want to send a broadcast message to users when the cluster services start.

Startup Cluster Lock Services?
Set to **false** unless you have a concurrent access configuration or if your application uses the Cluster Lock Manager; in this case set to **true**.

Startup Cluster Information Daemon?
Indicate whether you want to start the **clinfo** daemon. If your application uses Clinfo or if you use the **clstat** monitor, set this field to **true**. Otherwise, set it to **false**.

The value that you enter in the **Startup Cluster Information Services?** field works in conjunction with the value you enter in the **Start now, on system restart or both** field. If you set the startup field to **true** and the **Start now, on system restart or both** field to **both**, then this daemon is also started whenever the **clstrmgr** and **clsmuxpd** daemons are started.

3. Press Enter. The system starts the cluster services, activating the cluster configuration that you have defined, if you specified that you wanted cluster services started immediately.

   After a few seconds, a message appears on the console indicating that the **/usr/es/sbin/cluster/events/node_up** script has begun. The time that it takes this script to run depends on your configuration (that is, the number of disks, the number of interfaces to configure, the number of file systems to mount, and the number of applications being started). When the **node_up** script completes, a message appears on the console indicating that the **/usr/es/sbin/cluster/events/node_up_complete** is running.

   When the scripts complete, HACMP/ES is up and running on the first node. SMIT displays a command status window.

   If cluster services fail to start, check the **cluster.log** or **hacmp.out** log files for error messages.

4. Ensure that the HACMP-controlled applications are available.

   Access the applications controlled by HACMP/ES to verify that they behave as expected.

Repeat this procedure on the other cluster nodes. To verify that the nodes are up, use the **/usr/es/sbin/cluster/clstat** or HAView utility, described in Chapter 21, Monitoring an HACMP/ES Cluster.

Reintegrating Nodes

A node reintegrating into a cluster is just like a node starting up except that the Cluster Manager is already running on other cluster nodes. The cluster's fallback configuration determines whether the node joining the cluster takes back shared resources other nodes took over during fallover. If the resources are defined as rotating or cascading without fallback, the node joining the cluster (the original owner) does not take back the shared resources.

Ensuring that Network Daemons Start as Part of HACMP/ES Start-Up

Clusters that have both rotating and cascading resource groups, where the cascading resource group does not have IP address takeover configured (that is, the cascading resources are not tied to a service label), can experience problems starting the network daemons. At cluster startup, the first node to join the cluster acquires the rotating address and runs the **telinit -a** command. The second node to join the cluster, since it does not migrate from a boot to a service address, does not run the **telinit -a** command. As a result, the network-related daemons (NFS and NIS for example) do not get started. If your cluster requires this combination of resource groups, customize the pre- and post-event processing to issue the **telinit -a** command.

Starting Cluster Services in C-SPOC Clusters

The steps below describe the procedure for starting cluster services on a single node or on all nodes in a cluster by executing the C-SPOC **/usr/es/sbin/cluster/sbin/cl_rc.cluster** command on one of the cluster nodes.

Note: All cluster nodes should be running. For cold boots, let one node fully boot before powering up the other nodes.

As the root user, perform the following steps to start the cluster services on a node.

1. Enter:

```
smit cl_admin
```

2. Select the **HACMP for AIX Cluster Services** option and press Enter.
3. Select the **Start Cluster Services** option and press Enter. SMIT displays the following screen:
4. Enter field values as follows:

Start now, on system restart or both	Indicate whether you want the **clstrmgr, clresmgr,** and **clsmuxpd** daemons to start when you commit the values on this screen (now), when the operating system reboots (restart), or on both occasions. Choosing on system restart or both means that the cluster services are always brought up automatically after a system reboot.
BROADCAST message at startup?	Indicate whether you want to send a broadcast message to users when the cluster services start. The messages appear on all the nodes being started.
Startup Cluster Lock Services?	If you have a concurrent access configuration or an application that uses the Lock Manager, set this field to **true.** Otherwise, set it to **false.**

Startup Cluster Information Daemon?

Indicate whether you want to start the **clinfo** daemon. If your application uses Clinfo or if you use the clstat monitor, set this field to **true.** Otherwise, set it to **false**.

The value that you enter in the **Startup Cluster Information Services?** field works in conjunction with the value you enter in the **Start now, on system restart or both** field. If you set either (or both) of the startup fields to **true** and the **Start now, on system restart or both** field to **both**, then the **clinfo** daemon is also started whenever the **clstrmgr** and **clsmuxpd** daemons are started.

5. Press Enter. The system starts the cluster services on the nodes specified, activating the cluster configuration that you have defined. The time that it takes the commands and scripts to run depends on your configuration (that is, the number of disks, the number of interfaces to configure, the number of file systems to mount, and the number of applications being started).

   When command execution completes and HACMP/ES cluster services are started on all nodes specified, SMIT displays a command status window similar to the following:

   If cluster services fail to start on any cluster node, check the C-SPOC utility log file, named **/tmp/cspoc.log**, for error messages. This file contains the command execution status of the C-SPOC command executed on each cluster node.

Stopping Cluster Services

The following sections describe how to stop cluster services on a single cluster node or multiple nodes in a cluster of up to eight nodes using the C-SPOC utility.

Note: Minimize activity on the system. If the node you are stopping is currently providing highly available services, notify users of your intentions if their applications will be unavailable. Let them know when services will be restored.

Stopping Cluster Services on a Single Cluster Node

The steps below describe the procedure for stopping HACMP/ES on a node. If all nodes need to be stopped, follow the same steps on all nodes, but stop them sequentially, not in parallel.

To stop cluster services on a single cluster node, perform the following procedure.

1. Type `smit clstop`.

   The Stop Cluster Services screen appears.

2. Enter field values as follows:

Stop now, on system restart or both	Indicate whether you want the cluster services to stop now, at restart (when the operating system reboots), or on both occasions. If you choose restart or both, the cluster services will no longer come up automatically after a reboot.
BROADCAST cluster shutdown?	Indicate whether you want to send a broadcast message to users before the cluster services stop.
Shutdown mode	Indicate the type of shutdown: - **graceful**—Shutdown after the **/usr/es/sbin/cluster/events/node_down_complete** script is run on this node to release its resources. The other nodes *do not* take over the resources of the stopped node. - **graceful with takeover**—Shutdown after the **/usr/es/sbin/cluster/events/node_down_complete** script runs to release its resources. The other nodes do take over the resources of the stopped node.

3. Press Enter. HACMP/ES stops the cluster services according to your specifications.

   SMIT displays a command status window.

Note: When you stop the cluster and restart it immediately, there is a two-minute delay before the **clstrmgr** daemon becomes active.

AIX Shutdown and Cluster Services

When the AIX operating system is stopped with the shutdown command, HAMCP/ES is stopped gracefully, without takeover.

Using the stopsrc Command to Stop an HACMP/ES Daemon

You can also stop an HACMP/ES daemon by using the standard AIX **stopsrc** command with the **-g** flag, specifying the cluster group:

```
stopsrc -g cluster
```

Note: Using the **stopsrc** command to stop an individual HACMP/ES daemon (by specifying the **-s** flag with the name of the daemon) is not recommended and may cause unexpected behavior.

Stopping Cluster Services in a C-SPOC Cluster

The steps below describe the procedure for stopping cluster services on a single node or on all nodes in a cluster by executing the C-SPOC **/usr/es/sbin/cluster/sbin/cl_clstop** command on one of the cluster nodes. When stopping multiple nodes, C-SPOC stops them sequentially, not in parallel. If any node specified to be stopped is inactive, the shutdown operation aborts.

To stop cluster services in a C-SPOC cluster:

1. Enter:

   ```
   smit cl_admin
   ```

   Note: Do not confuse the C-SPOC **HACMP for AIX Cluster Services** SMIT screen with the **Cluster Services** SMIT screen available as an option on the main HACMP for AIX SMIT screen.

2. Select the **HACMP for AIX Cluster Services > Stop Cluster Services** options and press Enter.

3. Enter field values in the SMIT screen as follows:

   Stop now, on system restart or both

   Indicate whether you want the cluster services to stop now, at restart (when the operating system reboots), or on both occasions. If you choose restart or both, the entry in the **/etc/inittab** file that starts cluster services is removed. Cluster services will no longer come up automatically after a reboot.

   BROADCAST cluster shutdown?

   Indicate whether you want to send a broadcast message to users before the cluster services stop. If you specify true, a message is broadcast on each cluster node being stopped.

   Shutdown mode

   Indicate the type of shutdown:

   -graceful: Shutdown after the **/usr/es/sbin/cluster/events/node_down_complete** script is run on the node to release its resources. Other cluster nodes *do not* take over the resources of the stopped node.

   graceful with takeover: Shutdown after the **/usr/es/sbin/cluster/events/node_down_complete** script runs to release its resources. Other nodes take over the resources of the stopped node. You cannot shut down cluster services on other cluster nodes if you select this type of shutdown.

4. Press Enter. The system stops the cluster services on the nodes specified.

   SMIT displays a command status window.

If the stop operation fails, check the C-SPOC utility log file, named **/tmp/cspoc.log**, for error messages. This file contains the command execution status of the C-SPOC command executed on each cluster node.

Starting and Stopping Cluster Services on Clients

The cluster services on clients consist solely of the **clinfo** daemon, which provides clients with status information about the cluster. This daemon is under the control of the SRC. Although **clinfo** is under the control of the SRC, its abnormal exit does not cause a **halt -q** by default.

Note that the **/etc/inittab** file is modified when the HACMP/ES software is installed to start the **clinfo** daemon whenever the system is rebooted.

Starting Clinfo on a Client

Use the **/usr/es/sbin/cluster/etc/rc.cluster** script or the **startsrc** command to start **clinfo** on a client, as shown below:

```
/usr/es/sbin/cluster/etc/rc.cluster
```

You can also use the standard AIX **startsrc** command:

```
startsrc -s clinfoES
```

Stopping Clinfo on a Client

Use the standard AIX **stopsrc** command to stop **clinfo** on a client machine:

```
stopsrc -g cluster
```

Maintaining Cluster Information Services

In order for the **clinfo** daemon to get the information it needs, you must edit the **/usr/es/sbin/cluster/etc/clhosts** file. This file should contain IP service and boot labels or addresses of any HACMP/ES servers to which **clinfo** may communicate, including servers from any clusters accessible through logical connections. The installed version of the **clhosts** file on an HACMP/ES client node differs from that on an HACMP/ES server node.

Maintaining the clhosts File on a Client

As installed, the **clhosts** file on an HACMP/ES client node contains no hostnames or addresses. HACMP/ES server addresses must be provided by the user at installation time. This file should contain all boot and service names or addresses of HACMP/ES servers from any cluster accessible through logical connections with this client node. Upon startup, **clinfo** uses these names or addresses to attempt communication with a **clsmuxpd** process executing on an HACMP/ES server.

An example list of hostnames/addresses in a **clhosts** file follows:

```
n0_cl83        #  n0 service
n2_cl83        #  n2 service
n3_cl83        #  n3 service
```

Maintaining the clhosts File on a Cluster Node

As installed, the **clhosts** file on an HACMP/ES server node contains a loopback address. **clinfo** first attempts to communicate with a **clsmuxpd** process locally. If it succeeds, it then acquires an entire cluster map, including a list of all HACMP/ES server interface addresses. From then on, it uses this list rather than the provided loopback address to recover from **clsmuxpd** communication loss.

However, if **clinfo** does not succeed in communicating with a **clsmuxpd** process locally, it continues indefinitely to attempt communication periodically only to the local address. For this reason, you should add all HACMP/ES server boot and service addresses accessible through logical connections with this node, just as on an HACMP/ES client node. The loopback address is provided only as a convenience.

Enabling Clinfo for Asynchronous Event Notification

By default, **clinfo** polls the **clsmuxpd** periodically for information. You can optionally choose to have **clinfo** receive notification of events as asynchronous messages (traps).

Only one SNMP application can receive traps. If you are running NetView you cannot enable **clinfo** to receive traps.

To enable asynchronous event notification, use the following procedure.

1. Start **clinfo** with the **-a** option, by entering the following:

   ```
   chssys -s clinfoES -a "-a".
   ```
   To verify that the SRC has the correct command line arguments for clinfo, enter the following:

   ```
   lssrc -Ss clinfoES | awk -F: '{print $3}'
   ```
2. Edit the **snmpd.conf** file on the nodes that will send traps. As installed, traps are directed to the loopback address (**clinfo** receives those traps generated by the **clsmuxpd** on the same node).

   a. Find the trap line at the end of the file. It looks like this:

   ```
   view   1.17.2  system enterprises view
   trap   public  127.0.0.1       1.2.3   fe      # loopback
   ```
   b. Add trap lines as desired. Multiple **clinfo** processes can receive traps from a single **clsmuxpd** process. Make sure that the “1.2.3 fe” field is unique.

   An entry might look like the following example, with two more trap lines added:

   ```
   trap   public  127.0.0.1       1.2.3   fe      #loopback
   trap   public  123.456.789.1                   #adam
   trap   public  123.456.789.2                   #eve
   ```
   c. Stop and restart the **snmpd** and **clsmuxpd** processes on the hosts where you made the changes in the **snmpd.conf** file:

   ```
   stopsrc -s clsmuxpdES
   stopsrc -s snmpdES
   startsrc -s snmpdES
   startsrc -s clsmuxpdES
   ```

Chapter 21 Monitoring an HACMP/ES Cluster

This chapter describes tools you can use to monitor an HACMP/ES cluster.

Note: The default locations of log files are used in this chapter. If you redirected any logs, check the appropriate location.

Periodically Monitoring an HACMP/ES Cluster

By design, HACMP/ES compensates for various failures that occur within a cluster. For example, HACMP/ES compensates for a network adapter failure by swapping in a standby adapter. As a result, it is possible that a component in the cluster has failed and that you are unaware of the fact. The danger here is that, while HACMP/ES can survive one or possibly several failures, *a failure that escapes your notice threatens a cluster's ability to provide a highly available environment.*

To avoid this situation, you should customize your system by adding event notification to the scripts designated to handle the various cluster events. You can specify a command that sends you mail indicating that an event is about to happen (or that an event has just occurred), along with information about the success or failure of the event. The mail notification system enhances the standard event notification methods.

In addition, HACMP/ES offers application monitoring capability that you can configure and customize in order to monitor the health of specific applications and processes.

Use the AIX Error Notification facility to add an additional layer of high availability to an HACMP/ES environment. The combination of HACMP/ES and the inherent high availability features built into the AIX system keeps single points of failure to a minimum; the Error Notification facility can further enhance the availability of your particular environment. See Chapter 13, Tailoring AIX for HACMP/ES*,* for suggestions on customizing error notification.

See Chapter 8, Cluster Events: Tailoring and Creating, for detailed information on predefined events and on customizing event handling. Also, be sure to consult your worksheets, to document any changes you make to your system, and to periodically inspect the key cluster components to make sure they are in full working order.

Tools for Monitoring an HACMP/ES Cluster

Note: The directory **/usr/sbin/cluster** and subdirectories have symbolic links to the **/usr/es/sbin/cluster** directory and subdirectories. However, files in these directories are *not* linked as they were in releases prior to 4.3.1.

HACMP/ES provides the following tools for monitoring a cluster:

- The HAView utility extends NetView services so you can monitor HACMP/ES clusters and cluster components across a network from a single node. Using HAView, you can also view the full cluster event history in the **/usr/es/sbin/cluster/history/cluster.*mmdd*** file. The event history (and other cluster status and configuration information) is accessible through NetView's menu bar.
- Cluster Monitoring with Tivoli allows you to monitor clusters and cluster components through your Tivoli Framework interface.
- The **/usr/es/sbin/cluster/clstat** utility reports the status of key cluster components—the cluster itself, the nodes in the cluster, and the network adapters connected to the nodes.
- Application Monitoring allows you to monitor specific applications and processes and define action to take upon detection of process death or other application failures.
- The Event Emulator provides an emulation of cluster events.
- The SMIT Show Cluster Services screen shows the status of the HACMP/ES daemons
- Log files allow you to track cluster events and history: The **/usr/es/adm/cluster.log** file tracks cluster events; the **/tmp/hacmp.out** file records the output generated by configuration scripts as they execute; the **/usr/es/sbin/cluster/history/cluster.*mmdd*** log file logs the daily cluster history; the **/tmp/cspoc.log** file logs the status of C-SPOC commands executed on cluster nodes.

Monitoring a Cluster With HAView

HAView is a cluster monitoring utility that allows you to monitor HACMP/ES clusters using TME 10 NetView for AIX. Using NetView, you can monitor clusters and cluster components across a network from a single management station.

HAView creates and modifies NetView objects that represent clusters and cluster components. It also creates submaps that present information about the state of all nodes, networks, network interfaces, and resource groups associated with a particular cluster. This cluster status and configuration information is accessible through NetView's menu bar.

HAView monitors cluster status using the Simple Network Management Protocol (SNMP). It combines periodic polling and event notification through traps to retrieve cluster topology and state changes from the HACMP management agent, the Cluster SMUX peer daemon (**clsmuxpd**).

You can view cluster event history using the HACMP Event Browser and node event history using the Cluster Event Log. Both browsers can be accessed from the NetView menu bar. The **/usr/es/sbin/cluster/history/cluster.*mmdd*** file contains more specific event history. This information is helpful for diagnosing and troubleshooting fallover situations. See page 29-4 and page 29-15 for more information about this log file.

HAView Installation Considerations

HAView has a client/server architecture. You must install both an HAView server image and an HAView client image, on the same machine or on separate server and client machines. For more information on installation considerations, see Chapter 14, Installing the HACMP/ES Software, in Volume 1 of this guide. In particular, refer to the section Installation Server on page 14-4.

HAView File Modification Considerations

Certain files may need to be modified in order for HAView to monitor your cluster properly. When configuring HAView, you should check and/or edit the following files:

- **haview_start**
- **clhosts**
- **snmpd.conf**

The haview_start File

The **haview_start** file must be edited by the user so that it includes the name of the node that has the HAView server executable installed. This is how the HAView client knows where the HAView server is located. Regardless of whether the HAView server and client are on the same node or different nodes, you are required to specify the HAView server node in the **haview_start** file.

The **haview_start** file is loaded when the HAView client is installed and is stored in **/usr/haview**. Initially, the **haview_start** file contains only the following line:

```
"${HAVIEW_CLIENT:-/usr/haview/haview_client}" $SERVER
```

You must add the following line to the file:

```
SERVER="${SERVER:-<your server name>}"
```

For example, if the HAView server is installed on *mynode*, the edited **haview_start** file appears as follows:

```
SERVER="${SERVER:-mynode}"
"${HAVIEW_CLIENT:-/usr/haview/haview_client}" $SERVER
```

where *mynode* is the node that contains the HAView server executable.

The clhosts File

HAView monitors a cluster’s state within a network topology based on cluster-specific information in the **/usr/sbin/cluster/etc/clhosts** configuration file. The **clhosts** file must be present on the NetView management node. Make sure this file contains the IP address or IP label of the service and/or boot adapters of the nodes in each cluster that HAView is to monitor.

Note: Make sure the hostname and the service label of your NetView nodes are exactly the same. (If they are not the same, add an alias in the **/etc/hosts** file to resolve the name difference.)

Warning: If an invalid IP address exists in the **clhosts** file, HAView will fail to monitor the cluster. Make sure the IP addresses are valid, and there are no extraneous characters in the **clhosts** file.

The snmpd.conf File

The NetView management node must also be configured in the list of trap destinations in the **snmpd.conf** files on the cluster nodes of all clusters you want it to monitor. This makes it possible for HAView to utilize traps in order to reflect cluster state changes in the submap in a timely manner. Also, HAView can discover clusters not specified in the **clhosts** file on the nodes in another cluster.

The format for configuring trap destinations is as follows:

```
trap  <community name> <IP address of NetView management node>1.2.3 fe
```

For example, enter:

```
trap          public                   140.186.131.121 1.2.3 fe
```

Note the following:

- You can specify the name of the management node instead of the IP address.
- You can include multiple trap lines in the **snmpd.conf** file.

NetView Hostname Requirements for HAView

The following hostname requirements apply to using HAView in a NetView environment. If you change the hostname of an adapter, the NetView database daemons and the default map are affected.

Hostname Effect on the NetView Daemon

The hostname required to start NetView daemons must be associated with a valid interface name or else NetView fails to start.

Hostname Effect on the NetView Default Map

If you change the hostname of the NetView client, the new hostname does not match the original hostname referenced in the NetView default map database and NetView will not open the default map. Using the NetView **mapadmin** command, you need to update the default map (or an invalid map) to match the new hostname.

See the *NetView for AIX Administrator's Guide* for more information about updating or deleting an invalid NetView map.

Starting HAView

Once you've installed the HAView client and server, HAView is started and stopped when you start or stop NetView, but there are some conditions to verify on the management node before starting HAView.

Before starting NetView/HAView, check the management node as follows:

- Make sure both client and server components of HAView are installed. See Chapter 14, Installing the HACMP/ES Software, or Chapter 15, Upgrading an HACMP/ES Cluster, for more information.
- Make sure access control has been granted to remote nodes by running the **xhost** command with the plus sign (+) or with specified nodes:

  ```
  xhost +
  ```
  (to grant access to all computers)

  or, to grant access to specific nodes only:

  ```
  xhost <computers to be given access>
  ```
- Make sure the DISPLAY variable has been set to the monitoring node and to a label that can be resolved by and contacted from remote nodes:

  ```
  export DISPLAY=<monitoring node>:0.0
  ```

These actions allow you to access the HAView Cluster Administration option.

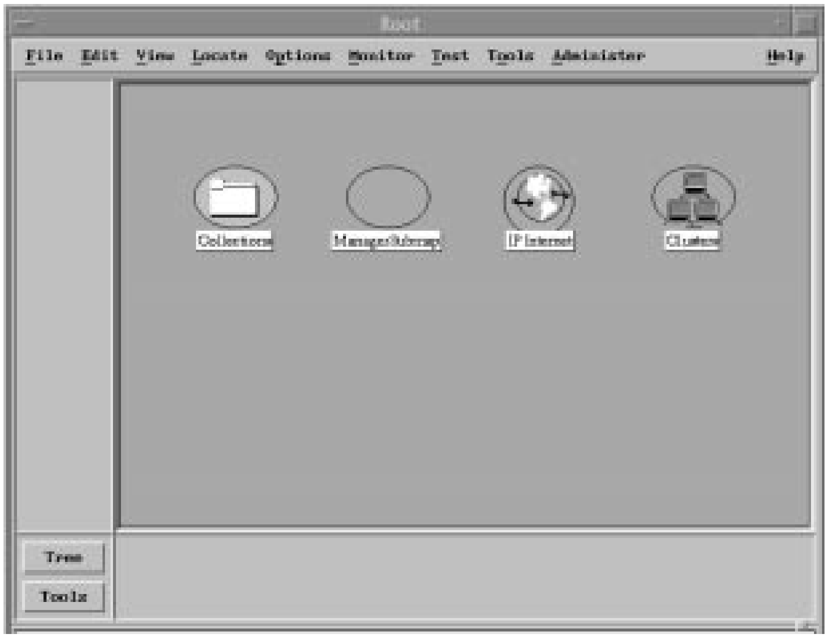
After ensuring these conditions are set, type the following to start NetView:

```
/usr/OV/bin/nv6000
```

Refer to the *NetView for AIX User's Guide for Beginners* for further instructions about starting NetView.

When NetView starts, HAView creates objects and symbols to represent a cluster and its components. Through submaps, you can view detailed information about these components.

HAView places the Clusters symbol on the NetView map after NetView starts. As shown in the following figure, the Clusters symbol is placed alongside the NetView Collections, Manager Submap, and IP Internet symbols.



HAView Clusters Symbol

Viewing Clusters and Components

To see which clusters HAView is currently monitoring, double-click the Clusters symbol. The Clusters submap appears. You may see one or more symbols that represent specific clusters. Each symbol is identified by a label indicating the cluster's name. Double-click a cluster symbol to display symbols for nodes, networks, and resource groups within that cluster.

Note that the cluster status symbol may remain unknown until the next polling cycle, even though the status of its cluster components is known. See HAView Polling Intervals on page 21-10 for more information about the default intervals and how to change them using SMIT.

Note: You can view component details at any time using the shortcut **ctrl-o**. See Obtaining Component Details in HAView on page 21-10 for information and instructions.

Read-Write and Read-Only NetView Maps

Normally, you have one master monitoring station for NetView/HAView. This station is supplied with new information as cluster events occur, and its map is updated so it always reflects the current cluster status.

In normal cluster monitoring operations, you will probably not need to open multiple NetView stations on the same node. If you do, and you want the additional stations to be updated with current cluster status information, you must be sure they use separate maps with different map names. For more information on multiple maps and changing map permissions, see the *NetView for AIX Administrators Guide*.

Interpreting Cluster Topology States

When using HAView to view cluster topology, symbols for clusters and cluster components such as nodes and networks are displayed in various colors depending on the object's state. The following table summarizes colors you may see when monitoring a cluster. (For information about the resource group symbol colors, see the Interpreting Resource Group Symbol Colors table on page 21-9.).

Status	**Meaning**	**Symbol Color**	**Connection Color (network submap)**
Critical	The object has failed or is not functioning. If the symbol is a node or network, the node or network is DOWN.	Red	Red
Normal	The object is functioning correctly. If the symbol is a node object, the node is UP.	Green	Black
Marginal	Some object functions are working correctly; others are not.	Yellow	Red
Unknown	The object's state cannot be determined. It may not be currently monitored by HAView.	Blue	Blue

Note: You can select **Legend** at any time from the Help pull-down menu to view NetView and HAView symbols and to understand their associative colors.

The Navigation Tree and Submap Windows

In addition to the submap window, the NetView Navigation Tree Window can help you keep track of your current location in the HAView hierarchy. Press the Tree button to see the Navigation Tree Window. In the Navigation Tree, the blue outline indicates where you are in the map, that is, which submap you are in.

The Symbols Legend

At any time, you can select **Legend** from the Help pull-down menu to view all NetView and HAView symbols and the meanings of the symbols and their various colors.

The Help Menu

To view help topics, select **Help > Index > Tasks > HAView Topics**.

Viewing Networks

To view the state of the nodes and addresses connected to a network associated with a specific cluster, double-click a network symbol in the specific Cluster submap. A network submap appears displaying symbols for all nodes connected to the network. The symbols appear in a color that indicates the nodes' current state. The vertical line representing a network is called the network connection. Its color indicates the status of the connection between the node and the network.

See Interpreting Cluster Topology States on page 21-6 in the next section for a table of symbol colors and how they reflect a cluster and its components' state.

Viewing Nodes

To view the state of nodes associated with a particular network, double-click a network symbol. A submap appears displaying all nodes connected to the network. Each symbol's color indicates the associated node's current state.

You can also view the state of any individual node associated with a cluster by double-clicking on that node's symbol in the specific cluster submap.

Viewing Addresses

To view the status of addresses serviced by a particular node, double-click a node symbol from either a cluster or network submap. A submap appears displaying symbols for all addresses configured on a node. Each symbol's color indicates the associated address's current state.

Note: When viewing adapters in a node submap from a network submap, all adapters relating to that node are shown, even if they are not related to a particular network.

Viewing Resource Groups and Resources

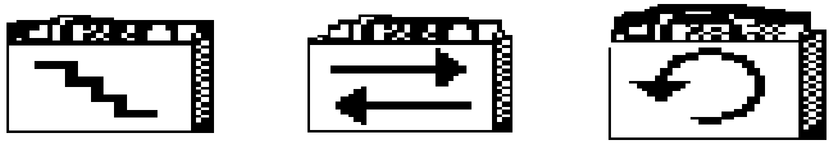
To view the status of a resource group, double-click a cluster or node symbol. A submap appears displaying symbols for all resource groups configured in the cluster, or on the specified node. Resource groups are symbolized by basket icons with a different logo to distinguish cascading, rotating, or concurrent resource group configurations.

To view individual resources, double-click a resource group symbol from either a cluster or node submap. A submap appears indicating the resources associated with the specified resource group.

Note: Individual Resources always appear in the unknown (blue) state in the submap. HAView indicates the *presence* of resources in a resource group, but does not monitor the *state* of individual resources.

Resource Group Symbols

The resource group symbols are shown below.



cascading resource group

concurrent resource group

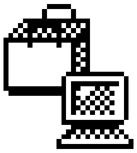
rotating resource group

HAView Resource Group Symbols

Resource Group Ownership Symbol

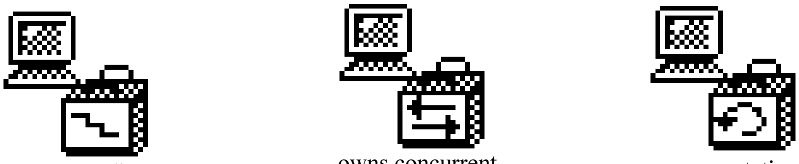
HAView indicates the current ownership of a resource group in both the Resource Group and Node submaps by showing the owner node together with the resource group as follows.

In the Resource Group submap, ownership is shown with this symbol:



Resource Group Ownership Symbol in Resource Group Submap

In the Node submap, these symbols indicate nodes that own a resource group, and the type of resource group owned:



owns cascading resource group

owns concurrent resource group

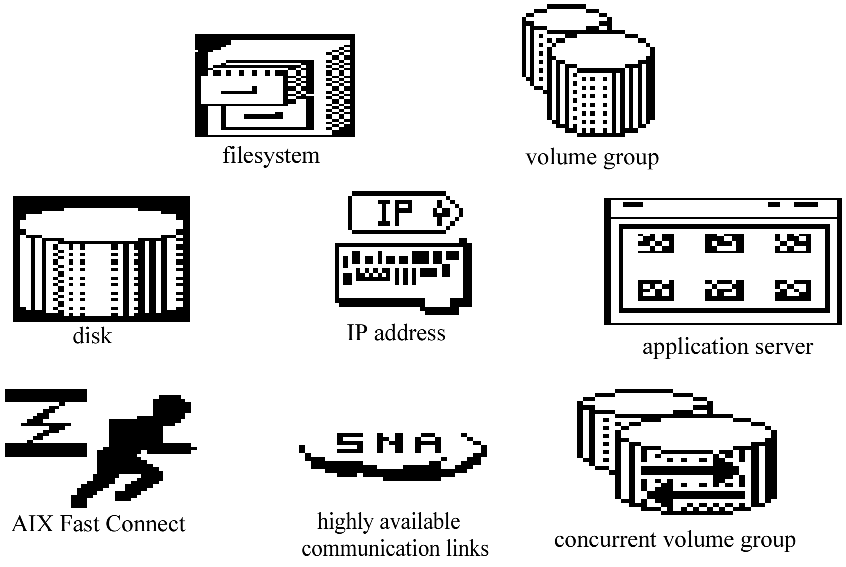
owns rotating resource group

Resource Group Ownership Symbols in Node Submap

Resource Submap—Individual Resource Symbols

The HAView Resource Group submap displays all the individual resources configured as part of a given resource group. Each type of resource has its own symbol, as shown below.

Note: Resource symbols will appear in the unknown (blue) state, since HAView does not monitor their state, only their presence.



Symbols for Individual Resource Types

Remember that individual resource symbols always appear in the blue (unknown) state, regardless of their actual state; HAView does not monitor the status of individual resources, only their presence and location.

Interpreting Resource Group Symbol Colors in HAView

Each symbol's color indicates the current state of the associated resource group, as follows:

Resource Group Status	Symbol Color	What is Occurring
Online/UP	green	The resource group is currently operating properly on one or more nodes in the cluster
Offline/DOWN	red	The resource group is not operating in the cluster and is not in an error condition.
Acquiring	yellow	The resource group is currently trying to come up on one of the nodes in the cluster.
Releasing	yellow	The resource group is in the process of being released from the ownership of a node.
Error	blue	The resource group has reported an error condition, and intervention is required.
Unknown	blue	The resource group's current status cannot be obtained, possibly due to a loss of communication between the monitoring node and the cluster.

Obtaining Component Details in HAView

NetView dialog boxes allow you to view detailed information about a cluster object. A dialog box can contain information about a cluster, network, node, network adapter, or resource group, or about cluster events. You can access an object's dialog box using the NetView menu bar or the Object Context menu, or by pressing **ctrl-o** at any time:

To view details about a cluster object using the NetView menu bar:

1. Click on an object in any submap.
2. Select the **Modify/Describe** option from the NetView Edit menu.
3. Select the **Object** option.

   An Object Description dialog window appears.
4. Select **HAView for AIX** and click on **View/Modify Object Attributes**.

   An Attributes dialog window appears.

   Note: You can view dialog boxes for more than one object simultaneously by either clicking the left mouse button and dragging to select multiple objects, or by pressing the **Alt** key and clicking on all object symbols for which you want more information.

To view details about a cluster object using the Object Context menu:

1. Click on an object in any submap.
2. Click on the symbol you have highlighted to display the object context menu, using:
   - Button 3 on a three-button mouse
   - Button 2 on a two-button mouse.
3. Select **Edit** from the object context menu.
4. Select **Modify/Describe** from the Edit cascade menu.
5. Select the **Object** option.

   An Object Description dialog window appears.
6. Select **HAView for AIX** and click on **View/Modify Object Attributes**.

   An Attributes dialog window appears.

HAView Polling Intervals

To ensure that HAView is optimized for system performance and reporting requirements, you can customize these two parameters:

- The polling interval (in seconds) at which HAView polls the HACMP/ES clusters to determine if cluster configuration or object status has changed. The default is 60 seconds.
- The polling interval (in minutes) at which HAView polls the **clhosts** file to determine if new clusters have been added. The default for Cluster Discovery polling is 120 minutes.

You can change the HAView polling intervals using the SMIT interface as follows:

1. On the HAView server node, open a SMIT screen by typing:

   ```
   smitty haview
   ```

   The Change/Show Server Configuration window opens.

2. Enter the polling interval numbers you want (between 1 and 32000) and press OK.

Note: If the **snmpd.conf** file is not properly configured to include the NetView server as a trap destination, HAView can detect a trap that occurs as a result of a cluster event, but information about the network topology may not be timely. Refer back to the section HAView File Modification Considerations on page 21-3 for more information on the snmpd.conf file.

Removing a Cluster from HAView

If a cluster does not respond to status polling, you can use the Remove Cluster option to remove the cluster from the database. To remove a cluster, it must be in an UNKNOWN state, represented by a blue cluster symbol. If the cluster is in any other state, the Remove Cluster option is disabled.

Warning: The **Remove Cluster** option is the only supported way to delete HAView objects from submaps. Do not delete an HAView symbol (cluster or otherwise) through the Delete Object or Delete Symbol menu items. If you use these menu items, HAView continues to poll the cluster.

When you remove a cluster, the following actions occur:

- The cluster name is removed from the NetView object database and HAView stops polling the cluster.
- The symbol for the cluster is deleted.
- The symbols for all child nodes, networks, addresses, and resource groups specific to that cluster are deleted.

If you are removing the cluster permanently, remember to remove the cluster addresses from the **/usr/es/sbin/cluster/etc/clhosts** file. If you do not remove the cluster addresses from the clhosts file, New Cluster Discovery polling continues to search for the cluster.

To remove a cluster:

1. Click on the cluster symbol you wish to remove. The cluster must be in an UNKNOWN state, represented by a blue cluster symbol.
2. Select **HAView** from the Tools pull-down menu.
3. Select **Remove Cluster** from the HAView cascade menu.

Using the HAView Cluster Administration Utility

HAView allows you to start a **smit hacmp** session to perform cluster administration functions from within the NetView session. The administration session is run on an aixterm opened on the chosen node through a remote shell. You can open multiple sessions of smit hacmp while in HAView.

Note: You can start an administration session for any node that is in an UP state (the node symbol is green). If you attempt to start an administration session when the state of the node is DOWN or UNKNOWN, no action occurs.

When bringing a node up, the HAView node symbol may show green before all resources are acquired. If you select the node symbol and attempt to open an administration session before all resources are acquired, you may receive an error.

Opening and Closing a Cluster Administration Session

To open a cluster administration session:

1. Click on an available node symbol (one that is green).
2. Select the **Tools > HAView > Cluster Administration**.
3. Proceed with your tasks in SMIT.
4. Select **F10** to exit the Cluster Administration session. The aixterm session will also close when using either of these choices.

Cluster Administration Notes and Requirements

Keep the following considerations in mind when using the Cluster Administration option:

- Be sure you have run the **xhost** command prior to starting NetView, so that a remote node can start an aixterm session on your machine.
- Be sure you have set the DISPLAY variable to a label that can be resolved and contacted from remote nodes.
- For the cluster administration session to proceed properly, the current NetView user (the account that started NetView) must have sufficient permission and be authenticated to perform an rsh to the remote node in the **/.rhosts** file or through Kerberos. (For more information, see the following sections: Editing the /.rhosts File on page 13-3, and Configuring Cluster Security on page 18-31.)
- If an IP Address Takeover (IPAT) occurs while a cluster administration session is running, the route between the remote node that the HAView monitoring node may be lost.

HAView Browsers

HAView provides two browsers which allow you to view the event history of a cluster, the Cluster Event Log and the HACMP Event Browser.

Cluster Event Log

Using the Cluster Event Log you can view the event history for a cluster as recorded by a specific node. The Log browser is accessible through the NetView Tools menu, and is only selectable if an active node symbol is highlighted.

For more detailed information on a node's event history, log onto the specific node and check the Cluster Message Log Files. See the Examining Cluster Log Files chapter in the *HACMP for AIX Troubleshooting Guide* for more information on Cluster Message Log Files.

Note: To ensure that the header for the Cluster Event Log displays properly, install all the NetView fonts on your system.

1. Click on the node symbol for which you wish to view a Cluster Event Log.
2. Select **HAView** from the Netview Tools menu.
3. Select the **Cluster Event Log** option.
4. Set the **number of events to view** field. You can use the up and down arrows to change this number or you can enter a number directly into the field. The possible range of values is 1 to 1000 records. The default value is 100.
5. Press the **Issue** button to generate the list of events. The message area at the bottom of the dialog box indicates when the list is done generating.

   When the list is done generating, the dialog box displays the following view-only fields:

   Event ID This field displays a numeric identification for each event which occurred on the cluster.

   Node Name The name of the node on which the event occurred.

   Time The date and time the event occurred. This field is in the format MM DD hh:mm:ss.

   Description A description of the event.
6. Press the **Dismiss** button to close the dialog box.

HACMP Event Browser

HAView provides a NetView browser which allows you to view the accumulative event history of a cluster. The browser shows the history of all nodes in the cluster, broadcast through an assigned primary node. If the primary node fails, another node will assume the primary role and continue broadcasting the event history.

The HACMP Event Browser provides information on cluster state events. A filter is used to block all redundant traps.

The HACMP Event Browser is available through the NetView menu bar. The menu item is always active, and when selected will start a NetView browser showing the event history for all active clusters. Note that you can access only one instantiation of the Event Browser at a time.

To view the HACMP Event Browser:

1. Select **HAView** from the Netview Tools menu.
2. Select the **HACMP Event Browser** option.

   The HACMP Event Browser appears. Note that only one instantiation of the Event Browser can be accessed at a time. See the *NetView for AIX User's Guide for Beginners* for more information on the NetView browser functions.

3. Select the **Close** option from the File menu of the HACMP Event Browser menu bar to close the browser.

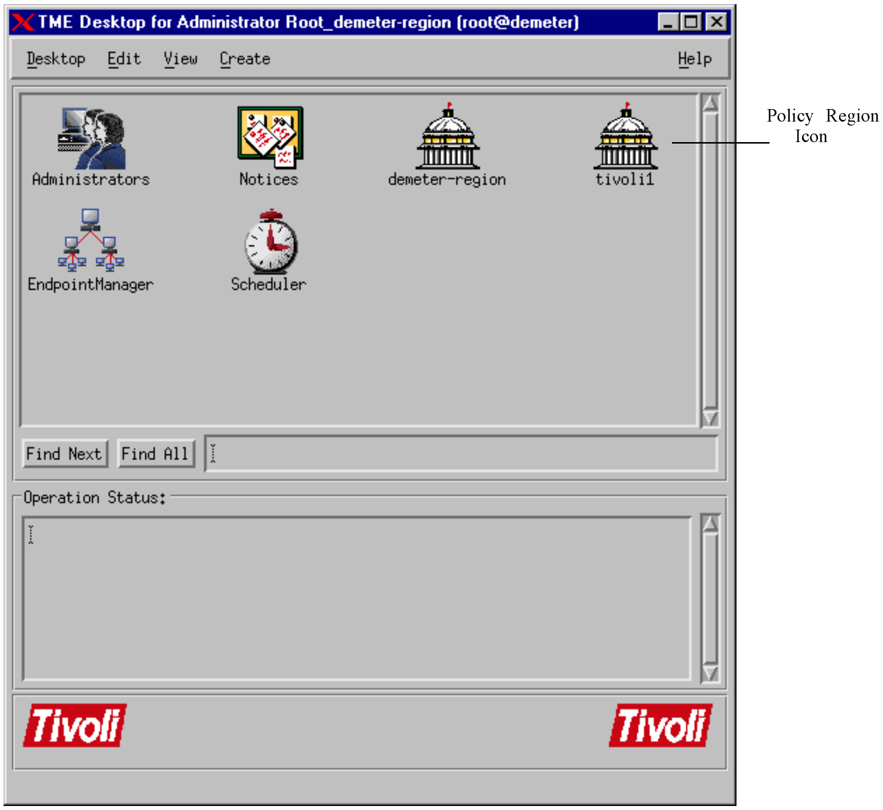
   Note: When you exit the Event Browser, the HAView application restarts. At this time, the HACMP cluster icon turns blue, disappears, and then reappears.

Monitoring HACMP/ES Clusters with Tivoli

You can monitor the state of an HACMP 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
- Configured resource group location and state
- Individual resource location (not state)

The initial TME Desktop view is shown here:



The Tivoli Desktop Initial Screen

Tivoli's thermometer icons (see figure on page 21-18) provide a visual indication of whether components are up, down, in transition, or in an unknown or error state. From the window for a selected Policy Region, you can go a cluster's Indicator Collection window, which displays thermometer icons indicating the state of all cluster components.

The cluster status information shown by the thermometers is updated every three minutes by default or at another interval you specify. (Further information on changing the default polling interval appears later in this chapter.)

In order to integrate HACMP/ES with Tivoli, you must configure your HACMP/ES cluster nodes as subscriber (client) nodes to the Tivoli server node, or Tivoli Management Region (TMR). Each cluster node can then maintain detailed node information in its local Tivoli database, which the TMR accesses for updated node information to display.

Complete installation instructions are located in Appendix D, Installing and Configuring Cluster Monitoring with Tivoli, in Volume 1 of this manual.

Consult Your Tivoli Documentation

The following sections provide information on monitoring an HACMP cluster through the Tivoli interface. Descriptions of Tivoli components and processes are provided here as needed, but for full information on installing, configuring, and using the Tivoli software itself, consult your Tivoli product user documentation.

Prerequisites and Considerations

As you planned and configured Cluster Monitoring with Tivoli, the following points should have been considered:

- The Tivoli Management Region (TMR) should be located on an AIX node outside the cluster.
- The Tivoli Framework, Distributed Monitoring, and AEF components must be installed on the Tivoli Management Region node and on each cluster node.
- In the Policy Region that contains HACMP cluster monitoring tasks and monitors, the Task Library and Profile Managers must be configured as managed resources.
- To ensure accurate monitoring of IP address takeover, consider using a dedicated network for communication between the TMR and the cluster nodes. If you do not have a separate network dedicated to Tivoli, you must set up your cluster to have one physical network with an extra subnet in order for IP address takeover and related post-events to occur. This subnet is used for the TMR adapter and for the cluster node's alias IP address.
- Also for proper IPAT monitoring, you create an alias for the IP address of each cluster node's standby adapter. You must place this alias in the **/etc/hosts** file and the **/usr/sbin/hativoli/ipaliases.conf** file. It is also recommended that it be placed in the Tivoli **/etc/wlocalhost** file.

For further detail on all prerequisites and installation issues for cluster monitoring with Tivoli, see Appendix D, Installing and Configuring Cluster Monitoring with Tivoli, in Volume 1 of this guide.

Installing and Configuring Cluster Monitoring with Tivoli

Preparing to monitor a cluster with Tivoli involves properly installing the Tivoli software, defining the necessary Tivoli components to handle cluster monitoring, and then installing the proper HACMP/ES filesets to make Tivoli aware of your HACMP/ES cluster nodes. In addition, there are some prerequisite tasks such as customizing post-event scripts in order to facilitate IP address takeover.

For complete instructions in installing HACMP/ES Tivoli filesets and making Tivoli aware of your cluster, see Appendix D, Installing and Configuring Cluster Monitoring with Tivoli, in Volume 1 of this manual.

Using Tivoli to Monitor the Cluster

Once you have properly installed your **hativoli** files and defined your nodes to Tivoli, you can view information on the status of your HACMP cluster components.

Note: **When all nodes are down, node status may not be displayed accurately.** You should be aware that in the event that your last remaining cluster node goes down, Tivoli may still indicate that the cluster is up. This can occur when HACMP is unable to contact the MIB for updated information. In this case, the Tivoli display will show information *as of the last successful poll*.

When you monitor your cluster through Tivoli, you can access cluster information in both icon and text form, in a number of different Tivoli windows. The next few sections are meant to orient you to the flow of Tivoli cluster monitoring information.

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.

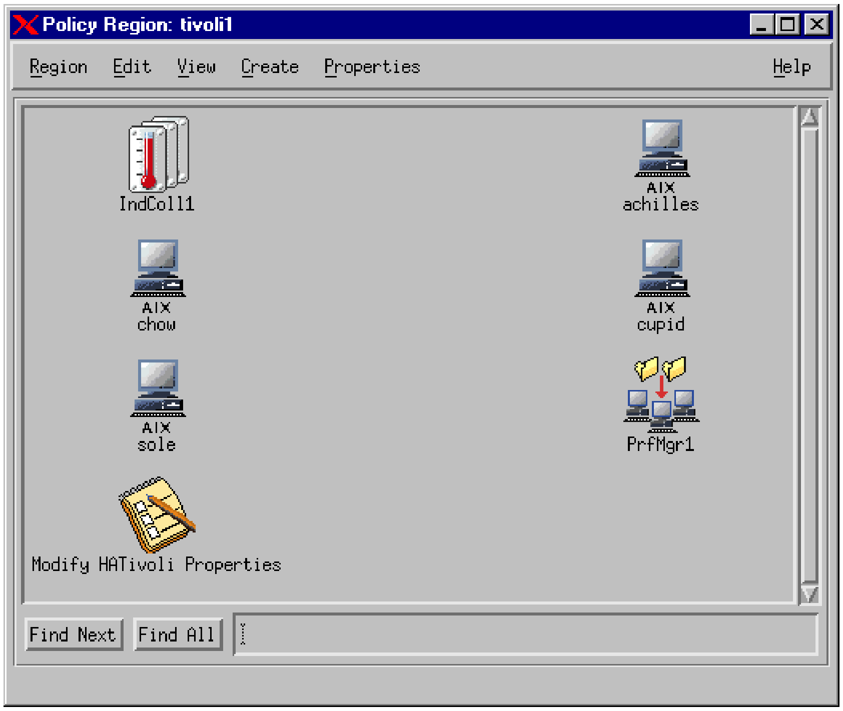
Note that there may be a delay as Tivoli adds the indicators for the cluster.

Tivoli Policy Regions

A Tivoli *Policy Region* groups together all entities related to a specific area of Tivoli monitoring. In this case, that area will be the HACMP/ES cluster or clusters. The HACMP/ES Policy Region will encompass the nodes, clusters, indicator icons, and tasks related to your HACMP/ES configuration.

You create this Policy Region during the process of defining your cluster nodes to Tivoli, as described in Appendix D, Installing and Configuring Cluster Monitoring with Tivoli, in Volume 1 of this guide.

Policy Region icons appear in the initial Tivoli Desktop window (shown on page 21-14). Clicking on a Policy Region icon opens the Policy Region window, in which you see the thermometer icons of the Indicator Collections, as well as icons for the profiles and tasks associated with the HACMP Policy Region.



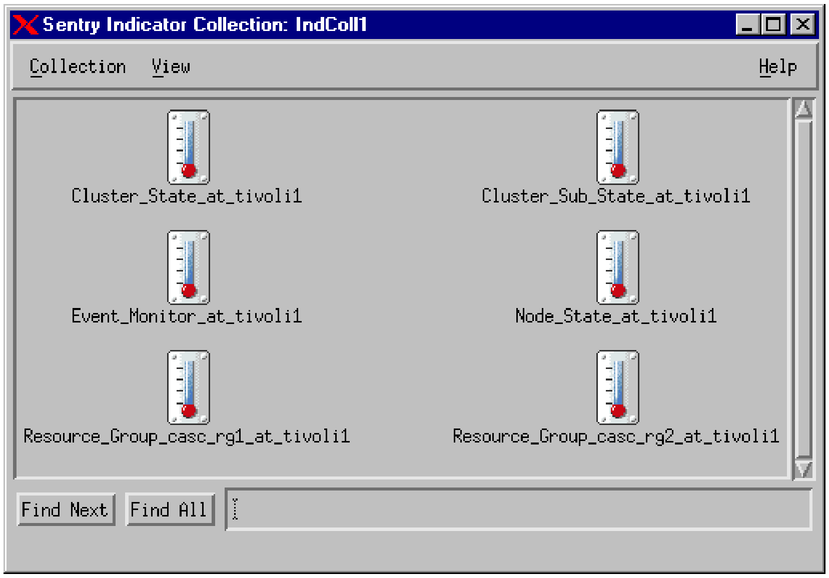
HAMCP Policy Region Window

Tivoli Distributed Monitors and Indicator Collections

For each cluster, a group of Tivoli *distributed monitors* is created that query the HACMP cluster node at set intervals for information about the various cluster components. The group of monitors is associated with an *indicator collection* that displays the state of the cluster components. If a change is detected in the state of the cluster or one of its components, the distributed monitor takes action, changing an icon in the associated Indicator Collection window. This provides the Tivoli administrator with a visual representation of any changes in the status of the monitored items.

New monitors are added whenever new cluster components are configured. When cluster components are removed from the cluster configuration, the associated monitors are also removed.

The icon for cluster and cluster component status is a thermometer figure with varying levels of red color depending on the severity of the component status. When you click on a cluster's Indicator Collection icon in the Policy Region window, the Indicator Collection window appears showing status icons for that cluster's components.


Sentry Indicator Collection: IndColl1
Collection
View
Help
Cluster_State_at_tivoli1
Cluster_Sub_State_at_tivoli1
Event_Monitor_at_tivoli1
Node_State_at_tivoli1
Resource_Group_casc_rg1_at_tivoli1
Resource_Group_casc_rg2_at_tivoli1
Find Next
Find All

Indicator Collection Window

Interpreting Indicator Displays for Various Cluster Components

The Indicator icons reflect varying degrees of severity of problems, depending on the height of the red color in the thermometer and the color-coded marker alongside it. The following tables list the indicator displays for various cluster component states:

CLUSTER STATE Indicator Display	Cluster State
Normal	UP
Fatal	DOWN
Severe	UNKNOWN

CLUSTER SUBSTATE Indicator Display	Cluster Substate
Normal	STABLE
Warning	UNSTABLE
Severe	RECONFIG
Critical	ERROR

NODE Indicator Display*	Node State(s)
Normal	All nodes ONLINE
Warning	One or more nodes OFFLINE

*Note that node state is displayed in the Distributed Monitor Indicator Collection as a composite view of all nodes rather than an individual node view.

RESOURCE GROUP Indicator Display	Resource Group State
Normal	ONLINE
Warning	ACQUIRING
Warning	RELEASING
Critical	ERROR
Fatal	OFFLINE

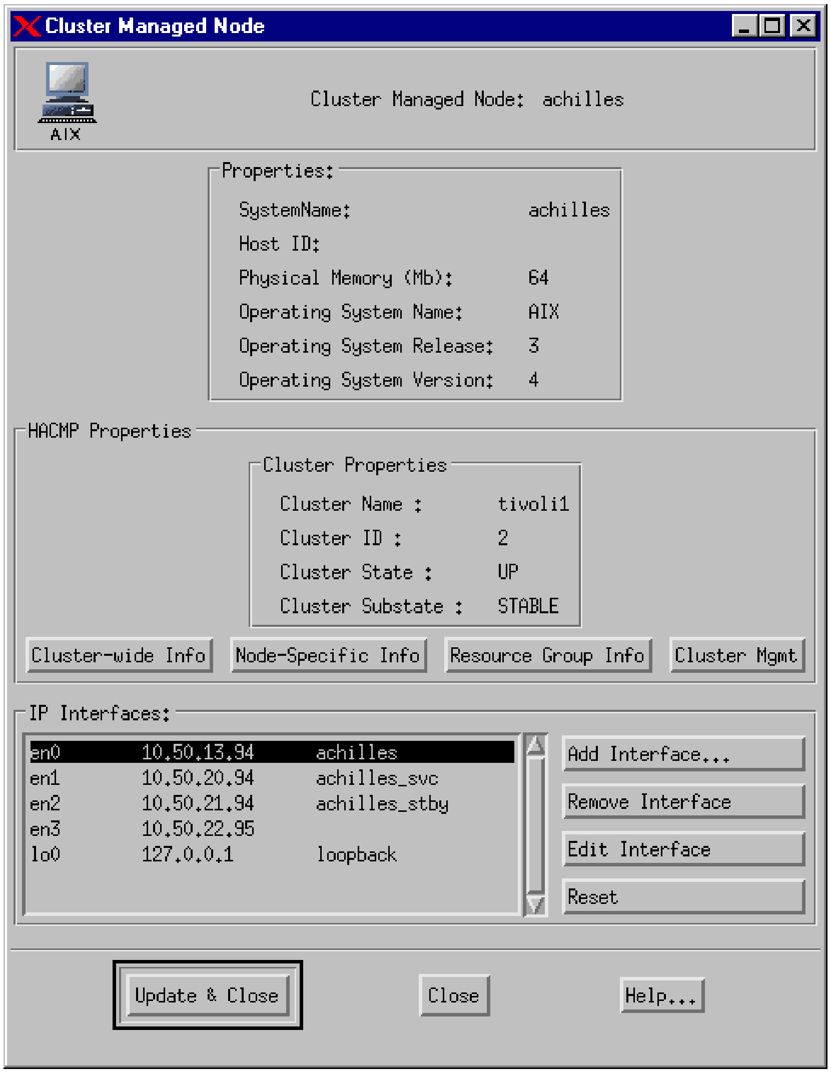
Viewing Cluster Information

The Cluster Managed Node window gives you all information about the current cluster topology and configuration.

The Properties section displays standard system properties information for the managed node, and the IP Interfaces section at the bottom shows standard IP Interface information for the node.

With the addition of the HACMP cluster monitoring feature, the standard Tivoli node icon is extended to capture and display additional information specific to your HACMP/ES cluster. The items that you see in the HACMP Properties portion of the window are detailed in the following sections.

To view the Managed Node window, right-click on a node icon in the Policy Region window.



Cluster Managed Node: achilles

Properties:

```
SystemName:                   achilles
Host ID:
Physical Memory (Mb):         64
Operating System Name:        AIX
Operating System Release:     3
Operating System Version:     4
```

Cluster Properties

```
Cluster Name :       tivoli1
Cluster ID :         2
Cluster State :      UP
Cluster Substate :   STABLE
```

IP Interfaces:

```
en0      10.50.13.94      achilles
en1      10.50.20.94      achilles_svc
en2      10.50.21.94      achilles_stby
en3      10.50.22.95
lo0      127.0.0.1        loopback
```

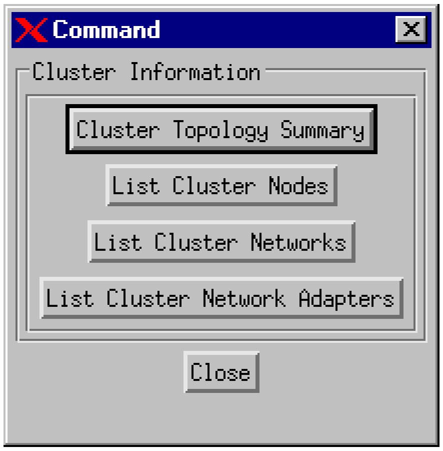
Cluster Managed Node Window

In the HACMP Properties portion of the Managed Node window shown above, you see the following four items of HACMP-specific top-level cluster information:

- Cluster name
- Cluster ID
- Cluster state
- Cluster substate

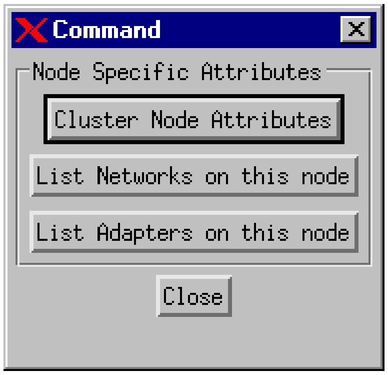
In addition, the HACMP Properties buttons lead you to further cluster and component details. Selecting a button brings up a new popup window with options for retrieving specific information. These buttons and the choices within them are detailed below.

Cluster-wide Information Button



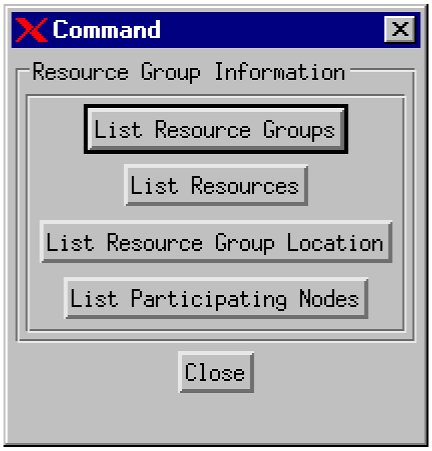
Clicking this button gives you the options shown above to view details on the configuration and status information for the cluster as a whole.

Node-Specific Information Button



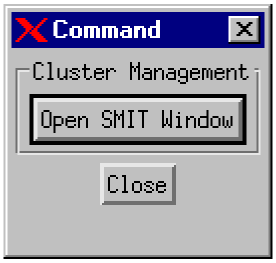
From the Node Specific Attributes window, you can access further information about the attributes, networks, and adapters associated with the node.

Resource Group Information Button



From the Resource Group Information window, you can access further details about the resource groups in your cluster and their associated resources and nodes.

Cluster Management Button



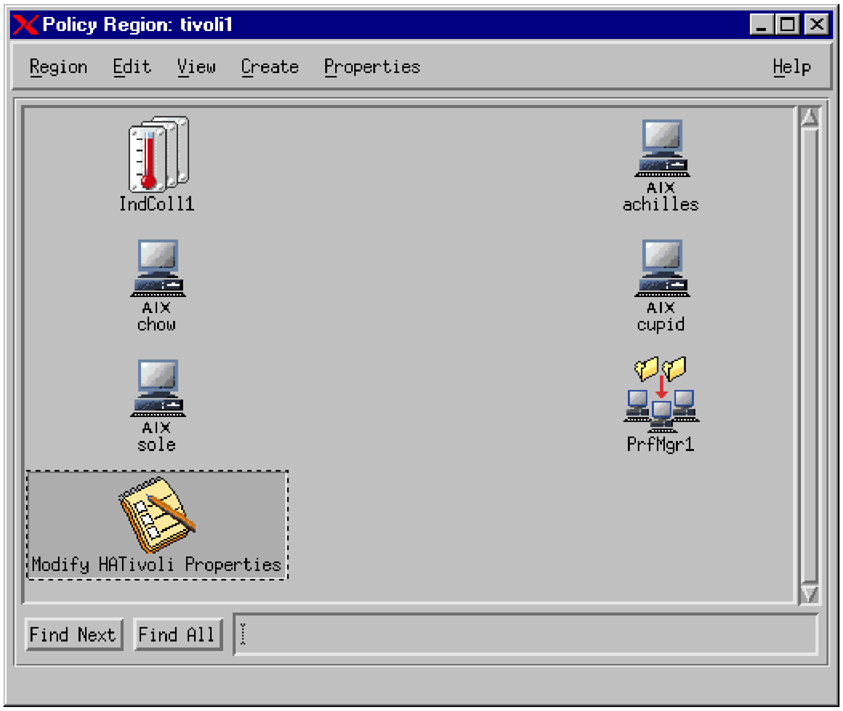
From the Cluster Management window, you can open a SMIT window to perform all of your normal cluster management tasks.

Note: In order to open a SMIT window from within Tivoli, you must have run the **xhost** command to grant access to remote nodes. See instructions in the section Starting Tivoli on page 21-16.

The HACMP Task Library

The Task Library contains options to perform cluster management tasks such as configuring, modifying, and deleting various items associated with the cluster, and refreshing the cluster view.

Task categories appear as “notepad” icons in the Policy Region window.



Policy Region Window with Task Icon Highlighted

Polling Intervals

The Distributed Monitors poll the cluster nodes periodically for cluster topology and status changes. The default polling interval is three minutes. If this interval is too short for your particular cluster monitoring needs, you can change this interval through an HACMP Task found in the Policy Region window. It is not recommended to make the polling interval shorter than the default.

Note: As mentioned earlier, be aware that in the event that your last remaining cluster node goes down, Tivoli may still indicate that the cluster is up. This can occur when HACMP is unable to contact the MIB for updated information. In this case, the Tivoli display will show information *as of the last successful poll*.

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.

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 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 node on the left, and click to move it to the right side.
    3. Click **Set & Close** to unsubscribe the nodes.

Monitoring Clusters with clstat

The **/usr/es/sbin/cluster/clstat** utility is a Clinfo client program that uses the Clinfo API to retrieve information about the cluster. Clinfo must be running on a node for the utilities to work properly. This utility runs on both ASCII and X Window Display clients in either single-cluster or multi-cluster mode. Multi-cluster mode requires that you use the **-i** flag when invoking the **clstat** utility. The client display automatically corresponds to the capability of the system. For example, if you run **clstat** on an X Window client, a graphical display for the utility appears; however, you can run an ASCII display on an X-capable machine by specifying the **-a** flag.

The **clstat** utility reports whether the cluster is up, down, or unstable. It also reports whether a node is up, down, joining, leaving, or reconfiguring, and the number of nodes in the cluster. For each node, the utility displays the IP label and address of each network interface attached to the node, and whether that interface is up or down. See the **clstat** man page for additional information about this utility.

The LPP contains both executables and source code for the clstat utility. If you want to recompile clstat, run the **make** command in the directory **/usr/es/lpp/cluster/samples/clstat.**

Single-Cluster clstat ASCII Display Mode

In single-cluster ASCII display mode, the **clstat** utility displays information about *only* one cluster. To invoke the **clstat** utility in single-cluster (non-interactive) mode, enter:

```
clstat
```

A screen similar to the following appears:

```
              clstat - HACMP for AIX Cluster Status Monitor
                ---------------------------------------------

Cluster: ibm_26c        (666)           Thu Jul  9 16:47:24 EDT 1998
                State: DOWN             Nodes: 2
                SubState: UNSTABLE
        Node: poseidon          State: DOWN
           Interface: poseidon_enboot (0)       Address: 140.186.70.106
                                                State:   DOWN

        Node: venus             State: DOWN
           Interface: venus_enboot (0)          Address: 140.186.70.107
                                                State:   DOWN


***************** f/forward, b/back, r/refresh, q/quit ***************
```

The **clstat** Single-Cluster ASCII Display Mode

The cluster information displayed shows both a cluster name and an ID. In this example, the cluster is down and has two nodes, both of which are down. Each node has one network adapter. Note that the *forward* and *back* menu options apply when more than one page of information is available to display.

If more than one cluster exists when you run the **clstat** command, the utility notifies you of this fact and requests that you retry the command specifying one of the following options:

clstat [-c *cluster ID*] [-r *seconds*] [-i]

where:

-c *ID* Displays information about the cluster with the specified ID if that cluster is active. This option cannot be used with the **-n** option.

If the cluster is not available, the **clstat** utility continues looking for it until it is found or until the program is cancelled. Note that this option cannot be used if the **-i** option (for multi-cluster mode) is used.

-r *seconds* Updates the cluster status display at the specified number of seconds. The default is 1 second; however, the display is updated only if the cluster state changes.

-a Causes **clstat** to display in ASCII mode.

-i Displays information about clusters interactively. Only valid when running **clstat** in ASCII mode.

To see cluster information about a specific cluster, enter:

```
clstat [-c cluster ID]
```

Multi-Cluster clstat ASCII Display Mode

The multi-cluster (interactive) mode lets you monitor all clusters that Clinfo can access from the list of active service IP labels or addresses found in the **/usr/es/sbin/cluster/etc/clhosts** file. In multi-cluster mode, the **clstat** utility displays this list of recognized clusters and their IDs, allowing you to select a specific cluster to monitor. For example, to invoke the **clstat** utility in multi-cluster mode, enter:

```
clstat -i
```

where the **-i** indicates multi-cluster (interactive) ASCII mode. A screen similar to the following appears.

```
              clstat - HACMP for AIX Cluster Status Monitor
                ---------------------------------------------

Number of clusters active: 1

                ID      Name            State

               666   ibm_26c            DOWN

Select an option:

              # - the Cluster ID                     x- quit
```

The **clstat** Multi-Cluster Mode Menu

This screen displays the ID, name, and state of each active cluster accessible by the local node. You can either select a cluster to see detailed information, or quit the **clstat** utility.

When you enter a cluster ID, a screen appears similar to the one that follows.

```
              clstat - HACMP for AIX Cluster Status Monitor
                ---------------------------------------------

Cluster: ibm_26c        (666)           Thu Jul  9 18:35:46 EDT 1998
                State: DOWN             Nodes: 2
                SubState: UNSTABLE
        Node: poseidon          State: DOWN
           Interface: poseidon_enboot (0)       Address: 140.186.70.106
                                                State:   DOWN

        Node: venus             State: DOWN
           Interface: venus_enboot (0)          Address: 140.186.70.107
                                                State:   DOWN
***************** f/forward, b/back, r/refresh, q/quit ************
```

The clstat Multi-Cluster ASCII Display Mode

After viewing this screen, press **q** to exit the display. The multi-cluster mode returns you to the cluster list so you can select a different cluster. Note that you can use all menu options displayed. The *forward* and *back* options allow you to scroll through displays of active clusters without returning to the previous screen.

The clstat Utility X Window System Display

When you start the **/usr/es/sbin/cluster/clstat** utility on a node capable of displaying X Window System applications, the **clstat** utility displays its graphical interface, if the client's DISPLAY environment variable is set to the value of the X server's node address.

To invoke the **clstat** utility X Window System display, enter the **clstat** command:

```
clstat [-n name] [-c Id] [ -r #] [-D debug_level]
```

where:

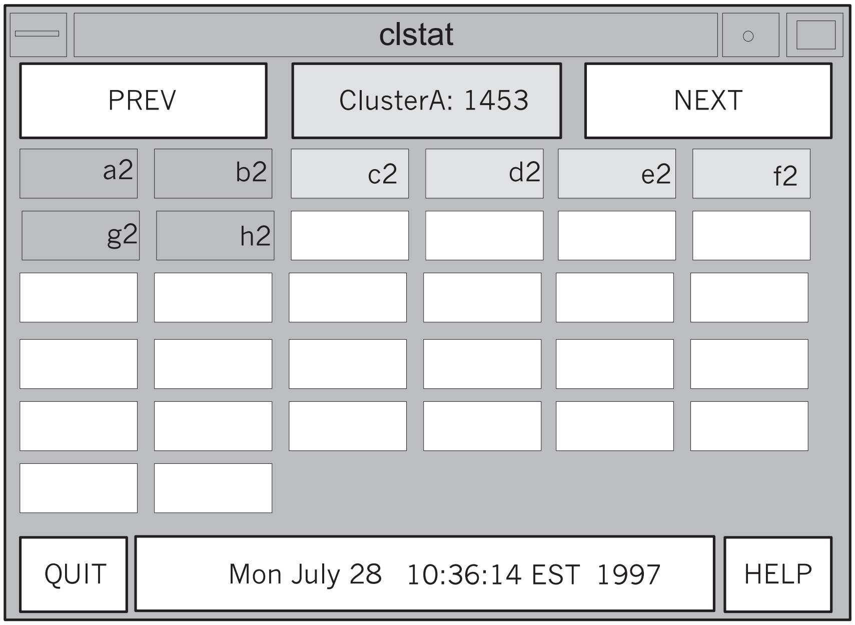
-n *name* The cluster name.

-c *ID* Displays information about the cluster with the specified ID if that cluster is active. This option cannot be used with the **-n** option.

-r # The interval at which the **clstat** utility updates the display. For the graphical interface, this value is interpreted in tenths of seconds. By default, **clstat** updates the display every 0.10 seconds.

-D *debug_level* The level of debugging to be performed. The levels range from 1 to 9 in increasing amounts of information. The default (0) turns debugging off.

The **clstat** utility graphical interface uses windows to represent cluster nodes.



The clstat X Window System Display

The middle box in the top row indicates the cluster name and ID. If the cluster is stable, this box appears green. If the cluster destabilizes for any reason, this box changes to red.

The large boxes in other rows represent nodes. A node name appears in a box for each active node in the cluster. You can see up to sixteen nodes per cluster. Nodes that are up are shown in green, nodes that are down are shown in red, nodes that are joining or leaving the cluster are shown in yellow (topology changes), and nodes that are undefined are shown in the background color. Colors are configured in the **Xclstat** X Window resource file in the **/usr/lpp/cluster/samples/clstat** directory.

On a monochrome display, gray shading represents the colors as follows:

red	dark gray
yellow	gray
green	light gray

Five buttons are available on the **clstat** display:

PREV Displays the previous cluster (loops from end to start).

NEXT Displays the next cluster (loops from start to end).

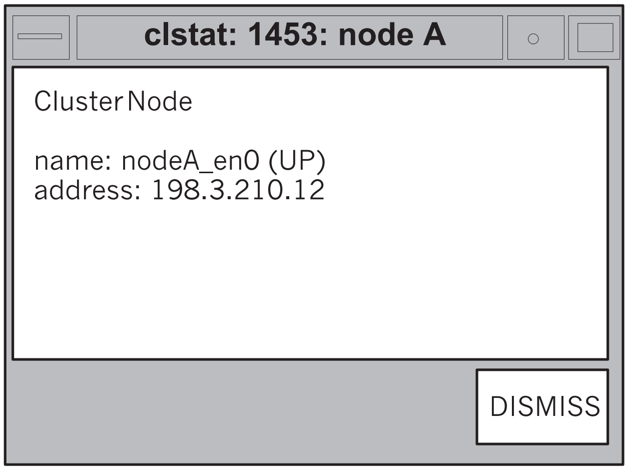
cluster:ID The refresh bar. Pressing this bar updates the status display.

QUIT Cancels the **clstat** utility.

HELP Displays help information.

Viewing Network Interface Information on an X Window System Display

To view information about network interfaces attached to a node, click mouse button 1 on the appropriate node box in the **clstat** display. A pop-up window similar to the following appears. The title in the example shows that you are viewing nodeA in cluster *1453*.



```
Cluster Node

name: nodeA_en0 (UP)
address: 198.3.210.12
```

The clstat Node Information Display

Press the DISMISS button to close the pop-up window and to return to the **clstat** display window. Do not use the **Close** option in the pull-down menu in the upper left corner of the window to close this display; it terminates the **clstat** utility.

Monitoring Applications

HACMP/ES can monitor specified applications and attempt to restart them upon detecting process death or application failure. Application monitoring works in one of two ways:

- **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, it 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.

In either case, 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 this retry count, HACMP/ES takes one of two actions, which you specify when configuring the application monitor:

- Choosing **fallover** causes the resource group containing the application to fall over to the node with the next-highest priority according to the resource policy.
- Choosing **notify** causes HACMP to generate a server_down event, similar to a network_down event, to inform the cluster of the failure.

When you configure an application monitor, you use the SMIT interface to specify 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.

Note: Be aware that if you specify the fallover option, which may cause a resource group to migrate from its original node, the possibility exists that even when the highest priority node is up, the resource group remains down. Unless you bring the resource group up manually, it could remain in an inactive state. See the troubleshooting section Common Problems and Solutions on page 29-24 for more information.

For complete information on configuring application monitoring, see Chapter 18, Configuring an HACMP/ES Cluster, in Volume 1 of this manual.

Using Resource Group Information Commands

You can monitor resource group status and location by using the HAView utility, as discussed earlier in this chapter. You can also locate resource groups using the command line.

You can use either of two commands to monitor resource group status and location:

- The **clRGinfo** command
- The **clfindres** utility.

The commands give similar information; the **clfindres** command tells you the current location and if a resource group has a “sticky” location, that is listed also. See the section Resource Migration Types: Sticky and Non-sticky on page 24-31 for more information on sticky locations. See the man pages for complete information on using these commands.

Using the clRGinfo Command

Running the **clRGinfo** command gives you a report on the location and status of one or more specified resource groups. A resource group can be in any one of the following states:

- *Online* - The resource group is currently operating properly on one or more nodes in the cluster.
- *Offline* - The resource group is not operating in the cluster and is currently not in an error condition.
- *Acquiring* - A resource group is currently coming up on one of the nodes in the cluster.
- *Releasing* - The resource group is in the process of being released from ownership by one node. Under normal conditions after being successfully released from a node the resource group's status changes to offline.
- *Error* - The resource group has reported an error condition. User interaction is required.
- *Unknown* - The resource group's current status cannot be attained, possibly due to loss of communication, or the fact that all nodes in the cluster are not up.

Sample Output for clRGinfo Command

The following command produces a report in normal format. The command asks for the status and location of three resource groups, named group1, group2, and group3.:

```
clRGinfo group1 group2 group3
---------------------------------------------------
GroupName             State                   Location
----------------------------------------------------
group1                ONLINE                  node1
                      ONLINE                  node2
                      ACQUIRING               node3
                      OFFLINE                 node4
group2                ONLINE                  node3
group3                RELEASING               node2
```

The same command with the **-s** (shortened) flag produces a condensed colon-separated format:

```
clRGinfo -s group1 group2 group3

group1:ONLINE:node1
group1:ONLINE:node2
group1:ACQUIRING:node3
group1:OFFLINE:node4
group2:ONLINE:node3
group3:RELEASING:node2
```

See Cluster Resource Group Information Commands on page E-19 for more information.

Using the clfindres Command

To help you check the location and state of resources placed on a specific node, the DARE Resource Migration utility includes a command, **clfindres**, that displays the state and location of specified resource groups. It also indicates whether a resource group has a sticky location, and it identifies that location.

See Appendix A, HACMP for AIX Commands in the *HACMP for AIX Administration Guide* for the complete syntax and typical output of the **clfindres** command.

Using the clRMupdate Command

The cluster events scripts provided with HACMP/ES update the **clstrmgr** daemon with resource group status and location. Under normal operation no manual intervention is needed in this process. However, if you have manually changed the configuration or you note another instance where the resource group information displayed by a monitoring utility like HAView is incorrect, you can manually update the resource group information using the **clRMupdate** command. For example, if you modify a resource group outside the cluster by bringing it down manually on one node and starting it on another node, you should update the **clstrmgr** daemon to accurately reflect the new state and location of the resource group.

Warning: The resource group information is used by the event scripts to determine resource group movement during cluster events. Manually updating the resource group information may lead to incorrect fallover operations.

Sample Use of the clRMupdate Command

Use the **clRGinfo** command to examine the current resource group information. If you decide that you need to update this information, you can then use the **clRMupdate** command. For example, if you decide a particular resource group *group1* should be up on Node A, you would execute the following command from Node A:

```
/usr/es/sbin/cluster/events/utils/clRMupdate rg_up group1
```

See Cluster Resource Group Information Commands on page E-19 for complete information.

Monitoring Cluster Events—Event Emulation

HACMP/ES provides a utility that lets you emulate cluster events by running event scripts that produce output but do not affect the cluster configuration status. This allows you to predict a cluster's reaction to an event as though the event actually occurred.

The Event Emulator follows the same procedure used by the Cluster Manager given a particular event, but does not execute any commands that would change the status of the Cluster Manager. For descriptions of cluster events and how the Cluster Manager processes these events, see Chapter 18, Configuring an HACMP/ES Cluster. For more information on the cluster log redirection functionality, see Chapter 24, Changing the Cluster Configuration.

The event emulator runs the events scripts on every active node of a stable cluster. Output from each node is stored in an output file on the node from which you invoked the emulation. You can specify the name and location of the output file using the environment variable EMUL_OUTPUT or use the default output file, **/tmp/emuhacmp.out**.

Event Emulator Considerations

Keep the following cautions in mind when using the Event Emulator:

- You can only run one instance of the event emulator at a time. If you attempt to start a new emulation in a cluster while an emulation is already running, the integrity of the results cannot be guaranteed. Each emulation is a stand-alone process; one emulation cannot be based on the results of a previous emulation.
- **clinfo** must be running.
- You should add a cluster snapshot before running an emulation, just in case uncontrolled cluster events happen during emulation. Instructions for adding cluster snapshots are in Chapter 26, Saving and Restoring Cluster Configurations.
- The Event Emulator can run only event scripts that comply with the currently active configuration. For example:
  - The Emulator expects to see the same environmental arguments used by the Cluster Manager; if you define arbitrary arguments, the event scripts will run, but error reports will result.
  - In the case of swap_adapter, you must enter the ip_label supplied for service and standby adapters in the correct order, as specified in the usage statement. Both adapters must be located on the same node at emulation time. Both must be configured as part of the same HACMP/ES logical network.

  For other events, the same types of restrictions apply. If errors occur during emulation, recheck your configuration to ensure that the cluster state supports the event to be emulated.
- The Event Emulator runs customized scripts (pre- and post-event scripts) associated with an event, but does not execute commands within these scripts. Therefore, if these customized scripts change the cluster configuration when actually run, the outcome may differ from the outcome of an emulation.
- When emulating an event which contains a customized script, the Event Emulator uses the **ksh** flags **-n** and **-v**. The **-n** flag reads commands and checks them for syntax errors, but does not execute them. The **-v** flag indicates verbose mode. When writing customized

scripts that may be accessed during an emulation, be aware that the other **ksh** flags may not be compatible with the **-n** flag and may cause unpredictable results during the emulation. See the **ksh** man page for flag descriptions.

Running Event Emulations

You can run the event emulator through SMIT or the command line. See the **cl_emulate** man page for information on emulating Cluster Events from the command line. Complete the following steps to emulate a cluster event from SMIT.

1. To start system management for HACMP/ES, enter:

   ```
   smit hacmp
   ```

2. Select **RAS Support** and press Enter.
3. Select **Event Emulator** and press Enter.

SMIT displays a screen with options. Each option provides a different cluster event to emulate. The following sections provide more information about each option.

Emulating a Node Up Event

To emulate a Node Up event:

1. Select **Node Up Event** from the **Select Event to Emulate** screen, and press enter. SMIT displays the screen.
2. Enter field data as follows:

   Node Name Enter the name of the node to use in the emulation.

3. Press Enter to start the emulation.

Emulating a Node Down Event

To emulate a Node Down event:

1. Select **Node Down Event** from the **Select Event to Emulate** screen, and press Enter. SMIT displays the screen.
2. Enter field data as follows:

   Node Name Enter the node to use in the emulation.

   Shutdown Mode Indicate the type of shutdown to emulate:

   - **graceful**: Shutdown after the **/usr/es/sbin/cluster/events/node_down_complete** script is run on this node to release its resources. The other nodes do not take over the resources of the stopped node.

   - **graceful with takeover:** Shutdown after the **/usr/es/sbin/cluster/events/node_down_complete** script runs to release its resources. The other nodes do take over the resources of the stopped node.

3. Press Enter to start the emulation.

Emulating a Network Up Event

To emulate a Network Up event:

1. From the **Select Event to Emulate** screen, select **Network Up Event** and press enter. SMIT displays the screen.
2. Enter field data as follows:

   Network Name Enter the network to use in the emulation.

   Node Name (Optional) Enter the node to use in the emulation.

3. Press Enter to start the emulation.

Emulating a Network Down Event

To emulate a Network Down event:

1. From the **Select Event to Emulate** screen, select **Network Down Event** and press enter. SMIT displays the screen.
2. Enter field data as follows:

   Network Name Enter the network to use in the emulation.

   Node Name (Optional) Enter the node to use in the emulation.

3. Press Enter to start the emulation.

Emulating a Fail Standby Event

To emulate a Fail Standby event:

1. Select **Fail Standby Event** from the **Select Event to Emulate** screen, and press enter. SMIT displays the Fail Standby Event screen.
2. Enter field data as follows:

   Node Name Enter the node to use in the emulation.

   IP Label Enter the IP label to use in the emulation.

3. Press Enter to start the emulation.

Note: The following messages are displayed on all active cluster nodes when emulating the Fail Standby and Join Standby events:

```
Adapter $ADDR is no longer available for use as a standby, due to
either a standby adapter failure or IP address takeover.

Standby adapter $ADDR is now available.
```

Emulating a Join Standby Event

To emulate a Join Standby event:

1. From the **Select Event to Emulate** screen, select **Join Standby Event** and press enter. SMIT displays the Join Standby Event screen.
2. Enter field data as follows:

Node Name	Enter the node to use in the emulation.
IP Label	Enter the IP label to use in the emulation.

3. Press Enter to start the emulation.

Note: The following messages are displayed on all active cluster nodes when emulating the Fail Standby and Join Standby events:

```
Adapter $ADDR is no longer available for use as a standby, due to
either a standby adapter failure or IP address takeover.

Standby adapter $ADDR is now available.
```

Emulating a Swap Adapter Event

To emulate a Swap Adapter event:

1. From the **Select Event to Emulate** screen, select **Swap Adapter Event** and press enter. SMIT displays the Swap Adapter Event screen.
2. Enter field data as follows:

Node Name	Enter the node to use in the emulation.
Network Name	Enter the network to use in the emulation.
IP Label (Standby Label used to Swap)	The name of the Standby Adapter to which you want to switch.
IP Label (Service Label to Fail)	The name of the Service Adapter that failed.

3. Press Enter to start the emulation.

Monitoring Dynamic Reconfiguration Events—Event Emulation

To run an emulation of a Dynamic Reconfiguration event, modify the cluster configuration to reflect the configuration to be emulated and use the SMIT screens explained in this section.

Note: The Event Emulator will not change the configuration of a cluster device. Therefore, if your configuration contains a process that makes changes to the Cluster Manager (disk fencing, for example), the Event Emulator will not show these changes. This could lead to a different output, especially if the hardware devices cause a fallover.

You should add a cluster snapshot before running an emulation, just in case uncontrolled cluster events happen during emulation. Instructions for adding cluster snapshots are in Chapter 26, Saving and Restoring Cluster Configurations.

To emulate synchronizing a cluster resource event:

1. Enter the SMIT fastpath:

   ```
   smit hacmp
   ```

2. Select **Cluster Configuration** > **Cluster Resources > Synchronize Cluster Resources** and press Enter.
3. Enter field data as follows:

   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.

4. Press Enter to start the emulation.

After you run the emulation, if you do not wish to run an actual DARE, you can restore the original configuration using the SMIT screen option **Restore System Default Configuration from Active Configuration**.

Monitoring Cluster Services

After checking cluster, node, and network interface status, check the status of the HACMP/ES daemons on both nodes and clients.

Monitoring Cluster Services on a Node

Use the SMIT Show Cluster Services screen to check the status of the HACMP/ES daemons on a node. Check the following daemons:

- Cluster Manager (**clstrmgr**) daemon
- Cluster SMUX Peer (**clsmuxpd**) daemon

  Note: If the **clsmuxpd** daemon is not active, also check the status of the SNMP (**snmpd**) daemon.

- Clinfo (**clinfo**) daemon (optional on a node).
- Cluster Lock Manager (**cllockd)** daemon (optional)

Use the **smit clshow** fastpath to view cluster services on a node.

1. Enter:

   ```
   smit clshow
   ```

   A screen similar to that shown in the figure below appears.

```
                      COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

Subsystem             Group            PID     Status
clstrmgrES            cluster          18524   active
clinfoES              cluster          15024   active
clsmuxpdES            cluster          16926   active
cllockdES             lock                     inoperative
```

The Command Status Display of Node Information

The screen indicates the:

- HACMP/ES subsystem
- Group to which the subsystem belongs
- Process ID number of the subsystem, if it is running
- Status of the subsystem.

Monitoring Cluster Services on a Client

The only HACMP/ES process that can run on a client is the **clinfo** daemon. (Not all clients run this daemon.) You can use the AIX **lssrc** command with either the **-g cluster** or **-s clinfo** arguments to check the status of the **clinfo** daemon on a client. The output looks similar to the following:

```
Subsystem      Group    PID     Status

clinfoES       cluster  9843    active
```

You can also use the **ps** command and **grep** for “clinfo”. For example:

```
ps -aux | grep clinfo
```

HACMP/ES Log Files

HACMP/ES writes the messages it generates to the system console and to several log files. Because each log file contains a different subset of the types of messages generated by HACMP/ES, you can get different views of cluster status by viewing different log files. HACMP/ES writes messages into the log files described below. See Chapter 29, Troubleshooting HACMP/ES Clusters, for more information about these files.

Note: The default locations of log files are used in this chapter. If you redirected any logs, check the appropriate location.

/usr/es/adm/cluster.log File

The **/usr/es/adm/cluster.log** file is the main HACMP/ES log file. HACMP/ES error messages and messages about HACMP/ES-related events are appended to this log with the time and date at which they occurred.

/tmp/hacmp.out File

The **/tmp/hacmp.out** file records the output generated by the configuration and startup scripts as they execute. This information supplements and expands upon the information in the **/usr/es/adm/cluster.log** file. To receive verbose output, the Debug Level run-time parameter should be set to *high* (the default). See Chapter 28, Additional Tasks: NFS and Run-Time Parameters, for details on setting run-time parameters.

/usr/es/sbin/cluster/history/cluster.*mmdd* File

The **/usr/es/sbin/cluster/history/cluster.*mmdd*** file contains time stamped, formatted messages generated by HACMP/ES scripts. The system creates a cluster history file whenever cluster events occur, identifying each file by the file name extension *mmdd*, where *mm* indicates the month and *dd* indicates the day.

While it is more likely that you will use these files during troubleshooting, you should occasionally look at them to get a more detailed picture of the activity within a cluster.

System Error Log File

The system error log file contains time stamped, formatted messages from all AIX subsystems, including HACMP/ES scripts and daemons. Cluster events are logged as operator messages (error id: AA8AB241) in the system error log.

/tmp/clstrmgr.debug Log File

The **clstrmgr.debug** log file contains time-stamped, formatted messages generated by HACMP/ES **clstrmgr** activity. This file is typically used by IBM support personnel.

/tmp/cspoc.log File

The **cpoc.log** file contains time-stamped, formatted messages generated by HACMP/ES C-SPOC commands. The **/tmp/cspoc.log** file resides on the node that invokes the C-SPOC command.

/var/ha/log/grpsvcs.<filename> File

Contains time-stamped messages in ASCII format. These track the execution of internal activities of the **grpsvcs** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

/var/ha/log/topsvcs.<filename> File

The **/var/ha/log/topsvcs.<filename>** log file contains time-stamped messages in ASCII format. These track the execution of internal activities of the **topsvcs** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

/var/ha/log/grpglsm File

The **/var/ha/log/grpglsm** file tracks the execution of internal activities of the **grpglsm** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

/tmp/emuhacmp.out File

The **/tmp/emuhacmp.out** file records the output generated by the event emulator scripts as they execute. The **/tmp/emuhacmp.out** file resides on the node from which the event emulator is invoked. You can use the environment variable EMUL_OUTPUT to specify another name and location for this file, but the format and information remains the same.

Chapter 22 Maintaining Shared LVM Components

This chapter explains how to maintain LVM components shared by nodes in an HACMP/ES cluster. This chapter includes specific procedures for managing volume groups, filesystems, logical volumes, and physical volumes. The chapter also includes information about using the C-SPOC utility to maintain LVM components in a cluster.

For specific information on AIX commands and SMIT screens, see your *AIX System Management Guide*.

Overview

A key element of any HACMP/ES cluster is the data used by the highly available applications. This data is stored on AIX Logical Volume Manager (LVM) entities. HACMP/ES clusters use the capabilities of the LVM to make this data accessible to multiple nodes.

In an HACMP/ES cluster, a *shared volume group* is a volume group that resides entirely on the external disks shared by cluster nodes. A *shared physical volume* is a disk that resides in a shared volume group. A *shared logical volume* is a logical volume that resides entirely in a shared volume group. A *shared filesystem* is a filesystem that resides entirely in a shared logical volume.

Common Maintenance Tasks

As a system administrator of an HACMP/ES cluster, you may be called upon to perform any of the following LVM-related tasks:

- Creating a new shared volume group
- Extending, reducing, changing, or removing an existing volume group
- Creating a new shared logical volume
- Extending, reducing, changing, or removing an existing logical volume
- Creating a new shared filesystem
- Extending, changing, or removing an existing filesystem
- Adding, removing physical volumes

When performing any of these maintenance tasks on shared LVM components, make sure that ownership and permissions are reset when a volume group is exported and then re-imported. After exporting and importing, a volume group is owned by root and accessible by the system group. Applications, such as some database servers, that use raw logical volumes may be affected by this if they change the ownership of the raw logical volume device. You must restore the ownership and permissions back to what is needed after this sequence.

Using the C-SPOC Utility to Maintain Shared LVM Components

HACMP/ES provides the Cluster-Single Point of Control (C-SPOC) utility, that simplifies maintenance of shared LVM components in clusters of up to 32 SP nodes in a single partition, or 8 nodes in clusters that include stand-alone RS/6000s or that span partitions on an SP. C-SPOC commands provide comparable functions in a cluster environment to the standard AIX commands that work on a single node. For example, the C-SPOC utility includes a command called **cl_chlv** that provides similar functions to the AIX **chlv** command. (The C-SPOC command calls the AIX command.) By automating repetitive tasks, C-SPOC eliminates a potential source of errors, and speeds up the process.

LVM maintenance tasks that you can perform using C-SPOC:

Note: The C-SPOC commands only operate on both shared and concurrent LVM components that are defined as part of an HACMP/ES resource group.

- Shared volume groups
  - Create a shared volume group
  - Import a volume group
  - Extend a volume group
  - Reduce a volume group
  - Mirror a volume group
  - Unmirror a volume group
  - Synchronize volume group mirrors
  - List all shared volume groups
  - List all active shared volume groups
  - Display characteristics of a shared volume group.
- Shared logical volumes
  - Create a logical volume
  - Make a copy of a logical volume
  - Remove a copy of a logical volume
  - List all shared logical volumes by volume group
  - Change or view the characteristics of a shared logical volume (name, size)
  - Remove a shared logical volume.
- Shared filesystems
  - Create a shared filesystem
  - List all shared filesystems
  - Change or view the characteristics of a shared filesystem
  - Remove a shared filesystem.
- Physical Volumes
  - Add a definition to cluster nodes
  - Remove a definition from cluster nodes

For more information about performing these tasks, see later sections of this chapter. (See Chapter 29, Troubleshooting HACMP/ES Clusters, for information about using the list commands.)

When you execute a C-SPOC command, the utility determines on which node to perform the operation and then executes the required commands on that node. Typically, C-SPOC executes the command on the node that owns the LVM component (has it varied on). However, you can use C-SPOC commands (on the command line, not from SMIT) on an LVM component that is not currently activated on any cluster node. In this case, C-SPOC determines which node will own the LVM component when it is activated, as specified for the HACMP/ES resource group, and performs the operation on that node.

Understanding C-SPOC and its Relation to Resource Groups

The C-SPOC commands that modify LVM components, such as **cl_chlv**, require that you specify a resource group name as an argument. The LVM component that is the target of the command *must* be configured in the resource group specified. C-SPOC uses the resource group information to determine on which nodes it must execute the operation specified.

To illustrate, consider a cluster with two resource groups defined: resgrp1 and resgrp2. The following summarizes the resource group configuration with filesystems defined as resources and participating nodes:

Resource Group	Filesystem	Participating Nodes
resgrp1	/fs1	Nodes A - G
resgrp2	/fs2	Nodes A - G

Note: In a C-SPOC cluster, any number of nodes can be defined as "participating" in a resource group.

To change the mount point of filesystem **/fs1**, you could enter the following C-SPOC command:

```
cl_chfs -cspoc "-g resgrp1" -m /fsnew /fs1
```

The example specifies resgrp1 in which /fs1 is configured as a resource. However, the command will also succeed if you specify **resgrp2** as the resource group because both resource groups have the same set of participating nodes.

Removing a Filesystem or Logical Volume

When removing a filesystem or logical volume using the C-SPOC **cl_rmfs** and **cl_rmlv** commands, the target filesystem or logical volume must *not* be configured as a resource in the resource group specified. You must unconfigure it from the resource group before removing it.

Updating LVM Components in an HACMP/ES Cluster

When you change the definition of a shared LVM component in a cluster, the operation updates the LVM data that describes the component on the local node and in the Volume Group Descriptor Area (VGDA) on the disks in the volume group. AIX 4.3 LVM enhancements allow all nodes in the cluster to be aware of changes to a volume group, logical volume, and filesystem, at the time the changes are made, rather than waiting for the information to be retrieved during a lazy update.

If for some reason the node is not updated via the C-SPOC enhanced utilities, due to an error condition (a node is down, for example), the volume group will be updated and the change will be taken care of during the “lazy update” mechanism or during execution of the **clvaryonvg** command.

If node failure does occur during a C-SPOC operation an error is displayed to the screen and the error messages are recorded in the C-SPOC error log (**/tmp/cspoc.log** is the default location of this log).

Error Reporting provides detailed information about inconsistency in volume group state across the cluster. If this happens, you must take manual corrective action.

Forcing an Update Before Fallover

In certain circumstances, you may want to update the LVM definition on remote cluster nodes before a fallover occurs. For example, if you rename a logical volume using C-SPOC, the LVM data describing the component is updated on the local node and is updated in the VGDA on the disks, as previously described. If you attempt to rename the logical volume a second time using C-SPOC; the operation will fail if the LVM data on any other cluster node has not been updated.

To update the LVM data on a remote node, use the C-SPOC **/usr/es/sbin/cluster/sbin/cl_updatevg** command. This command causes the specified remote node to export and import the LVM data whether the time-stamp associated with the LVM component is the same or different. For more information about the **cl_updatevg** command, see its man page.

Note: The volume group must be varied off on all nodes accessing the shared LVM component before running the **cl_updatevg** command.

Maintaining Shared Volume Groups

The following administrative tasks involve shared volume groups:

- Creating a shared volume group
- Extending a shared volume group
- Importing a shared volume group
- Reducing a shared volume group
- Making a copy of a volume group
- Removing a copy of a volume group
- Mirroring a volume group
- Unmirroring a volume group
- Removing a shared volume group
- Synchronizing volume group mirrors.

The following section describes how to do these tasks using the standard AIX commands.The next section describes how to do these tasks using C-SPOC.

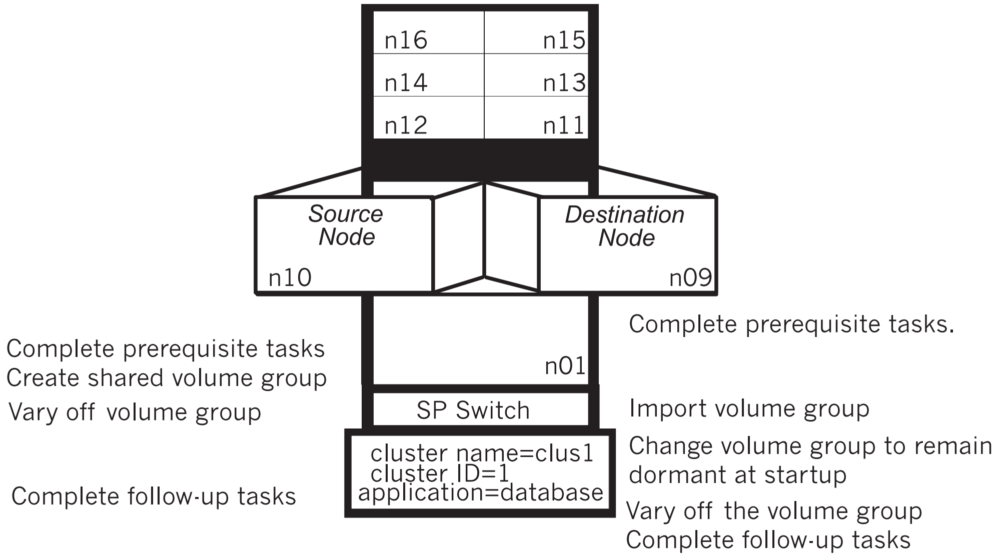
Using C-SPOC simplifies the steps required for all tasks. Moreover, you do not have to stop and restart cluster services in order to do the tasks.

For creating a shared volume group or adding nodes to an existing volume group, you can also use the TaskGuide, a graphical interface that leads you through all of the necessary steps. The TaskGuide can reduce the time and effort of these shared volume group tasks and can help prevent user errors, as it does not allow you to proceed with certain steps if you do not have the necessary configuration in place. For more information on the TaskGuide for creating shared volume groups see Chapter 12, Defining Shared LVM Components.

Using AIX Commands to Maintain Shared Volume Groups

Creating a Shared Volume Group with AIX Commands.

The figure below summarizes the steps you must complete on each cluster node to create a shared volume group. Perform these tasks on each destination node, one node at a time.



Creating a Shared Volume Group

The physical volumes (**hdisks**) should be installed, configured, and available. You can verify the disks’ status with the **lsdev -Cc disk** command.

The following procedure provides more detail about each of the step.

Note: It is essential that you perform all the steps in the correct order so the data does not become corrupted.

1. Complete prerequisite tasks. On the source node, create the volume group, using the SMIT **mkvg** fastpath.

   Enter the specific field values and press Enter. For other fields use the defaults or the appropriate entries for your operation:

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 this field 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 be sure to use the same major number on all nodes in the resource group chain. Use the **lvlstmajor** command on each node to determine a free major number common to all nodes.

2. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.
3. On each destination node, in turn, import the volume group using the SMIT **importvg** fastpath. When you import a volume group, the system copies data from the Volume Group Description Area (VGDA) on a physical volume in the volume group into kernel data structures.

   Enter specific field values and press Enter. For other fields use the defaults or the appropriate entries for your operation:

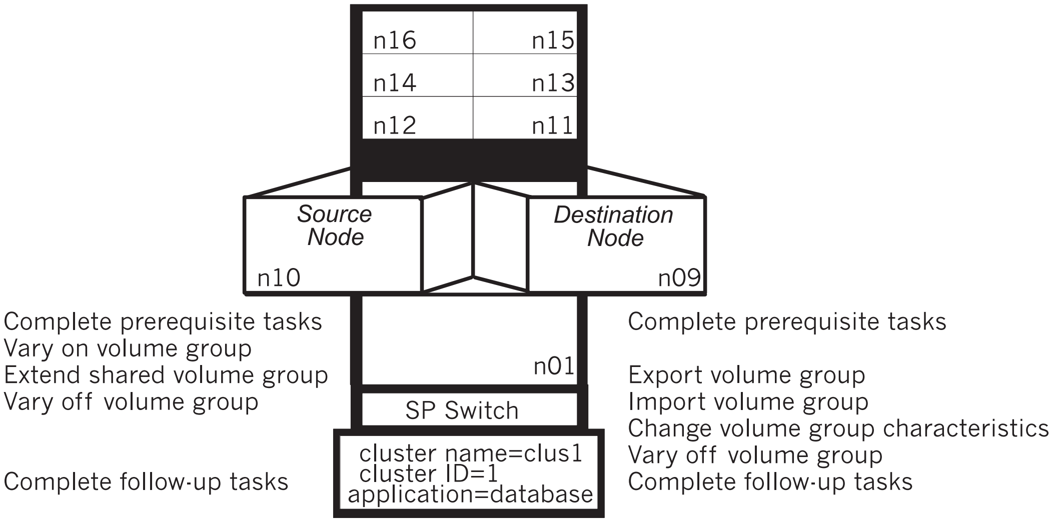
VOLUME GROUP name	Enter the name of the volume group that you are importing. Make sure the volume group name is the same name that you used on the source node.
PHYSICAL VOLUME name	Enter the name of one of the physical volumes that resides in the volume group. *Note that a disk may have a different hdisk 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 **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 be 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.

4. On each destination node, in turn, change the volume group to remain dormant at startup, using the **chvg** command. Set the **Activate volume group AUTOMATICALLY at system restart?** option to **no**.
5. On each destination node, in turn, vary off the volume group, using the SMIT **varyoffvg** fastpath.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Extending a Shared Volume Group with AIX

The figure below summarizes the steps you must complete on all cluster nodes to extend (add one or more physical volumes to) an existing shared volume group.



Extending a Shared Volume Group

Note: The physical volumes (**hdisks**) being added to the volume group must be installed, configured, and available. They must have PVIDs on all nodes that can own the volume group.

The following procedure provides more detail about each step.

1. Complete prerequisite tasks. Stop HACMP/ES cluster services on the nodes in the resource group chain. See Chapter 21, Monitoring an HACMP/ES Cluster, for more information.
2. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
3. On the source node, extend the volume group, using the SMIT **extendvg** fastpath. SMIT displays the Add a Physical Volume to a Volume Group screen.

   Press the F4 key to obtain a list of volume groups from which you can select one. Press the F4 key again to obtain a listing of physical volumes from which you can pick one or more to add to the volume group. Press Enter and AIX adds the physical volumes to the volume group.
4. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.
5. On each destination node, in turn, export the volume group to remove its definition from the system, using the SMIT **exportvg** fastpath.

6. On each destination node, in turn, import the volume group to make it known to the system using the SMIT **importvg** fastpath. On the **Import a Volume Group** SMIT screen, you must enter data for the following fields:

   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 **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 be 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.

7. On each destination node, in turn, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath. Set the **Activate volume group AUTOMATICALLY at system restart?** option to **no**.

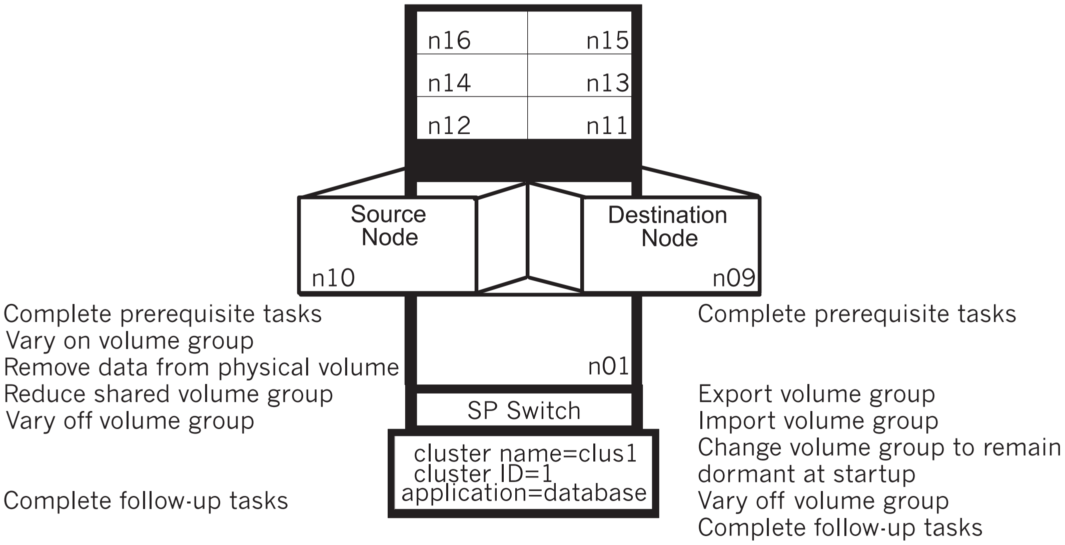
8. On each destination node, in turn, vary off the volume group, using the SMIT **varyoffvg** fastpath.

9. Restart cluster services on all nodes.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Reducing a Shared Volume Group with AIX

The figure below summarizes the steps you must complete on all cluster nodes to reduce (remove one or more physical volumes from) a shared volume group.



Reducing a Shared Volume Group

The following procedure provides more detail about each step.

1. Complete prerequisite tasks. Stop HACMP/ES cluster services on nodes in the resource group chain. See Chapter 20, Starting and Stopping Cluster Services, for more information.
2. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
3. On the source node, move data from the physical volume (or volumes) being removed from the volume group, using the SMIT **migratepv** fastpath. Set the **Move only data belonging to this LOGICAL VOLUME** field to **no**.

   Warning: If you do not perform this step, data will be lost.

4. On the source node, reduce the size of the volume group, using the SMIT **reducevg** fastpath. SMIT displays the Reduce a Volume Group screen.
5. Select the **Remove a Physical Volume from a Volume Group** option and press Enter. SMIT prompts you to choose the volume group you want to reduce. Press the F4 key to obtain a list of volume groups.

   After you select the volume group, SMIT displays the Remove a Physical Volume from a Volume Group screen.

   Specify the names of the physical volumes you want to remove from the volume group. Set the value of the **FORCE deallocation of all partitions on this physical volume** field to **no**.

6. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.
7. On each destination node, in turn, export the volume group to remove its definition from the system, using the SMIT **exportvg** fastpath.

8. On each destination node, in turn, import the volume group to make it known to the system, using the SMIT **importvg** fastpath. 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 **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 be 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.

9. On each destination node, in turn, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath. Set the **Activate volume group AUTOMATICALLY at system restart?** option to no.

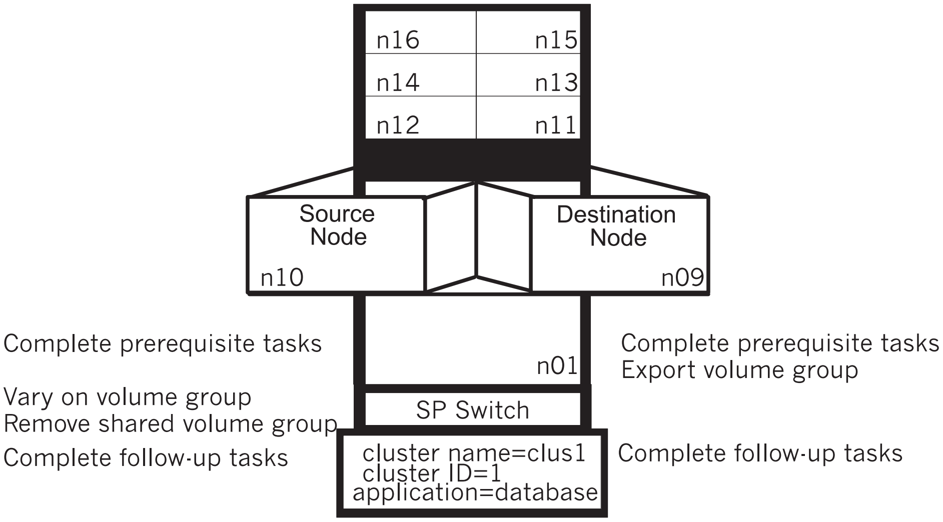
10. On each destination node, in turn, vary off the volume group, using the SMIT **varyoffvg** fastpath.

11. Restart cluster services on all nodes.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Removing a Shared Volume Group with AIX

The figure below summarizes the steps you must complete on all cluster nodes to remove a shared volume group.



Removing a Shared Volume Group

The following procedure provides more detail about each step.

1. Complete prerequisite tasks. Stop HACMP/ES cluster services on nodes in the resource group chain. See Chapter 20, Starting and Stopping Cluster Services, for more information.
2. On each destination node, export the volume group to remove its definition from the system, using the SMIT **exportvg** fastpath.
3. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
4. On the source node, remove the shared volume group by using the SMIT **reducevg** fastpath. SMIT displays the Reduce a Volume Group screen:
5. Select the **Remove a Volume Group** option and press Enter. In the next screen, you must specify the volume group you want to remove. Press the F4 key to obtain a list of volume groups from which to choose. After making your selection, press Enter and AIX deletes the volume group.
6. Vary off the volume group on the source node.
7. Restart cluster services on all nodes on all cluster nodes.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Using C-SPOC to Maintain Shared Volume Groups

You can use C-SPOC to do all the shared volume maintenance tasks.

Creating a Shared Volume Group with C-SPOC

Before creating a shared volume group for the cluster using C-SPOC, check that:

- All disk devices are properly attached to the cluster nodes
- All disk devices are properly configured on all cluster nodes and the device is listed as available on all nodes
- Disks have a PVID.

Take the following steps to create a shared volume group for a selected list of cluster nodes:

1. Enter the fastpath **smitty cl_admin**
2. Select **Cluster Logical Volume Manager** > **Shared Volume Groups** > **Add a Shared Volume Group**.

   SMIT displays a list of cluster nodes.
3. Select two or more nodes from the list and press Enter.

   The system correlates a list of all free physical disks that are available to all nodes selected. (Free disks are those disks that currently are not part of a volume group and have PVIDs.) SMIT displays the list of free physical disks in a multi-pick list by PVIDs.
4. Select one or more disks from the list and press Enter.

   SMIT displays the **Add a Volume Group** screen.
5. Complete the selections as follows and press Enter.

   VOLUME GROUP name — Enter the name for this volume group.

   Physical partition SIZE im megabytes — Accept the default.

   PHYSICAL VOLUME names — Selected physical volumes are listed.

   Volume group MAJOR NUMBER — The system displays the number C-SPOC has determined to be correct.

   Warning: Changing the volume group major number may result in the command's inability to execute on a node that does not have that major number currently available. Please check for a commonly available major number on all nodes before changing this setting.

   C-SPOC verifies communication paths and version compatibility and then executes the command on all nodes in selection

Note: If the major number entered on the SMIT panel was not free at the time that the system attempted to make the volume group the command will display an error for the node that did not complete the execution and continue to the other nodes. At the completion of the command the volume group will not be active on any node in cluster.

Extending a Shared Volume Group with C-SPOC

Take the following steps to add a physical volume to a shared volume group using C-SPOC.

1. The physical volumes (**hdisks**) being added to the volume group must be installed, configured, and available. They must have PVIDs on all nodes that can own the volume group.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management -> Cluster Logical Volume manager > Shared Volume Groups > Set Characteristics of a Shared Volume Group > Add a Physical Volume to a Shared Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. You can pick one or more to add to the volume group. Select the ones you want to add to the shared volume group and Press Enter.
7. SMIT displays the Add a Physical Volume to a Shared Volume Group screen, with the following entries filled in.

Resource Group name The cluster resource group to which this shared volume group belongs.

Volume Group name Name of the shared volume group where hdisks are to be added.

Reference node Name of the node where the hdisks are found.

Physical Volume names Names of the hdisks to be added to the volume group.

8. If this screen reflects the correct information, press Enter to add the disks to the shared volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the shared volume group varied on, vary off the volume group on that node.

Importing a Shared Volume Group with C-SPOC

Take the following steps to import a volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available on all nodes that can own the volume group.

2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Shared Volume Groups > Import a Shared Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the **Import a Shared Volume Group** screen. Values for fields you have selected are displayed. For other fields, use the defaults or the appropriate entries for your operation:

   Resource Group name The cluster resource group to which this shared volume group belongs.

   VOLUME GROUP name The name of the volume group that you are importing.

   PHYSICAL VOLUME name The name of one of the physical volumes that resides in the volume group. This is the hdisk name on the reference node.

   Reference node Node from which the physical disk was retrieved.

   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 be 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.

   Make this VG concurrent capable? The default is **no**. For a non-concurrent VG, leave this **no**.

   Make default varyon of VG Concurrent? The default is **no**. For a non-concurrent VG, leave this **no**.

8. If this screen reflects the correct information, press Enter to import the shared volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the shared volume group varied on, vary off the volume group on that node.

Removing a Physical Volume from a Shared Volume Group with C-SPOC

Take the following steps to remove a physical volume from a shared volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).

3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Shared Volume Groups > Set Characteristics > Remove a Physical Volume from a Shared Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the **Remove a Physical Volume from a Shared Volume Group** screen, with the following entries filled in.:

   Resource Group name The cluster resource group to which this shared volume group belongs.

   VOLUME GROUP name The name of the volume group that you are reducing.

   Reference node Node from which the name of the physical disk was retrieved.

   PHYSICAL VOLUME name The name of the physical volume that you want to remove.This is the hdisk name on the reference node.

8. If this screen reflects the correct information, press Enter to reduce the shared volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the shared volume group varied on, vary off the volume group on that node.

Mirroring a Volume Group Using C-SPOC

Take the following steps to mirror a shared volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Shared Volume Groups > Mirror a Shared Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the **Mirror a Shared Volume Group** screen, with the selected entries filled in. For other fields, use the defaults or the appropriate entries for your operation:

   Resource Group Name The name of the resource group to which this shared volume group belongs is displayed.

   VOLUME GROUP name The name of the volume group that you want to mirror is displayed.

Reference node	Node from which the name of the physical disk was retrieved is displayed.
PHYSICAL VOLUME names	The name of a physical volume on the volume group that you want to mirror.This is the hdisk name on the reference node.
Number of COPIES of each logical partition	The default is **2**. You can also select **3**.
Keep Quorum Checking On?	The default is **no.** You can also choose **yes.**
Create Exact LV Mapping?	The default is **no**.

8. If this screen reflects the correct information, press Enter to mirror the shared volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the shared volume group varied on, vary off the volume group on that node.

Unmirroring a Volume Group Using C-SPOC

Take the following steps to unmirror a shared volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Shared Volume Groups > Unmirror a Shared Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the Unmirror a Shared Volume Group screen, with the chosen fields filled in. For other fields, use the defaults or the appropriate entries for your operation:

Resource Group Name	The name of the resource group to which this shared volume group belongs is displayed.
VOLUME GROUP name	The name of the volume group that you want to mirror is displayed.
Reference node	Node from which the name of the physical disk was retrieved is displayed.
PHYSICAL VOLUME names	The name of a physical volume on the volume group that you want to unmirror.This is the hdisk name on the reference node.

Mirror sync mode — **Foreground** is the default. Other choices are **Background** and **No Sync.**

Number of COPIES of each logical partition — The default is **2**. You can also select **3**.

8. If this screen reflects the correct information, press Enter to unmirror the shared volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the shared volume group varied on, vary off the volume group on that node.

Synchronizing Volume Group Mirrors

Take the following steps to synchronize shared LVM Mirrors by volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Synchronize Shared LVM Mirrors > Synchronize By Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the **Synchronize LVM Mirrors by Volume Group** screen, with the chosen entries filled in. For other fields, use the defaults or the appropriate entries for your operation:

Resource Group Name — The name of the resource group to which this shared volume group belongs is displayed.

VOLUME GROUP name — The name of the volume group that you want to mirror is displayed.

Reference node — Node from which the name of the physical disk was retrieved is displayed.

Number of Partitions to Sync in Parallel — .Leave empty.

Synchronize All Partitions — The default is **no.**

Delay Writes to VG from other cluster nodes during this Sync — The default is **no**.

8. If this screen reflects the correct information, press Enter to synchronize LVM mirrors by the shared volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the shared volume group varied on, vary off the volume group on that node.

Synchronizing a Shared Volume Group Definition

Take the following steps to synchronize a shared volume group definition using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Synchronize A Shared Volume Group Definition** and press Enter.
5. SMIT displays the **Synchronize A Shared Volume Group Definition** screen. Here you must enter the name of the shared volume group to synchronize. Press F4 for a picklist, select the desired volume group and press Enter.
6. The command is run. All nodes in the cluster receive this updated information.
7. If you did this task from a cluster node that does not need the shared volume group varied on, vary off the volume group on that node.

Maintaining Shared Filesystems

The following administrative tasks involve shared filesystems:

- Creating a shared filesystem
- Extending a shared filesystem
- Changing a shared filesystem
- Removing a shared filesystem.

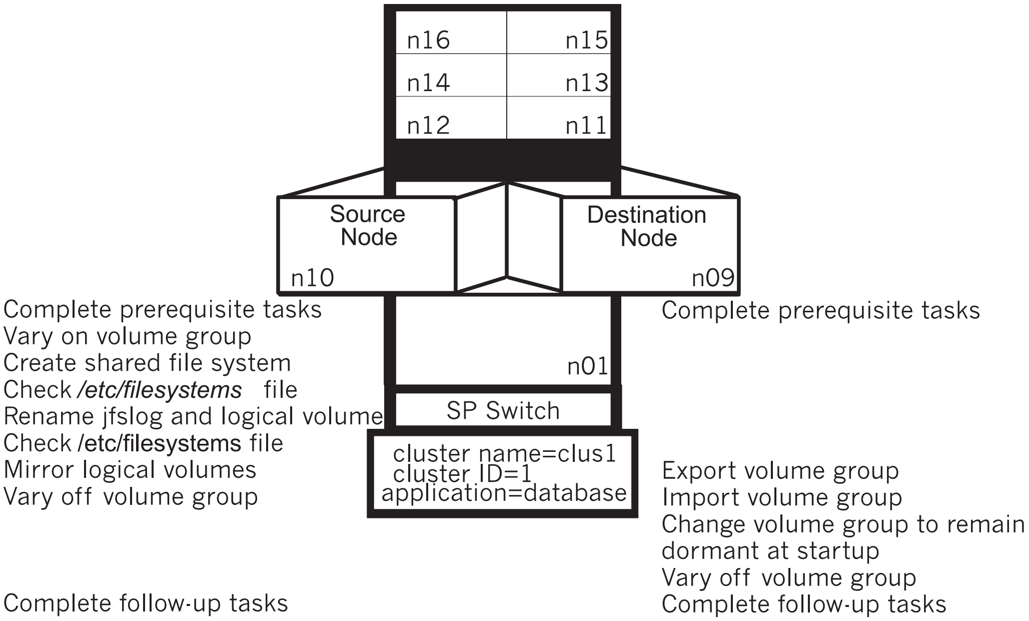
Each operation is described below. The sections also describe how to use the C-SPOC utility to create, change or remove a shared filesystem in a cluster.

Creating a Shared Filesystem Using AIX Commands

This section describes how to use the **smit crjfs** to create a shared journaled filesystem. The **crfs** command also creates the logical volume associated with the journaled filesystem. If you prefer, you can use the **mklv** command to create a logical volume and then use the **mkfs** command to create the filesystem for the logical volume.

Note: You must have previously created the shared volume group that will contain the shared filesystem.

The figure below summarizes the steps you must complete on all cluster nodes to create a shared filesystem.



Creating a Shared Filesystem

The following procedure provides more detail about each step.

1. Complete prerequisite tasks. Stop HACMP/ES cluster services on nodes in the resource group chain. See Chapter 20, Starting and Stopping Cluster Services, for more information.
2. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
3. On the source node, create the filesystem, using the `crjfs` SMIT fastpath. Select **Add a Standard Journaled Filesystem**. SMIT then prompts you to specify the volume group on which you want to create the filesystem.

   After you choose the volume group, SMIT displays the following screen.

   You must specify the size of the filesystem and the name of the mount point. Make sure the **Mount AUTOMATICALLY at system restart?** field is set to **no**.
4. On the source node, verify that the mount attribute for the filesystem in **/etc/filesystems** file is set to false. Setting the **mount** attribute to **false** prevents AIX from mounting the filesystem at boot time.
5. On the source node, rename the jfslog and logical volumes to ensure that they will be consistent across all cluster nodes. If you do not specify names for the jfslog and logical volume, AIX assigns them a name which could be vary on each node.

   Use the **lsvg -l volume_group** command to determine the name of the jfslog and the logical volume associated with the filesystem. In the output from this command, look for the logical volume with the type jfs and the log logical volume that has type jfslog.

   Use the SMIT **chlv** fastpath to rename these logical volumes. Use a naming convention that makes the relationship of the logical volumes clear.

6. On the source node, check the dev and log Attributes in the **/etc/filesystems** file to make sure they reflect the change.

7. On the source node, add copies to logical volume, using the SMIT **mklvcopy** fastpath. 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.

   Note: This step does not apply to RAID devices using RAID levels 3 or 5. They provide their own data redundancy.

   Enter the specific field values as follows and press Enter to create the copies. For other fields use the defaults or the appropriate entries for your operation:

Allocate each logical partition copy on a SEPARATE physical volume?	Specify **yes**. Doing so ensures that copies reside on separate disks.
SYNCHRONIZE the data in the new logical partition copies	Specify **yes**.

   - To 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.

   - 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.

8. On the source node, run a consistency check on each filesystem. Enter:

   ```
   fsck /filesystem_name
   ```

   Verify that you can mount the filesystem by entering:

   ```
   mount /filesystem_name
   ```

   Verify that you can unmount the filesystem by entering:

   ```
   unmount /filesystem_name
   ```

9. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.

10. On each destination node, in turn, export the volume group to remove its definition from the system, using the SMIT **exportvg** fastpath.

11. On the destination nodes, import the volume group to make it known to the system, using the SMIT **importvg** fastpath. 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 **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 be 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.

12. On the destination nodes, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath. Set the **Activate volume group AUTOMATICALLY at system restart?** option to **no**.
13. On the destination nodes, vary off the volume group, using the SMIT **varyoffvg** fastpath.
14. Restart cluster services on all nodes.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Creating a Shared Filesystem on a C-SPOC Cluster

Before creating a journaled filesystem for the cluster using C-SPOC, check that:

- All disk devices are properly attached to the cluster nodes
- All disk devices are properly configured and available on all cluster nodes
- The volume group that will contain the filesystem must be varied on on at least one cluster node.

You can add a journaled filesystem to either:

- A shared volume group (no previously defined cluster logical volume)
- A previously defined cluster logical volume (on a shared volume group).

To add a filesystem where no logical volume is currently defined:

1. Enter the fastpath **smitty cl_admin**
2. Select **Cluster Logical Volume Manager** > **Shared Filesystems** > **Add a Journaled Filesystem**.

   SMIT displays a list of filesystem types (Standard, Compressed or Large File Enabled).
3. Select the desired filesystem type from the list.

   SMIT generates a list of all volume groups in the cluster.
4. Select the volume group where the filesystem will be added.

   SMIT displays the AIX SMIT screen for selecting filesystem attributes.

Node Names	The names of the selected cluster nodes are displayed.
Volume Group Name	The selected volume group name is displayed.
MOUNT POINT	Enter the mount point for the filesystem.
PERMISSIONS	Set as desired.
Mount OPTIONS	Set as desired.
Start Disk Accounting?	Set as desired. Default is **no**.
Fragment Size (Bytes)	4096 is the default.
Number of Bytes per node	4096 is the default.
Compression algorithm	Default is **no**.

5. Select the filesystem attributes and press Enter.

   SMIT checks the node list for the resource group that contains the volume group, creates the logical volume (on an existing log logical volume if present, otherwise it will create a new log logical volume) and adds the filesystem to the node where the volume group is varied on. All other nodes in the resource group will run an **importvg -L**.

Take the following steps to add a filesystem to a previously defined cluster logical volume:

1. Enter the fastpath **smitty cl_admin**
2. Select **Cluster Logical Volume Manager** > **Shared Filesystems** > **Add a Journaled Filesystem to a Previously Defined Logical Volume**.

   SMIT displays a list of filesystem types (Standard, Compressed or Large File Enabled).
3. Select the desired filesystem type from the list.

   SMIT generates a list of all free logical volumes in the cluster and nodes they are on. SMIT reports a logical volume as free if:

   - the logical volume is part of a parent volume group that is configured as a resource in the cluster

- the logical volume is varied on prior to and during the system polling the disk for logical volume information
- the logical volume does not have a filesystem mount point.

4. Select a logical volume where the filesystems will be added.

   SMIT displays the AIX SMIT screen for selecting filesystem attributes.

LOGICAL VOLUME name	The name of the selected logical volume is displayed.
Nodes	The names of the selected cluster nodes are displayed.
MOUNT POINT	Enter the mount point for the filesystem.
PERMISSIONS	Set as desired.
Mount OPTIONS	Set as desired.
Start Disk Accounting?	Set as desired. Default is **no**.
Fragment Size (Bytes)	4096 is the default.
Number of Bytes per inode	4096 is the default.
Compression algorithm	Default is **no**.

5. Select the filesystem attributes and press Enter.

   SMIT checks the node list for the resource group that contains the volume group where the logical volume is located and adds the filesystem to the node where the volume group is varied on. All other nodes in the resource group will run an **importvg -L**.

Changing a Shared Filesystem in a Cluster

As system administrator of an HACMP/ES cluster, you may need to change the characteristics of an existing filesystem. The following sections describe how to perform this task in a cluster using the C-SPOC utility, and on a cluster of any size using standard AIX commands.

Changing a Filesystem on a C-SPOC Cluster

Using the C-SPOC utility, you can change the characteristics of a shared filesystem on cluster nodes by executing a command on a single cluster node. The C-SPOC command changes the attributes of the shared filesystem on the node which currently has the shared volume group varied on. The definition of the filesystem on other cluster nodes is not updated immediately; it is updated when the filesystem is activated on these nodes.

To change the characteristics of a shared filesystem:

1. Vary on the volume group, if needed, by using the **lsvg** command. You can use the C-SPOC utility to change a filesystem even if the volume group on which it is defined is varied off. In this case, specify the **-f** flag. See the **cl_chfs** man page.
2. Use C-SPOC to change attributes of the filesystem. Enter the SMIT fastpath

```
smit hacmp
```

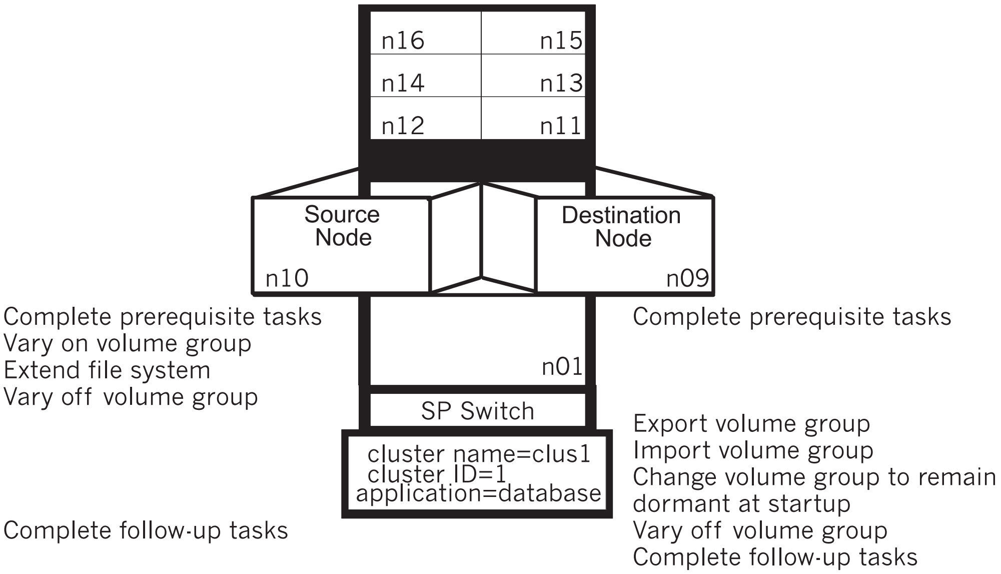
3. Select the **Cluster System Management** option and press Enter. (You can also use the SMIT fastpath `cl_admin` to get to this screen directly from the command line.)
4. Select the **Cluster Logical Volume Manager** option and press Enter.
5. Choose the **Shared Filesystems** option and press Enter.
6. Select the **Change/Show Characteristics of a Shared Filesystem in the Cluster** option and press Enter. SMIT displays a pick list of existing filesystems from which you must select one. After you select the filesystem, SMIT displays a screen containing the characteristics of the filesystem you can change using the C-SPOC utility.
7. Enter data in the fields you want to change and press Enter. The C-SPOC utility changes the filesystem characteristics on the local node. The filesystem characteristics are not updated on remote nodes until the volume group is activated on each node.

   To check the status of the C-SPOC command execution on cluster nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Changing a Shared Filesystem Using AIX Commands

To change the characteristics of a filesystem on a cluster larger than two nodes, you must make the changes on one cluster node, the source node, and then import the changed filesystems on all the other cluster nodes. The figure below summarizes the steps you must complete on all cluster nodes to change a shared filesystem.



Changing a Shared Filesystem on Cluster Larger Than Two Nodes

The following procedure provides more detail about each step.

1. Complete prerequisite tasks. Stop HACMP/ES cluster services on nodes in the resource group chain. See Chapter 20, Starting and Stopping Cluster Services, for more information.

2. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
3. On the source node, change the characteristics of the shared filesystem, using the SMIT **chjfs** fastpath.

   Enter data in the fields you want to change and press Enter.
4. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.
5. On each destination node, in turn, export the volume group to remove its definition from the system, using the SMIT **exportvg** fastpath.
6. On the destination nodes, import the volume group to make it known to the system, using the SMIT **importvg** fastpath. 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 **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 be 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.
7. On the destination nodes, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath. Set the **Activate volume group AUTOMATICALLY at system restart?** option to **no**.
8. On the destination nodes, vary off the volume group, using the SMIT **varyoffvg** fastpath.
9. Restart cluster services on all nodes.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Removing a Shared Filesystem

As system administrator of an HACMP/ES cluster, you may need to remove a filesystem. You can optionally remove the filesystem's mount point as part of the same operation. The following sections describe how to perform this task on a cluster using the C-SPOC utility, and on any cluster using standard AIX commands.

Removing a Shared Filesystem in a C-SPOC Cluster

Using the following procedure, you can remove a shared filesystem on any node in a cluster by executing a C-SPOC command on that node. The C-SPOC command deletes the shared filesystem on the node which currently has the shared volume group varied on. The command removes both the shared logical volume on which the filesystem resides and the associated stanza in the **/etc/filesystems** file. The filesystem is removed and the stanza removed from **/etc/filesystems** on other cluster nodes when the filesystem is activated on each node.

The filesystem being deleted must not be configured as a resource in the resource group specified as an argument in the C-SPOC command line.

1. On the source node, vary on the volume group, if needed, using the SMIT **varyonvg** fastpath. You can use the C-SPOC utility on a volume group that is varied off; however, you must specify the **-f** flag. See the **cl_rmfs** man page for more information.
2. Use C-SPOC to delete the filesystem. Enter the SMIT fastpath:

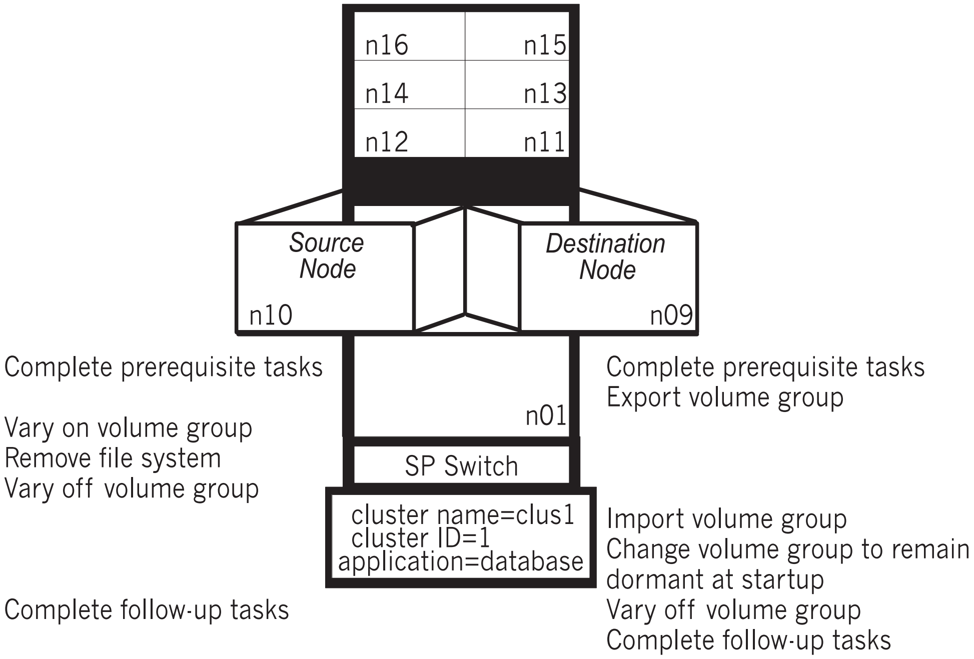
   ```
   smit hacmp
   ```
3. Select the **Cluster System Management** option and press Enter. (You can also use the SMIT fastpath `cl_admin` to get to this screen directly from the command line.)
4. Select the **Cluster Logical Volume Manager** option and press Enter.
5. Choose the **Shared Filesystems** option and press Enter.
6. Select the **Remove a Shared Filesystem** option and press Enter. SMIT displays the following screen.
7. Press the F4 key to obtain a pick list of existing filesystems from which you may select one. Toggle the **Remove Mount Point** option to **yes** if you want to remove the mount point in the same operation. When you finish entering data, press Enter. The C-SPOC utility removes the filesystem on the source (local) node. The filesystem is not removed on remote nodes until the volume group on which the filesystem is defined is activated.

   To check the status of the C-SPOC command execution on both nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Removing a Shared Filesystem Using Standard AIX Commands

The figure below summarizes the steps you must complete on the source and destination nodes to remove a shared filesystem.



Removing a Shared Filesystem

The following procedure provides more detail about each step.

1. Stop HACMP/ES cluster services on nodes in the resource group chain. See Chapter 20, Starting and Stopping Cluster Services, for more information.
2. On the destination nodes, export the volume group to remove its definition from the system, using the SMIT **exportvg** fastpath.
3. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
4. On the source node, delete the shared filesystem, using the SMIT **rmjfs** fastpath.

   Press the F4 key to obtain a pick list of existing filesystems from which you may select one. Toggle the **Remove Mount Point** option to **yes** if you want to remove the mount point in the same operation. When you finish entering data, press Enter.
5. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.
6. On the destination nodes, import the volume group to make it known to the system, using the SMIT **importvg** fastpath. 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 **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 be 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.

7. On the destination nodes, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath. Set the **Activate volume group AUTOMATICALLY at system restart?** option to **no**.
8. On the destination nodes, vary off the volume group, using the SMIT **varyoffvg** fastpath.
9. Restart cluster services on all nodes.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Maintaining Logical Volumes

The following administrative tasks involve shared logical volumes. You can perform all these tasks using AIX commands, or using the C-SPOC utility (except for creating a shared logical volume):

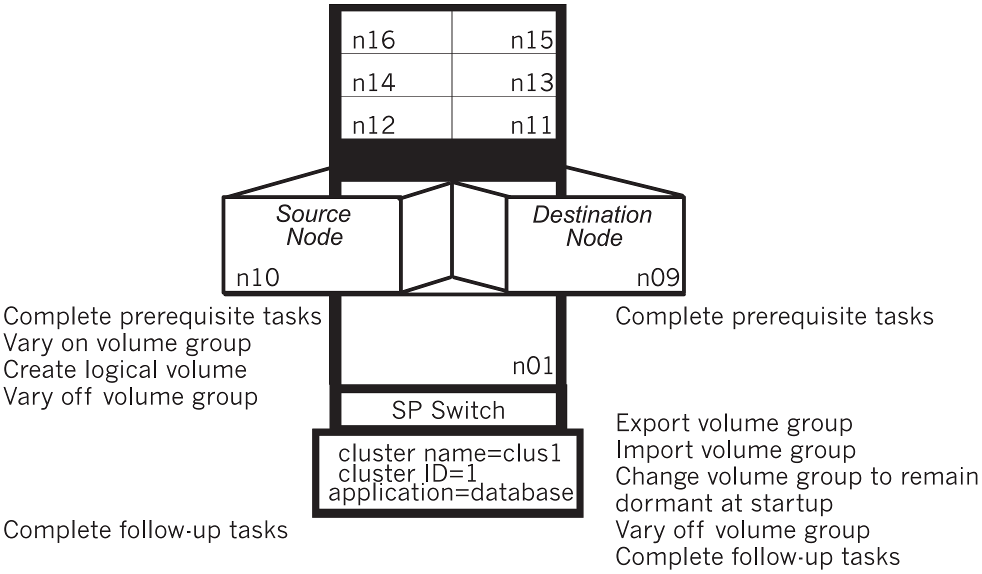
- Creating a shared logical volume
- Changing a shared logical volume (renaming, extending, adding or removing copy)
- Adding or removing a shared logical volume

Note: Regarding increasing or decreasing the number of copies (mirrors) of a shared logical volume: *This task does not apply to RAID devices.*

Creating a Logical Volume

The figure below summarizes the steps you must complete on the source and destination nodes to create a logical volume.

Note: The volume group on which you are creating the logical volume must already exist.



Creating a Logical Volume

The following procedure provides more detail about each step.

1. Stop HACMP/ES cluster services on nodes in the resource group chain. See the section Stopping Cluster Services on page 20-11 for more information.
2. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
3. On the source node, create the logical volume, using the SMIT **mklv** fastpath.

4. Enter the specific field values as follows. For other fields use the defaults or the appropriate entries for your operation:

   Logical Volume NAME Enter the name of the logical volume. Make sure that the name you assign to this logical volume is unique within the cluster.

   Number of COPIES of each logical partition Enter the number of logical volume copies (mirrors). Entering a value of three eliminates the physical volume as a single point of failure.

   If using a RAID device, set this field to 1. These devices provide their own data redundancy if you use RAID level 1, 3, or 5 (not 0).

   Mirror Write Consistency Set this field to **yes**, for nonconcurrent environments, or to **no** for concurrent environments.

   Allocate each logical partition copy on a SEPARATE physical volume Make sure this field is set to **yes**. (Does not apply to RAID devices.)

   ENABLE BAD BLOCK relocation Set this field to **no** to disable bad block relocation (applies to RAID devices).

5. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.

6. On the destination nodes, export the volume group, using the SMIT **exportvg** fastpath.

7. On the destination nodes, import the volume group, using the SMIT **importvg** fastpath. 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 **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 be 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.

8. On the destination nodes, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath.
9. On the destination nodes, vary off the volume group, using the SMIT **varyoffvg** fastpath.
10. Restart cluster services on all nodes.

Adding a Logical Volume to a Cluster Using C-SPOC

To add a logical volume to a cluster using C-SPOC, take the following steps:

1. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath. You can add a logical volume using C-SPOC when the volume group is varied off; however, you must specify the **cspoc -f** flag. See the **cl_mklv** command man page for more information.
2. Enter the SMIT C-SPOC fastpath:

   ```
   smit cl_admin
   ```

3. Select these options **Cluster Logical Volume Manager** >**Shared Logical Volumes** > **Add a Shared Logical Volume** and press Enter.
4. SMIT displays a popup list of shared volume groups. Select one and press Enter.
5. SMIT displays a list of physical volumes. Select one and press Enter. The **Add a Shared Logical Volume** screen appears, with chosen fields filled in as shown in the sample below.:

Resource Group name	casc_rg
VOLUME GROUP name	sharedvg
Reference node	a1
Number of LOGICAL PARTITIONS	[]
PHYSICAL VOLUME names	hdisk16
Logical volume NAME	[]
Logical volume TYPE	[]
POSITION on physical volume	middle
RANGE of physical volumes	minimum
MAXIMUM NUMBER of PHYSICAL VOLUMES to use for allocation	[]
Number of COPIES of each logical partition	1
Mirror Write Consistency?	yes
Allocate each logical partition copy on a SEPARATE physical volume?	yes
RELOCATE the logical volume during reorganization	yes

Logical volume LABEL	[]
MAXIMUM NUMBER of LOGICAL PARTITIONS	[512]
Enable BAD BLOCK relocation?	yes
SCHEDULING POLICY for writing logical partition copies	parallel
Enable WRITE VERIFY?	no
File containing ALLOCATION MAP	[]
Stripe Size?	[Not Striped]

6. The default LV characteristics are most common. Make changes if necessary for your system and press Enter. Other cluster nodes are updated with this information.

Changing the Characteristics of a Shared Logical Volume

As system administrator of an HACMP/ES cluster, you may need to change the characteristics of an existing logical volume. The following sections describe how to perform this task on a cluster both by using the C-SPOC utility and by using standard AIX commands.

Changing a Shared Logical Volume on a C-SPOC Cluster

Using the following procedure, you can change the characteristics of a shared logical volume on all nodes in a cluster by executing a C-SPOC command on any node.

1. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath. You can change a logical volume using C-SPOC when the volume group is varied off; however, you must specify the **cspoc -f** flag. See the **cl_chlv** command man page for more information.
2. Use the C-SPOC utility to change the logical volume. Enter the SMIT fastpath:

   ```
   smit cl_admin
   ```

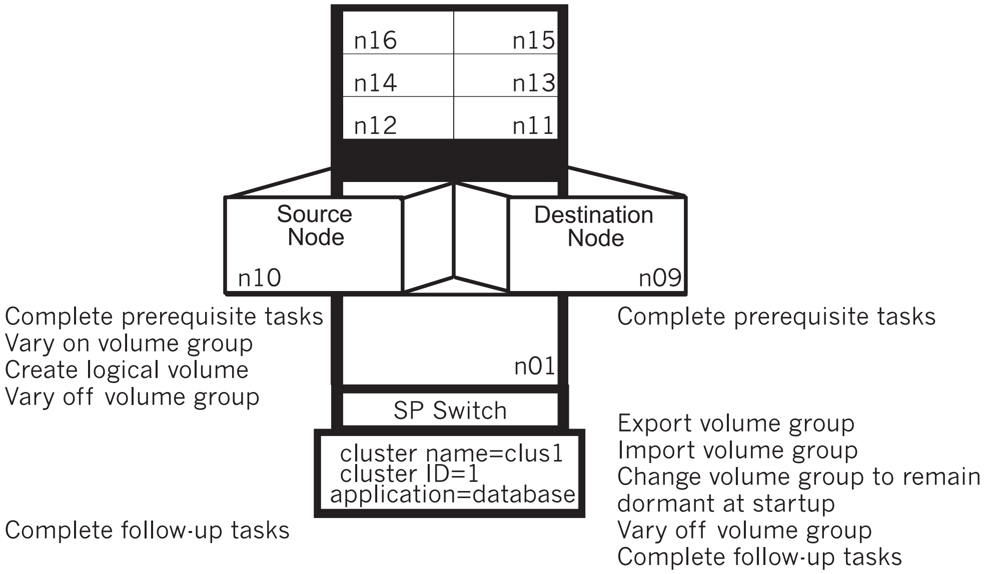
3. Select the **Cluster Logical Volume Manager** > **Shared Logical Volumes > Change a Shared Logical Volume** option and press Enter. SMIT displays a pick list of existing logical volumes.
4. Select the one you want. SMIT displays the screen, with the values of the selected logical volume attributes filled in.
5. Enter data in the fields you want to change and press Enter. The C-SPOC utility changes the characteristics on the local node. The logical volume definition is updated on remote nodes.

   To check the status of the C-SPOC command execution on all nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

   Caution: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Changing a Shared Logical Volume Using AIX Commands

To change the characteristics of a shared logical volume on large clusters, you must make the changes on one cluster node, the source node, and then import the changes on all other cluster nodes. The figure below summarizes the steps you must complete on all cluster nodes.



Changing a Shared Logical Volume on Cluster Larger Than Two Nodes

The following procedure provides more detail about each step.

1. Complete prerequisite tasks. Stop HACMP/ES cluster services on nodes in the resource group chain. See the section Stopping Cluster Services on page 20-11 for more information.
2. On the source node, vary on the volume group, if necessary, using the SMIT **varyonvg** fastpath.
3. On the source node, change the characteristics of the shared logical volume, using the SMIT **chlv** fastpath.
4. Select the **Change a Logical Volume** option and press Enter. SMIT displays the **Change a Logical Volume on the Cluster** screen.
5. Press the F4 key to obtain a pick list of existing logical volumes from which you may select one. When you finish entering data, press Enter. SMIT displays the screen with the values of the current logical volume attributes filled in.
6. Enter data in the fields you want to change and press Enter. The C-SPOC utility changes the filesystem characteristics on the local node.

7. On the destination nodes, import the volume group to make it known to the system, using the SMIT **importvg** fastpath. 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 **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 be 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.

8. On the destination nodes, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath.
9. On the destination nodes, vary off the volume group, using the SMIT **varyoffvg** fastpath.
10. Restart cluster services on all nodes.

Warning: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Removing a Logical Volume

As system administrator of an HACMP/ES cluster, you may need to remove a logical volume. The following sections describe how to perform this task on a cluster using the C-SPOC utility, and on any cluster using standard AIX commands.

Removing a Logical Volume on a C-SPOC Cluster

Using the following procedure, you can remove a logical volume on any node in a cluster by executing a C-SPOC command on one of the nodes.

1. On a cluster node, vary on the volume group, if needed, using the SMIT **varyonvg** fastpath. You can remove a logical volume using C-SPOC when the volume group is varied off; however, you must specify the **-f** flag. See the **cl_rmlv** command man page for more information.

Note: If the logical volume to be removed contains a filesystem, you first must remove the filesystem from any specified resource group before attempting to remove the logical volume. Afterwards, be sure to synchronize cluster resources on all cluster nodes.

2. Use the C-SPOC utility to change the logical volume. Enter the SMIT fastpath:

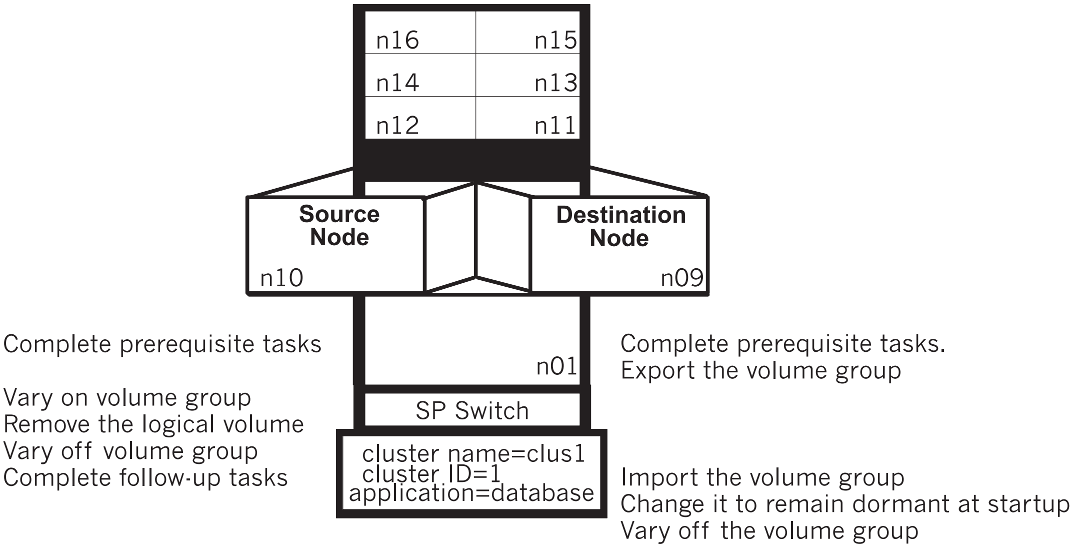
```
smit cl_admin
```

3. Select the **Cluster Logical Volume Manager** option and press Enter.
4. Choose the **Shared Logical Volumes** option and press Enter.
5. Select the **Remove a Shared Logical Volume** option and press Enter.
6. C-SPOC provides a list of shared logical volumes, organized by HACMP/ES resource group. Select the logical volume you want to remove and press Enter. Remote nodes are updated.

   To check the status of the C-SPOC command execution on other cluster nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Removing a Logical Volume on Clusters Using AIX Commands

To remove a logical volume, you must remove the logical volume on one cluster node, the source node, and then import the volume group on which the logical volume is defined on all the other cluster nodes. The figure below summarizes the steps you must complete on the source and destination nodes to remove a logical volume.



Removing a Logical Volume

1. Complete prerequisite tasks. Stop HACMP/ES cluster services on nodes in the resource group chain. See the section Stopping Cluster Services on page 20-11 for more information.
2. On the destination nodes, export the volume group to remove the definition from the system, using the SMIT **exportvg** fastpath.
3. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.

4. On the source node, remove the shared logical volume, using the SMIT **rmlv** fastpath. You can accept the defaults for all fields.
5. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.
6. On the destination nodes, import the volume group to make it known to the system, using the SMIT **importvg** fastpath. 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 Set the field 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 be 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.

7. On the destination nodes, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath.
8. On the destination nodes, vary off the volume group, using the SMIT **varyoffvg** fastpath.
9. Restart cluster services on all nodes.

After completing this procedure, verify fallover behavior.

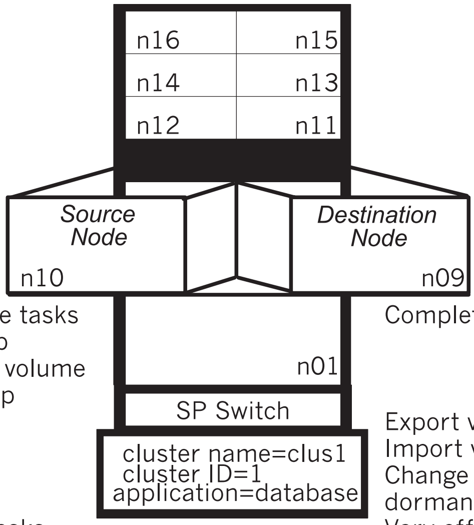
Adding Copies (Mirrors) to a Shared Logical Volume

You should use at least two copies–and preferably three–per logical volume, each on a separate physical volume, to maintain high availability.

Note: This task does not apply to RAID devices, which provide their own data redundancy if you use RAID level 1, 3, or 5 (not 0). Also, you should have created the logical volume to which you are adding copies.

Adding Copies to a Shared Logical Volume Using AIX Commands

The figure below summarizes the steps you must complete on the source and destination nodes to add copies to a logical volume.



Complete prerequisite tasks
Vary on volume group
Add copies to logical volume
Vary off volume group

Complete follow-up tasks

Complete prerequisite tasks

Export volume group
Import volume group
Change volume group to remain dormant at startup
Vary off volume group
Complete follow-up tasks

Mirroring a Logical Volume

The following procedure provides more detail about each step.

1. Stop HACMP/ES cluster services on nodes in the resource group chain. See the section Stopping Cluster Services on page 20-11 for more information.
2. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
3. On the source node, add the mirror copies to the logical volume, using the SMIT **mklvcopy** fastpath. Enter the specific field values as follows. For other fields use the defaults or the appropriate entries for your operation:

NEW TOTAL number of logical partition copies
Specify the total number of copies. As many as three copies are allowed, and three is the recommended number.

PHYSICAL VOLUME names
Specify the physical disk volumes on which the logical volume copies will reside. Press F4 to list the disks available in the volume group.

The copies, if possible, should also reside on disks that are controlled by different disk adapters and are located in separate drawers or units. Because copies should reside on separate disks, you should specify as many disks as there are copies.

Allocate each logical partition copy on a SEPARATE physical volume?
Specify **yes**. Doing so ensures that copies reside on separate disks.

Mirror Write Consistency	Set this field to **yes.**
ENABLE BAD BLOCK relocation	Set this field to **no** to disable bad block relocation (applies to RAID devices).
SYNCHRONIZE the data in the new logical partition copies	Specify **yes**.

4. Verify the creation of the logical volume copies, using the **lsvg -l volume_group** command. In the output from the command, check 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. To check that the copies were created on the correct physical volumes, use the **lspv -l hdiskx** command.

5. Test the filesystem (if the logical volume contains a filesystem) by running a consistency check on each filesystem. Enter:

   ```
   fsck /filesystem_name
   ```

   Verify that you can mount the filesystem by entering:

   ```
   mount /filesystem_name
   ```

   Verify that you can unmount the filesystem by entering:

   ```
   unmount /filesystem_name
   ```

6. On the destination nodes, export the volume group to remove its definition from the system, using the SMIT **exportvg** fastpath.

7. On the destination nodes, import the volume group to make it known to the system, using the SMIT **importvg** fastpath. 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 **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 be 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.

8. On the destination nodes, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath.
9. On the destination nodes, vary off the volume group, using the SMIT **varyoffvg** fastpath.
10. Restart cluster services on all nodes.

After completing the procedure, verify fallover behavior.

Setting Characteristics of a Shared Logical Volume Using C-SPOC

You can use C-SPOC to do the following tasks for all cluster nodes from one node:

- Rename a shared logical volume
- Increase the size of a shared logical volume
- Add copies to a shared logical volume
- Remove copies from a shared logical volume.

Renaming a Shared Logical Volume Using C-SPOC

Using the following procedure, you can rename a shared logical volume on all nodes in a cluster by executing a C-SPOC command on any node.

1. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath. You can rename a logical volume using C-SPOC when the volume group is varied off; however, you must specify the **cspoc -f** flag. See the **cl_chlv** command man page for more information.
2. Use the C-SPOC utility to rename the logical volume. Enter the SMIT fastpath:

   ```
   smit cl_admin
   ```

3. Select these options: **Cluster Logical Volume Manager** >**Shared Logical Volumes >Set Characteristics of A Shared Logical Volume > Rename a Shared Logical Volume** and press Enter.

   SMIT displays the **Rename a Logical Volume on the Cluster** screen.

4. Press Enter to obtain a pick list of existing logical volumes; select one and press Enter. SMIT displays a screen with the **Resource group name** and **Current logical volume name** filled in.
5. Enter the new name in the **NEW logical volume name** field and press Enter. The C-SPOC utility changes the name on all cluster nodes.

   Caution: After completing this procedure, confirm your changes by initiating failures and verifying correct fallover behavior before resuming normal user operations.

Increasing the Size of a Shared Logical Volume Using C-SPOC

Using the following procedure, you can increase the size of a shared logical volume on all nodes in a cluster by executing a C-SPOC command on any node.

1. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath. You can rename a logical volume using C-SPOC when the volume group is varied off; however, you must specify the **cspoc -f** flag. See the **cl_chlv** command man page for more information.

2. Enter the C-SPOC SMIT fastpath:

   ```
   smit cl_admin
   ```

3. Select these options: **Cluster Logical Volume Manager** >**Shared Logical Volumes >Set Characteristics of A Shared Logical Volume > Increase Size of a Shared Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.

4. Choose the desired logical volume from the pick list and press Enter. SMIT displays a list of physical volumes.
5. Choose a physical volume and press Enter. SMIT displays the **Increase Size of a Shared Logical Volume** screen with the Resource Group, Logical Volume, Reference Node and default fields filled.
6. Enter the new size in the **Number of ADDITIONAL logical partitions** field and press Enter. The C-SPOC utility changes the size of this logical volume.on all cluster nodes.

Adding a Copy to a Shared Logical Volume Using C-SPOC

Using the following procedure, you can add a copy to a shared logical volume on all nodes in a cluster by executing a C-SPOC command on any node.

1. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath. You can rename a logical volume using C-SPOC when the volume group is varied off; however, you must specify the **cspoc -f** flag. See the **cl_chlv** command man page for more information.
2. Enter the C-SPOC SMIT fastpath:

   ```
   smit cl_admin
   ```

3. Select these options: **Cluster Logical Volume Manager** > **Shared Logical Volumes >Set Characteristics of A Shared Logical Volume > Add a Copy to a Shared Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.

4. Choose the desired logical volume from the pick list and press Enter. SMIT displays a list of physical volumes.
5. Choose a physical volume and press Enter. SMIT displays the **Add a Copy to a Shared Logical Volume** screen with the Resource Group, Logical Volume, Reference Node and default fields filled.
6. Enter the new number of mirrors in the **NEW TOTAL number of logical partitions** field and press Enter. The C-SPOC utility changes the number of copies of this logical volume on all cluster nodes.

Removing a Copy from a Shared Logical Volume Using C-SPOC

Using the following procedure, you can remove a copy of a shared logical volume on all nodes in a cluster by executing a C-SPOC command on any node.

1. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath. You can rename a logical volume using C-SPOC when the volume group is varied off; however, you must specify the **cspoc -f** flag. See the **cl_chlv** command man page for more information.

2. Enter the C-SPOC SMIT fastpath:

   ```
   smit cl_admin
   ```

3. Select these options: **Cluster Logical Volume Manager** >**Shared Logical Volumes >Set Characteristics of A Shared Logical Volume > Remove a Copy from a Shared Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.

4. Choose the desired logical volume from the pick list and press Enter. SMIT displays a list of physical volumes.

5. Choose the physical volumes from which you want to remove copies and press Enter. SMIT displays the **Remove a Copy from a Shared Logical Volume** screen with the Resource Group, Logical Volume name, Reference Node and Physical Volume names fields filled in.

6. Enter the new number of mirrors in the **NEW maximum number of logical partitions copies** field and check the **PHYSICAL VOLUME name(s) to remove copies from** field to make sure it is correct and press Enter. The C-SPOC utility changes the number of copies of this logical volume on all cluster nodes.

   To check the status of the C-SPOC command execution on all nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Synchronizing LVM Mirrors by Logical Volume

Take the following steps to synchronize shared LVM Mirrors by logical volume using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.

2. On any cluster node that can own the shared volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).

3. On the source node, enter **smit hacmp**.

4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Synchronize Shared LVM Mirrors > Synchronize By Logical Volume**.

5. SMIT displays a list of logical volumes. Select the desired logical volume and press Enter.

6. SMIT displays a list of physical volumes. Pick one and Press Enter.

7. SMIT displays the Synchronize LVM Mirrors by Volume Group screen, with the chosen entries filled in. For other fields, use the defaults or the appropriate entries for your operation:

   Resource Group Name The name of the resource group to which this logical volume belongs is displayed.

   LOGICAL VOLUME name The name of the logical volume that you want to synchronize is displayed.

   Reference node Node from which the name of the physical disk was retrieved is displayed.

Number of Partitions to Sync in Parallel	.Leave empty.
Synchronize All Partitions	The default is **no.**
Delay Writes to VG from other cluster nodes during this Sync	The default is **no**.

8. If this screen reflects the correct information, press Enter to synchronize LVM mirrors by the shared logical volume. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the shared volume group varied on, vary off the volume group on that node.

Maintaining Physical Volumes

Administrative tasks that involve shared physical volumes include:

- Adding a disk definition to cluster nodes
- Removing a disk definition from cluster nodes
- Moving (migrating) the data on one physical volume to other physical volumes within the shared volume group.

Adding a Disk Definition to Cluster Nodes Using C-SPOC

The nodes must be configured as part of the cluster. Only SSA and SCSI disks are currently supported for this utility.

Take the following steps to add a raw disk on selected cluster nodes:

1. Enter the fastpath **smitty cl_admin**
2. Select **Cluster Disk Manager** > **Add a Disk to the Cluster**.

   SMIT displays a list of nodes in the cluster and prompts you to select the nodes where the disk definition should be added
3. Select one or more nodes where you want to have the new disk configured.

   The system generates a list of available disk types based on those disk types known to the first node in the list (above).
4. Select the type of disk you want to add to the cluster.

   The set of screens that follow depend on the disk type selected. Possible disk types include:

   - SCSI (various types listed)
   - SSA (available types listed)

Adding a SCSI Disk

If you are select a SCSI disk type to define, SMIT displays a multi-select list of parent adapter/node name pairs. You are prompted to select one parent adapter per node.

Select a parent adapter for each node connected to the disk and accept the selection.

SMIT displays the AIX SCSI **Add a Disk** screen with all entries filled in except Connection Address.

Disk type	1000mb
Disk interface	scsi
Description	1.0GB SCSI Disk Drive
Parent adapter	nodea:vscsi0,nodeb:vscsi1,...etc.
CONNECTION address	Use F4 to display the list; select the address.
ASSIGN physical volume identifier	**no** is the default.

Select the connection address and press Enter.

C-SPOC executes the necessary commands to define the disk on all selected nodes.

Adding an SSA Disk

If you select an SSA disk to define, SMIT displays the AIX **Add an SSA Logical Disk** screen.

Disk type	hdisk
Disk interface	ssa
Description	SSA Logical Disk Drive
Parent	ssar
CONNECTION address	Press F4 for a list; add the address.
Location Label	(Optional entry). Press F1 for information.
ASSIGN physical volume identifier	**no** is the default.
RESERVE disk on open	**yes** is the default.
Queue depth	(Optional entry). Press F3 for information on ranges allowed. Press F1 for help.
Maximum coalesce	(Optional entry). Press F3 for information on ranges allowed. Press F1 for help.

Select the connection address and other attributes as desired and press Enter.

C-SPOC executes the necessary commands to define the disk on all selected nodes.

Removing a Disk Definition on Cluster Nodes Using C-SPOC

Before removing a disk from the cluster using C-SPOC, check that the disk to be removed is not currently part of an existing volume group. If it is, use the C-SPOC **cl_reducevg** command to remove a physical volume from a volume group.

Take the following steps to remove a configured disk on all selected nodes in the cluster:

1. Enter the fastpath **smitty cl_admin**
2. Select **Cluster Disk Manager** > **Remove a Disk from the Cluster**.

   SMIT displays a list of nodes in the cluster that currently have the disk configured. and prompts you to select the nodes where the disk should be removed.
3. Select the nodes from which you want to remove the disk configuration. (You may have removed the cable from some of the nodes in the cluster and only want the disk configuration removed from those nodes.)

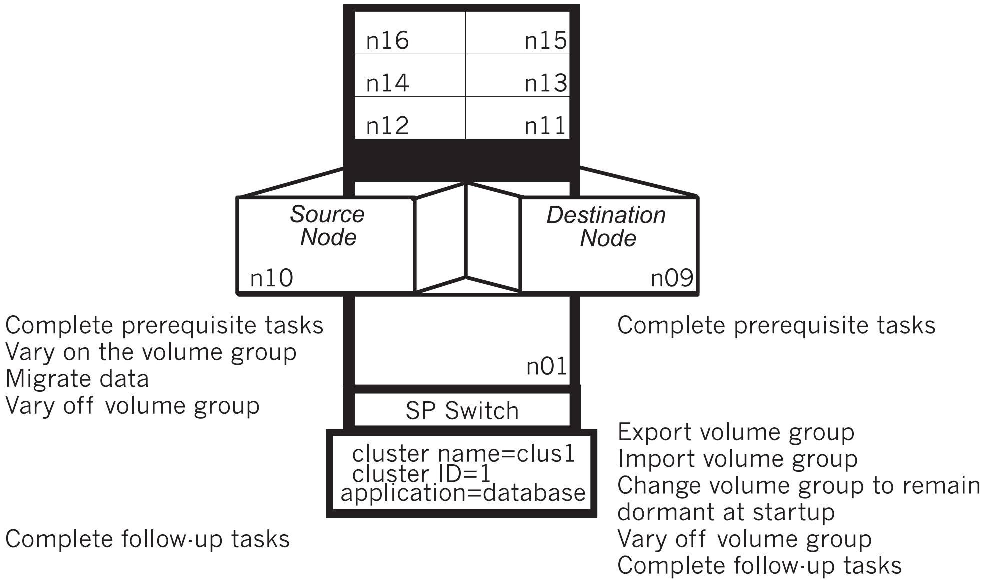
   SMIT displays the AIX **Remove a Disk** screen with the selected disks displayed.
4. For the entry “Keep the disk definition in database” Select **yes t**o keep the definition in the database; select **no** to delete the disk from the database. Press Enter.

   C-SPOC sends the **rmdev** command to all nodes listed to remove the selected disk.

Migrating Data

The figure below summarizes the steps you must complete on the source and destination nodes to migrate data from one physical volume to another physical volume.

Note: The physical volumes (**hdisks**) should be installed, configured, and available.



Migrating Data

To migrate data from one physical volume to another, perform the following procedure.

1. Stop HACMP/ES cluster services on nodes in the resource group chain.
2. On the source node, vary on the volume group, using the SMIT **varyonvg** fastpath.
3. On the source node, move data from the physical volume (or volumes) being removed from the volume group, using the SMIT **migratepv** fastpath. Set the **Move only data belonging to this LOGICAL VOLUME** field to **no**.

   Warning: If you do not perform this step, data will be lost.

4. On the source node, reduce the size of the volume group, using the SMIT **reducevg** fastpath. SMIT displays the **Reduce a Volume Group** screen.

   Select the **Remove a Physical Volume from a Volume Group** option and press Enter. SMIT prompts you to choose the volume group you want to reduce. Press the F4 key to obtain a list of volume groups.

5. On the source node, vary off the volume group, using the SMIT **varyoffvg** fastpath.
6. On the destination nodes, export the volume group, using the SMIT **exportvg** fastpath.
7. On the destination nodes, import the volume group, using the SMIT **importvg** fastpath.

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 **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 be 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.

8. On the destination nodes, change the volume group to remain dormant at startup, using the SMIT **chvg** fastpath.
9. On the destination nodes, vary off the volume group, using the SMIT **varyoffvg** fastpath.
10. Restart cluster services on all nodes.

Chapter 23 Maintaining Shared LVM Components in a Concurrent Access Environment

This chapter explains how to maintain shared LVM components in a concurrent access environment.

While the maintenance tasks for shared LVM components in concurrent access environments are similar to those of non-concurrent access environments, described in Chapter 22, Maintaining Shared LVM Components, there are some special considerations in a concurrent access environment.

Overview

In a *concurrent access* environment, two or more nodes can have simultaneous (concurrent) access to a volume group that resides on shared external disks.

Note: You cannot define filesystems on a concurrent access volume group.

HACMP/ES supports concurrent access volume groups and logical volumes defined on any of the following devices:

- IBM 7133 and 7131-405 Serial Storage Architecture (SSA) Disk Subsystems
- IBM 7135-110 and IBM 7135-210 RAIDiant Disk Arrays
- IBM 2105-B09 and 2105-100 Versatile Storage Servers

The following sections first describe how the HACMP/ES scripts handle concurrent access LVM components. Then the chapter includes sections that describe how to perform specific administrative tasks for LVM components in a concurrent access environment.

Many maintenance tasks can be performed using the HACMP/ES C-SPOC utility.

Creating a concurrent volume group or adding nodes to an existing volume group can also be done using the TaskGuide utility. For further information and instructions, see page 23-2.

Concurrent Access and HACMP/ES Scripts

You should seldom, if ever, need to intervene in a concurrent access cluster. In a concurrent access environment, as in a non-concurrent environment, the HACMP/ES event scripts control the actions taken by a node and coordinate the interactions between the nodes. However, as a system administrator, you should monitor the status of the concurrent access volume groups when HACMP/ES events occur.

To intervene in a cluster, you must understand how nodes in a concurrent access environment control their interaction with shared LVM components. For example, the HACMP/ES **node_up_local** script may fail before varying on a volume group in concurrent mode. After fixing whatever problem caused the script to fail, you may need to manually vary on the volume group in concurrent access mode. The following sections describe the processing performed by these scripts.

Nodes Join the Cluster

A node joining a cluster calls the **node_up_local** script which calls the **cl_mode3** script to activate the concurrent capable volume group in concurrent access mode.

The **cl_mode3** script calls the **varyonvg** command with the **-c** flag. For more information about this command and its flags, see “Activating a Concurrent Capable Volume Group in Concurrent Access Mode.” If the concurrent capable volume group is defined on a RAID disk array device, the scripts use the **convaryonvg** command to vary on the concurrent volume groups in concurrent mode.

Nodes Leave the Cluster

Nodes leaving the cluster do not affect the concurrent access environment. They simply vary off the volume groups normally. The remaining nodes take no action to change the concurrent mode of the shared volume groups.

When a node leaves the cluster gracefully, it executes the **node_down_local** script which calls the **cl_deactivate_vgs** script. This script uses the **varyoffvg** command to vary off the concurrent volume groups.

Maintaining Concurrent Access Volume Groups

The LVM enables you to create concurrent access volume groups that can be varied on in either concurrent access mode or non-concurrent access mode. Maintaining these volume groups may require you to perform any of the following tasks:

- Creating a concurrent capable volume group in concurrent access mode
- Varying on a concurrent capable volume group in concurrent access mode
- Determining if a volume group is a concurrent capable volume group
- Determining if a volume group is varied on in concurrent access mode
- Replacing a failed drive in a concurrent access volume group
- Maintaining a concurrent access environment while upgrading HACMP/ES software
- Restarting the LVM concurrent access daemon (**clvmd**).

The following sections describe how to perform these tasks.

Tasks you can perform with C-SPOC are described in a section at the end of the chapter.

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 and 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 (for non-concurrent access), as you would do manually when creating a non-concurrent shared volume group without the TaskGuide. In addition, the TaskGuide now displays the physical location of available disks.

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:

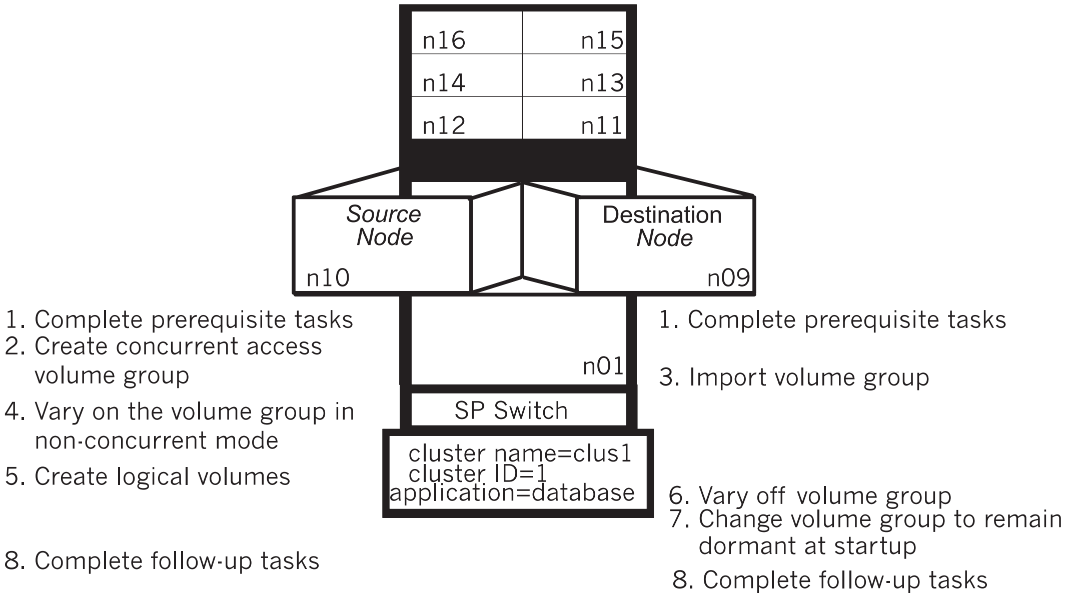
1. Type `smit hacmp`
2. From the SMIT main menu, choose **Cluster System Management > Cluster Logical Volume Manager >Taskguide for Creating a Shared Volume Group**

   After a pause, the TaskGuide Welcome panel appears.
3. Proceed through the panels to create or share a volume group.

   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.

Creating a Concurrent Volume Group

The figure below summarizes the steps you must complete on the source and destination nodes in an HACMP/ES cluster to create a concurrent volume group that HACMP/ES can vary on in concurrent access mode.


n16
n15
n14
n13
n12
n11
Source Node
n10
Destination Node
n09
n01
SP Switch
cluster name=clus1
cluster ID=1
application=database
1. Complete prerequisite tasks
2. Create concurrent access volume group
4. Vary on the volume group in non-concurrent mode
5. Create logical volumes
8. Complete follow-up tasks
1. Complete prerequisite tasks
3. Import volume group
6. Vary off volume group
7. Change volume group to remain dormant at startup
8. Complete follow-up tasks

Creating a Concurrent Access Volume Group

You are likely to plan to create concurrent volume groups on SSA disk subsystems as part of the procedure for replacing a failed drive in an SSA concurrent access volume group. In order to enable this procedure, you must assign unique non-zero node numbers on each node of the cluster.

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.

The following sections describe each step of creating a concurrent volume group in detail.

Step 1. Complete Prerequisite Tasks

The physical volumes (**hdisks**) should be installed, configured, and available. You can verify the disks’ status with the **lsdev -Cc disk** command.

Step 2. Create a Concurrent Access Volume Group on Source Node

The procedure you use to create a concurrent access volume group varies depending on the type of device you are using.

Creating a Concurrent Access Volume Group on Serial Disk Subsystems

To use a volume group in concurrent access mode on a serial disk subsystem, such as an IBM 7133 SSA 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 a logical volume 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, as in the following example:

```
mkvg -n -s 4 -c -V 35 -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 SMIT screen. The screen has additional fields in a concurrent access environment.

2. Enter the specific field values as follows:

   VOLUME GROUP name Specify name of volume group.

   Physical partition SIZE in megabytes Accept 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 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 You can accept the default; however, using the **lvlstmajor** command is recommended to find a free major number.

   Create VG concurrent capable? Set 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 to **no** so that the volume group can be activated as appropriate by the cluster event scripts.

3. Press Enter. The system asks if you are sure. Press Enter again.

4. After the command completes, press F10 to exit SMIT.

Creating a Concurrent Access Volume Group on RAID Disk Subsystems

On IBM 7135 RAIDiant Disk Arrays and IBM 2105 Versatile Storage Servers, use the **mkvg** command to create concurrent access volume groups. Do not use the **-c** flag, since RAID disks don’t use the CLVM.

Alternatively, you can 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 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.

Step 3. Vary On the Concurrent Access Volume Group in Non-Concurrent Mode

Use the **varyonvg** command to activate the volume group in non-concurrent mode. To create logical volumes, the volume group must be varied on in non-concurrent mode. For example, to vary on the concurrent access volume group **myvg** in non-concurrent mode, enter the following command:

```
varyonvg myvg
```

You can also use SMIT to vary on the volume groups, as follows:

1. To vary on a concurrent capable volume group in non-concurrent mode, enter:

   ```
   smit varyonvg
   ```

   SMIT displays the **Activate a Volume Group SMIT** screen; it has additional fields in a concurrent access environment.

2. Enter field values as follows.

   VOLUME GROUP name Specify name of volume group.

   RESYNCHRONIZE stale physical partitions? Set 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.

3. Press Enter. The system asks if you are sure. Press Enter again.
4. After the command completes, press F10 to exit SMIT.

Step 4. Create Logical Volumes on Concurrent Access 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. For more information about creating logical volumes, see the *AIX System Management Guide*.

1. To create a logical volume, enter:

   ```
   smit mklv
   ```

   SMIT displays the Add a Logical Volume menu.
2. In general, you can supply values for these fields following the guidelines in Chapter 12, Defining Shared LVM Components. However, specify the following values for these selected fields for logical volumes created on concurrent access volume groups:

Number of COPIES of each logical partition	Specify 1, 2, or 3 mirror copies. **Note:** If you defined the volume group on an IBM 7135-110 or IBM 7135-210 disk array, do not use LVM mirroring.
Mirror Write Consistency	Specify the value **no**.
Enable BAD BLOCK relocation?	Specify the value **no**.

3. Press Enter. The system asks if you are sure. Press Enter again.
4. After the command completes, press F10 to exit SMIT.

Step 5. 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 6. Import Volume Group Information on Destination Nodes

On each destination node in turn, import the volume group, using the **importvg** command, as in the following example:

```
importvg -V 35 -y vg_name physical_volume_name
```

Specify the name of any of the disks in the volume group as an argument to the **importvg** command.

Note: You must vary on concurrent capable volume groups defined on serial disk subsystems—they are not varied on automatically upon import by AIX. Concurrent access volume groups defined on RAID disk subsystems are varied on automatically when imported.

With SMIT use the following procedure.

1. To import a concurrent access volume group, enter:

   ```
   smit importvg
   ```

   SMIT displays the **Import a Volume Group** SMIT screen; it has additional fields in a concurrent access environment.

2. Enter field values as follows:

   VOLUME GROUP name Enter the name of the volume group that you are importing. Make sure the volume group name is the same name that you used on the 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 You can accept the default; however, using the **lvlstmajor** command to find a free major number is recommended.

   Make this VG concurrent capable Accept the default value (**no**). Only used when importing a volume group created on an earlier version of AIX. For more information, see “Maintaining Concurrent Access During a Cluster Upgrade.”

   Make default varyon of VG concurrent Accept the default value (**no**). Only used when importing a volume group created on an earlier version of AIX. For more information, see “Maintaining Concurrent Access During a Cluster Upgrade.”

3. Press Enter to commit the information.

4. 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, make sure there is no reserve on any of the disks in the volume group, by executing the following command:

   ```
   /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 7. Change Volume Group to Remain Dormant at Startup on Destination Nodes

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 change the volume group so that it remains dormant at startup.

To change this characteristic of a volume group enter the following:

```
chvg -a n volume_group_name
```

You can also use SMIT by using the following procedure:

1. Use the **smit chvg** fastpath to change the characteristics of a volume group.
2. 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**.
3. Press Enter to commit this change.
4. 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 access volume group as a concurrent cluster resource.

Activating a Volume Group in Concurrent Access Mode

As a system administrator, you may, at times, need to activate (vary on) a concurrent capable volume group in concurrent access mode. For example, if an HACMP/ES script fails before activating a volume group in concurrent access mode, you may want to activate the volume group manually after correcting the failure.

The method used to vary on a volume group in concurrent access mode varies, depending on the type of device. The following sections describe the procedure for IBM 7133 or 7131-405 SSA Disk Subsystems and IBM 7135 and IBM 2105 Versatile Storage Servers RAID devices.

For concurrent access volume groups, bad block relocation is automatically turned off.

Activating Volume Groups Defined on Serial Disk Subsystems

If the volume group is defined on an IBM 7133 or 7131-405 SSA Disk Subsystem, you can activate the volume group using **varyonvg** command, specifying the **-c** flag, as in the following example:

```
varyonvg -c myvg
```

Note: To vary on a volume group in concurrent access mode, the HACMP/ES software must be present on the node.

You can also use SMIT to activate a volume group in concurrent access mode. Use the following procedure.

1. Enter:

   ```
   smit varyonvg
   ```

SMIT displays the **Activate a Volume Group** SMIT screen; it has an additional field in a concurrent access environment.

2. Enter the 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 (**no**).
FORCE activation of the Volume Group?	Accept the default (**no**).
Varyon VG in concurrent mode?	Set to **yes**.

3. Press Enter. The system asks if you are sure. Press Enter again.
4. After the command completes, press F10 to exit SMIT.

Activating Volume Groups Defined on RAID Devices

To activate a volume group defined on a 7135 or 2105 Versatile Storage Server RAID device in concurrent access mode, you must use the **/usr/es/sbin/cluster/events/utils/convaryonvg** command. This command first checks to make sure the disks in the volume group are not reserved.

For example, to vary on the *sharedvg1* volume group in concurrent mode, enter:

```
/usr/es/sbin/cluster/events/utils/convaryonvg sharedvg1
```

Determining the Access Mode of a Volume Group

To determine if a volume group is a concurrent capable volume group and to determine its current mode, use the **lsvg** command specifying the name of the volume group as an argument. The **lsvg** command displays the following information about the volume group:

```
# lsvg myvg
VOLUME GROUP:   myvg               VG IDENTIFIER:  000025426733bc50
VG STATE:       active             PP SIZE:        16 megabyte(s)
VG PERMISSION:  read/write         TOTAL PPs:      119 (1904 megabytes)
MAX LVs:        256                FREE PPs:       119 (1904 megabytes)
LVs:            10                 USED PPs:       0 (0 megabytes)
OPEN LVs:       8                  QUORUM:         2
TOTAL PVs:      1                  VG DESCRIPTORS: 2
STALE PVs:      0                  STALE PPs       0
ACTIVE PVs:     1                  AUTO ON:        no
Concurrent:     Capable            Auto-Concurrent: disabled
VG Mode:        Concurrent
```

To determine if the volume group is concurrent capable, check the value of the **Concurrent** field. The volume group in the example was created as a concurrent capable volume group, as indicated by the value of this field. If this volume group was not a concurrent capable volume group, the value of this field would be *Non-Capable*.

To determine if the volume group is activated in concurrent access mode, check the value of the **VG Mode** field. In the example, the volume group is activated in concurrent access mode. If this volume group had not been varied on in concurrent access mode, the value of this field would be *Non-Concurrent*.

The **Auto-Concurrent** field indicates whether the volume group should be varied on in concurrent access mode when the volume group is started automatically at system reboot. The value of this field is determined by the value of the **-x** option to the **mkvg** command when the volume group was created. In an HACMP/ES environment, this option should always be disabled; HACMP/ES scripts control when the volume should be varied on.

Replacing a Failed Drive in a Concurrent Access Volume Group

The following five volume group reconfiguration commands allow you to replace a failed disk drive belonging to a concurrent access volume group:

extendvg Add a physical volume to an existing volume group.

reducevg Remove a physical volume from an existing volume group.

mklvcopy Add a mirror copy to an existing logical volume.

rmlvcopy Remove a mirror copy from an existing logical volume.

syncvg Brings all copies of physical partitions in a logical volume up to date.

The concurrent capable volume group must be varied on in concurrent access mode *on all nodes* for the changes made by these reconfiguration commands to be made known to all cluster nodes automatically. If a node has the volume group varied off, it must export the volume group and then re-import the volume group before varying it on again.

Note: Make sure that the concurrent access volume group is in a quiescent state (no I/O operations in progress) before executing these reconfiguration commands.

Two procedures follow. The first is for IBM 9333 serial disk subsystems and IBM RAIDiant Disk Arrays; the second is for IBM SSA disk subsystems.

Replacing a Failed Drive in a Concurrent Access Volume Group on an IBM RAIDiant Disk Array

The following example illustrates the procedure for replacing a failed drive in a concurrent access volume group on an IBM 7135 RAIDiant Disk Array or IBM 2105 Versatile Storage Server. In the example, three cluster nodes (named A, B, and C) share access to a concurrent access volume group, named *sharedvg*, on which three copies of the logical volume *sharelv* are maintained.

The example assumes that the system error log reports problems with *hdisk4.* The data in the volume group remains accessible because the number of disks in the volume group that are still available constitute a quorum. However, if the disk is not repaired or replaced, high availability is endangered.

To replace a failed drive in a concurrent access volume group:

1. Determine the contents of the failed disk drive by executing the **lsvg** command:

```
lsvg -M -n sharedvg
 hdisk2:1      sharelv:1:1
 hdisk2:2      sharelv:2:1
 hdisk2:3      sharelv:3:1
```

```
hdisk2:4-50
hdisk3:1      sharelv:1:2
hdisk3:2      sharelv:2:2
hdisk3:3      sharelv:3:2
hdisk3:4-50
hdisk4:1      sharelv:1:3   stale
hdisk4:2      sharelv:2:3   stale
hdisk4:3      sharelv:3:3   stale
hdisk4:4-50
```

The output indicates that the logical volume mirror copies on *hdisk4* are stale. This is the drive that has failed and must be replaced.

2. Reduce the number of copies of the physical partitions to only those that are currently available, using the **rmlvcopy** command:

   ```
   rmlvcopy sharelv 2 hdisk4
   ```

   This command reduces the number of copies of the logical volume *sharelv* to two and specifies that the copy to be removed is on the failed drive, *hdisk4.*

3. Reduce the size of the volume group to omit the failed drive, using the **reducevg** command:

   ```
   reducevg sharedvg hdisk4
   ```

   This command specifies that the physical volume *hdisk4* should be removed from the volume group *sharedvg.*

4. Execute the **rmdev** command on *each* cluster node to remove the definition of the failed drive from the device configuration database:

   ```
   rmdev -l hdisk4 -d
   ```

   The disk may not be named *hdisk4* on all the nodes. Use the **lspv** command on each node to obtain a list of disk PVIDs to accurately identify the failed disk.

5. Remove the failed drive and replace it with a new drive. Refer to the documentation that accompanied the disk subsystem for information about removing and replacing an individual drive. *Power down only the failed drive. Do not power down the entire disk subsystem.*

   Run the **diag** command to format the new drive, if necessary.

6. Configure the new drive on *each* cluster node. Do *not* use the **cfgmgr** command to reconfigure the disk drive; instead, use the following commands. The example configures the adapters and disks on an IBM 9333 disk subsystem:

   ```
   lsdev -Ccadapter -tserdasda | awk '{print $1}' | xargs -n1 HR>
   -i{} /etc/methods/cfgserdasda -l {}
   lsdev -Ccadapter -sserdasda | awk '{print $1}' | xargs -n1 HR>
   -i{} /etc/methods/cfgserdasdc -l {}
   lsdev -Ccdisk -sserdasdc | awk '{print $1}' | xargs -n1 HR>
   -i{} mkdev -l {}
   ```

7. Determine the logical name of the new disk, using the **lsdev** command:

   ```
   lsdev -Cc disk
   ```

   The system displays a list of all currently configured disks. To find out what name the system assigned to the replacement disk, you must know the location code of the drive. This code can be obtained by entering:

   ```
   lsdev -Ccdisk -sserdasdc
   ```

   The following steps assume that the system assigned the same logical name to the new drive (hdisk4).

8. Add the new disk to the concurrent access volume group, using the **extendvg** command:

```
extendvg sharedvg hdisk4
```

9. Recreate the mirrors of the logical volume on the new drive, use the **mklvcopy** command:

   ```
   mklvcopy sharelv 3 hdisk4
   ```

   Do not specify the **-k** option which synchronizes the logical volume mirrors as you create the copies.

10. Synchronize the new mirror copies of the logical volume with the current mirrors of the logical volume, using the **syncvg** command:

    ```
    syncvg -v sharevg
    ```

This completes the procedure for replacing a failed drive in a concurrent access volume group on an IBM RAIDiant Disk Array.

Replacing a Failed Drive in a Concurrent Access Volume Group on an IBM SSA Disk Subsystem

The following example illustrates the procedure for replacing a failed SSA drive in a concurrent access volume group. In the example, two cluster nodes (named A and B) share access to a concurrent access volume group named *sharedvg*, on which two copies of the logical volume *lvmarket* are maintained. The example assumes that an SSA disk configured as *hdisk6* has failed.

Warning: For this procedure to work, you must first have assigned unique non-zero node_numbers through the ssar on each cluster node. If you have specified (or are planning to specify) SSA disk fencing in your concurrent resource group, these node_numbers are assigned when you synchronize the resources. If, however, you do not specify SSA disk fencing, issue the following command to assign node_numbers: chdev -l ssar -a node_number=x, where x is the number to assign to the node. Reboot the system to make the node_number change effective. Changing node_numbers is best done after installing HACMP/ES and after attaching and configuring your SSA devices.

Complete the following steps to replace a failed SSA drive in a concurrent access volume group. You must be the root user.

Perform Steps 1 Through 5 on Node A.

1. Run the **lsvg** command to determine which logical volumes have partitions on the failed disk:

   ```
   # lsvg -M sharedvg
       hdisk6:126     lvmarket:72:2   stale
       hdisk6:127     lvmarket:73:2   stale
       hdisk6:128     lvmarket:74:2   stale
       hdisk6:129     lvmarket:75:2   stale
       hdisk6:130     lvmarket:76:2   stale
       hdisk6:131     lvmarket:77:2   stale
       hdisk6:132     lvmarket:78:2   stale
       hdisk6:133     lvmarket:79:2   stale
       hdisk6:134     lvmarket:80:2   stale
   ```

   The output indicates that the logical volume mirror copies on *hdisk6* are stale. This is the drive that has failed and must be replaced.

2. For each logical volume with partitions on the failed disk, use the **rmlvcopy** command to remove the mirror copies from that disk:

   ```
   # rmlvcopy lvmarket 1 hdisk6
   ```

   This command reduces the number of copies of the logical volume *lvmarket* to one and specifies that the copy to be removed is on the failed drive, *hdisk6.*

   Note: The **rmlvcopy** command also causes a synclvodm to be run on the other node, which resets the permissions on the raw device to the default values. A permissions change can affect applications using the device since it can cause them to lose access to the device.

3. Run the **reducevg** command to remove the failed disk from the volume group:

   ```
   # reducevg sharedvg hdisk6
   ```

   This command specifies that the physical volume *hdisk6* should be removed from the volume group *sharedvg.*

4. Determine which pdisk corresponds to the hdisk, if the disk has failed completely, by issuing the following command:

   ```
   lsdev -Ccpdisk -Fconnwhere,name|grep `lsdev -ClhdiskX HR>-Fconnwhere`
   ```

   This command gives the pdisk name and the “connwhere” value of the failing hdisk and pdisk. The serial number printed on the front of the failed SSA disk drive is represented by the fifth through the twelfth characters of the connwhere value.

   If the disk has not failed completely, run the **diag** command as follows:

   a. Choose the Service Aids option.

   b. Choose the SSA Service Aids option.

   c. Choose the Configuration Verification option.

   d. Choose the appropriate hdisk and press ENTER. The corresponding pdisk is displayed with the serial number on the disk drive. Remember the pdisk number, for example, pdisk1.

   e. Exit the diagnostic service aids.

   Use the pdisk number in the following step:

5. Remove the failed disk.

   If the disk has failed completely, its check light remains on, or all indicators on the device are off. The adjacent devices show flashing ready lights. You can now remove the drive without first setting Service Mode.

   If, however, the disk has not failed completely, run the diag command as follows:

   a. Choose the Service Aids option.

   b. Choose the SSA Service Aids option.

   c. Choose the Configuration Verification option.

   d. Choose “Set or Reset Service Mode”

      The drive’s yellow check light turns on.

   e. Remove this drive from the SSA enclosure.

   f. Exit the diagnostic service aids.

Perform Steps 6 Through 9 on Both Nodes.

6. Run the **rmdev** command to remove the hdisk and pdisk devices on both nodes. Note that, while the pdisk device number is always the same on both nodes, the hdisk number for the same disk may differ from node to node, depending on the disk configuration of each node.

   You can verify that you have the correct disk using the following commands:

   ```
   # lsdev -Cl hdiskX -Fconnwhere
   # lsdev -Cl pdiskX -Fconnwhere
   ```

   Then remove the hdisk and pdisk devices on both nodes as follows:

   ```
   # rmdev -d -l hdisk6
   # rmdev -d -l pdisk1
   ```

   The resulting character string should be identical for the pdisk and hdisk on node A and node B.

7. Replace the failed disk with a new one.

8. Run the following commands on each node to configure the new device. Do not run the **cfgmgr** command while the cluster is running.

   ```
   # /usr/lib/methods/cfgssar -l ssar >/dev/null
   # lsdev -Cssar -S0 iFname|xargs -n1
   ```

9. Run the **lspv** command to check the hdisk name of the newly configured disk. It should be named identical to the removed hdisk.

Perform Steps 10 Through 12 on Node A.

10. Run the **extendvg** command to add the disk back into the volume group:

    ```
    # extendvg sharedvg hdisk6
    ```

11. Run the **mklvcopy** command to recreate the mirror copies onto the new disk:

    ```
    # mklvcopy lvmarket 2 hdisk6
    ```

12. Run the **syncvg** command to synchronize the new mirror copies with the existing mirror copies for each affected logical volume:

    ```
    # syncvg -l lvmarket
    ```

This completes the procedure for replacing a failed SSA drive in a concurrent access volume group.

Maintaining Concurrent Access During a Cluster Upgrade

When upgrading a cluster, you typically bring down the cluster nodes one at a time, install the new HACMP/ES software, and then reintegrate the node into the cluster. As part of this process, you import the shared volume group definitions onto the node that is reintegrating.

When upgrading a cluster to HACMP/ES, Version 4.4, in which a concurrent access volume group is configured, you must specify the **-c** flag with the **importvg** command when importing the shared concurrent access volume groups, as in the following:

```
importvg -y volume_group_name -c physical_disk_name
```

Concurrent access volume groups created with versions of AIX before 4.2 do not identify the volume groups as concurrent capable in the Volume Group Descriptor Area (VGDA). If the volume group is not identified as concurrent capable, the new version of AIX activates the imported volume group automatically in non-concurrent access mode. When one node varies on a concurrent access volume group in non-concurrent mode, the access mode reverts to non-concurrent access for all nodes accessing the volume group, thus locking out the other cluster nodes. By specifying the **-c** flag with the **importvg** command, AIX updates the on-disk copy of the volume group's VGDA to indicate it is concurrent capable.

Note: Once you change a volume group to be concurrent capable, you cannot change it back to a standard volume group.

Restarting the Concurrent Access Daemon (clvmd)

As a system administrator, you may, at times, need to restart the concurrent access daemon (**clvmd**). The **clvmd** daemon normally gets started by the **varyonvg** command when you vary on a volume group in concurrent access mode by specifying the **-c** flag. You can restart the **clvmd** by re-executing the **varyonvg -c** command on the already-varied on concurrent access volume group. You cannot vary on an already-varied on volume group in a different mode, however.

You can also restart the **clvmd** daemon using the following SRC command:

```
startsrc -s clvmd
```

By using the SRC facility command to start the daemon, you ensure that only a single copy of the daemon is running on a node at any one time.

Maintaining Concurrent LVM Components with C-SPOC

In HACMP/ES version 4.4, C-SPOC uses the AIX 4.3 CLVM capabilities that allow changes to concurrent LVM components without stopping and restarting the cluster. See Chapter 22, Maintaining Shared LVM Components, for a general explanation of how C-SPOC works.

You can use the C-SPOC utility to do the following tasks:

- Concurrent volume groups tasks
  - Create a concurrent volume group on selected cluster nodes
  - List all concurrent volume groups in the cluster
  - Import a concurrent volume group
  - Extend a concurrent volume group
  - Reduce a concurrent volume group
  - Mirror a concurrent volume group
  - Unmirror a concurrent volume group
  - Synchronize concurrent LVM mirrors by volume group.
- Concurrent logical volumes tasks
  - List all concurrent logical volumes by volume group
  - Add a concurrent logical volume to a volume group
  - Make a copy of a concurrent logical volume
  - Remove a copy of a concurrent logical volume
  - Show the characteristics of a concurrent logical volume
  - Remove a concurrent logical volume

- Synchronize concurrent LVM mirrors by logical volume.

Note: The volume group must be varied on in concurrent mode in order to do these tasks.

Maintaining Concurrent Volume Groups Using C-SPOC

You can use C-SPOC for all concurrent volume group maintenance tasks.

Creating a Concurrent Volume Group on Cluster Nodes Using C-SPOC

Using C-SPOC simplifies the procedure for creating a concurrent volume group on selected cluster nodes. Before creating a concurrent volume group for the cluster using C-SPOC, check that:

- All disk devices are properly attached to the cluster nodes
- All disk devices are properly configured on all cluster nodes and listed as available on all nodes
- The cluster concurrent logical volume manager is installed
- All disks that will be part of the volume group must be concurrent capable.

Take the following steps to create a shared volume group for a selected list of cluster nodes:

1. Enter the fastpath **smitty cl_admin**
2. Select **Cluster Concurrent Logical Volume Manager** > **Concurrent Volume Groups** > **Add a Concurrent Volume Group**.

   SMIT displays a list of cluster nodes.
3. Select two or more nodes from the list of cluster nodes and press Enter.

   The system correlates a list of all free concurrent-capable physical disks that are available to all nodes selected. (Free disks are those disks that currently are not part of a volume group and have PVIDs.) SMIT displays the list of free physical disks in a multi-pick list by PVIDs
4. Select one or more disks from the list and press Enter.

   SMIT displays the **cl_mkvg** screen with a major number inserted into the Major Number data field. The system determines this free major number; do not change it.
5. Complete the selections as follows

Nodes	Names of the selected nodes are displayed.
VOLUME GROUP name	Enter a name for this concurrent capable volume group.
Physical partition SIZE in megabytes	Accept the default.
*** PHYSICAL VOLUME names**	The system displays the selected disks.

Volume Group MAJOR NUMBER	The system displays the number C-SPOC has determined to be correct. **Warning**: Changing the volume group major number may result in the command's inability to execute on a node that does not have that major number currently available. Please check for a commonly available major number on all nodes before changing this setting.

C-SPOC verifies communication paths and version compatibility and then executes the command on all the nodes you selected.

Note: If the major number entered on the SMIT panel was not free at the time that the system attempted to make the volume group the command will display an error for the node that did not complete the execution and continue to the other nodes. At the completion of the command the volume group will not be active on any node in cluster.

Listing all Concurrent Volume Groups in the Cluster

To list all concurrent volume groups in the cluster, take the following steps:

1. On the source node, enter **smit hacmp**.
2. From the main HACMP menu, choose **Cluster System Management > Cluster Concurrent Logical Volume manager > List all Concurrent Volume Groups** and press Enter.
3. SMIT displays a message asking if you want to view only active concurrent volume groups.

   Choose **yes** to see a list of active concurrent volume groups only, or choose **no** to see a list of all concurrent volume groups.

Extending a Concurrent Volume Group with C-SPOC

Take the following steps to add a physical volume to a concurrent volume group using C-SPOC.

1. The physical volumes (**hdisks**) being added to the volume group must be installed, configured, and available.
2. On any cluster node that can own the volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already in concurrent mode).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Concurrent Logical Volume manager > Concurrent Volume Groups > Set Characteristics of a Concurrent Volume Group > Add a Physical Volume to a Concurrent Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. You can pick one or more to add to the volume group. Select the ones you want to add to the volume group and Press Enter.

7. SMIT displays the Add a Physical Volume to a Concurrent Volume Group screen, with the following entries filled in.

    Resource Group name The cluster resource group to which this concurrent volume group belongs.

    Volume Group name Name of the volume group where hdisks are to be added.

    Reference node Name of the node where the hdisks are found.

    Physical Volume names Names of the hdisks to be added to the volume group.

8. If this screen reflects the correct information, press Enter to add the disks to the concurrent volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the concurrent volume group varied on, vary off the volume group on that node.

Importing a Concurrent Volume Group with C-SPOC

Take the following steps to import a concurrent volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Concurrent Logical Volume manager > Concurrent Volume Groups > Import a Concurrent Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the Import a Concurrent Volume Group screen. Values for fields you have selected are displayed. For other fields, use the defaults or the appropriate entries for your operation:

    VOLUME GROUP name The name of the volume group that you are importing.

    PHYSICAL VOLUME name The name of one of the physical volumes that resides in the volume group. This is the hdisk name on the reference node.

    Reference node Node from which the physical disk was retrieved.

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 be 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.

Make this VG concurrent capable? The default is **no**. Change to **yes** for concurrent VGs.

Make default varyon of VG concurrent? The default is **no**. Change to **yes** for concurrent VGs.

8. If this screen reflects the correct information, press Enter to import the concurrent volume group. All nodes in the cluster receive this updated information immediately (before lazy update).
9. If you did this task from a cluster node that does not need the concurrent volume group varied on, vary off the volume group on that node.

Removing a Physical Volume from a Concurrent Volume Group with C-SPOC

Take the following steps to remove a physical volume from a concurrent volume group using the C-SPOC utility.

1. Complete prerequisite tasks. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management >Cluster Concurrent Logical Volume manager > Concurrent Volume Groups >Set Characteristics > Remove a Physical Volume from a Concurrent Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the **Remove a Physical Volume from a Concurrent Volume Group** screen, with the following entries filled in.:

VOLUME GROUP name The name of the volume group that you are reducing.

Reference node Node from which the name of the physical disk was retrieved.

PHYSICAL VOLUME name The name of the physical volume that you want to remove.This is the hdisk name on the reference node.

8. If this screen reflects the correct information, press Enter to reduce the concurrent volume group. All nodes in the cluster receive this updated information immediately (before lazy update).

9. If you did this task from a cluster node that does not need the concurrent volume group varied on, vary off the volume group on that node.

Mirroring a Concurrent Volume Group Using C-SPOC

Take the following steps to mirror a concurrent volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Concurrent Logical Volume manager > Concurrent Volume Groups > Mirror a Concurrent Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the **Mirror a Concurrent Volume Group** screen, with the following entries filled in.

   For other fields, use the defaults or the appropriate entries for your operation:

   Resource Group Name The name of the resource group to which this concurrent volume group belongs is displayed.

   VOLUME GROUP name The name of the volume group that you want to mirror is displayed.

   Reference node Node from which the name of the physical disk was retrieved is displayed.

   PHYSICAL VOLUME names The name of a physical volume on the volume group that you want to mirror.This is the hdisk name on the reference node.

   Mirror sync mode **Foreground** is the default. Other choices are **Background** and **No Sync.**

   Number of COPIES of each logical partition The default is **2**. You can also select **3**.

   Keep Quorum Checking On? The default is **no.** You can also choose **yes.**

   Create Exact LV Mapping? The default is **no**.

8. If this screen reflects the correct information, press Enter to mirror the concurrent volume group. All nodes in the cluster receive this updated information.

9. If you did this task from a cluster node that does not need the concurrent volume group varied on, vary off the volume group on that node.

Unmirroring a Concurrent Volume Group Using C-SPOC

Take the following steps to unmirror a concurrent volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Concurrent Logical Volume manager > Concurrent Volume Groups > Unmirror a Concurrent Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the **Unmirror a Concurrent Volume Group** screen, with the chosen fields filled in.

   For other fields, use the defaults or the appropriate entries for your operation:

   Resource Group Name The name of the resource group to which this concurrent volume group belongs is displayed.

   VOLUME GROUP name The name of the volume group that you want to mirror is displayed.

   Reference node Node from which the name of the physical disk was retrieved is displayed.

   PHYSICAL VOLUME names The name of a physical volume on the volume group that you want to unmirror.This is the hdisk name on the reference node.

   Number of COPIES of each logical partition The default is **2**. You can also select **3**.

8. If this screen reflects the correct information, press Enter to unmirror the concurrent volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the concurrent volume group varied on, vary off the volume group on that node.

Synchronizing Concurrent Volume Group Mirrors

Take the following steps to synchronize concurrent LVM Mirrors by volume group using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Synchronize Concurrent LVM Mirrors > Synchronize By Volume Group**.
5. SMIT displays a list of volume groups. Select the desired volume group and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.
7. SMIT displays the **Synchronize Concurrent LVM Mirrors by Volume Group** screen, with the chosen entries filled in.

   For other fields, use the defaults or the appropriate entries for your operation:

   Resource Group Name The name of the resource group to which this concurrent volume group belongs is displayed.

   VOLUME GROUP name The name of the volume group that you want to mirror is displayed.

   Reference node Node from which the name of the physical disk was retrieved is displayed.

   Number of Partitions to Sync in Parallel Set the range from 1 to 32.

   Synchronize All Partitions The default is **no.**

   Delay Writes to VG from other cluster nodes during this Sync The default is **no**.

8. If this screen reflects the correct information, press Enter to synchronize LVM mirrors by the concurrent volume group. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the concurrent volume group varied on, vary off the volume group on that node.

Using C-SPOC to Maintain Concurrent Logical Volumes

You can use the C-SPOC utility to do many maintenance tasks on concurrent logical volumes. You can create a new logical volume or change the size of the logical volume. After you complete the procedure using SMIT, the other cluster nodes are updated with the new information.

Listing all Concurrent Logical Volumes in the Cluster

To list all concurrent logical volumes in the cluster, take the following steps:

1. On the source node, enter **smit hacmp**.
2. From the main HACMP menu, choose **Cluster System Management > Cluster Concurrent Logical Volume manager > List all Concurrent Logical Volumes by Volume Groups** and press Enter.
3. SMIT displays a message asking if you want to view only active concurrent volume groups.

   Choose **yes** to see a list of active concurrent volume groups only, or choose **no** to see a list of all concurrent volume groups.

Adding a Concurrent Logical Volume to a Cluster Using C-SPOC

To add a concurrent logical volume to a cluster using C-SPOC, take the following steps:

1. Enter the SMIT C-SPOC fastpath:

   ```
   smit cl_admin
   ```

2. Select these options **Cluster Concurrent Logical Volume Manager** >**Concurrent Logical Volumes** > **Add a Concurrent Logical Volume** and press Enter.
3. SMIT displays a popup list of concurrent volume groups. Select one and press Enter.
4. SMIT displays a list of physical volumes. Select one and press Enter. The **Add a Concurrent Logical Volume** screen appears, with chosen fields filled in as shown in the sample below.:

Resource Group name	ccur_rg
VOLUME GROUP name	concurrentvg
Reference node	a1
*** Number of LOGICAL PARTITIONS**	[]
PHYSICAL VOLUME names	hdisk16
Logical volume NAME	[]
Logical volume TYPE	[]
POSITION on physical volume	middle
RANGE of physical volumes	minimum

MAXIMUM NUMBER of PHYSICAL VOLUMES to use for allocation	[]
Number of COPIES of each logical partition	1
Mirror Write Consistency?	yes
Allocate each logical partition copy on a SEPARATE physical volume?	yes
RELOCATE the logical volume during reorganization	yes
Logical volume LABEL	[]
MAXIMUM NUMBER of LOGICAL PARTITIONS	[512]
Enable BAD BLOCK relocation?	yes
SCHEDULING POLICY for writing logical partition copies	parallel
Enable WRITE VERIFY?	no
File containing ALLOCATION MAP	[]
Stripe Size?	[Not Striped]

5. The default LV characteristics are most common. Make changes if necessary for your system and press Enter. Other cluster nodes are updated with this information.

Removing a Concurrent Logical Volume on a C-SPOC Cluster

Using the following procedure, you can remove a concurrent logical volume on any node in a cluster by executing a C-SPOC command on one of the nodes.

Note: If the logical volume to be removed contains a file system, you first must remove the file system from any specified resource group before attempting to remove the logical volume. Afterwards, be sure to synchronize cluster resources on all cluster nodes.

1. Use the C-SPOC utility to change the logical volume. Enter the SMIT fastpath:

```
smit cl_admin
```

2. Select the **Cluster Logical Volume Manager** option and press Enter.
3. Choose the **Concurrent Logical Volumes** option and press Enter.
4. Select the **Remove a Concurrent Logical Volume** option and press Enter.
5. C-SPOC provides a list of concurrent logical volumes, organized by HACMP/ES resource group. Select the logical volume you want to remove and press Enter.

To check the status of the C-SPOC command execution on other cluster nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Setting Characteristics of a Concurrent Logical Volume Using C-SPOC

You can use C-SPOC to do the following tasks:

- Add copies to a concurrent logical volume
- Remove copies from a concurrent logical volume.

Adding a Copy to a Concurrent Logical Volume Using C-SPOC

Using the following procedure, you can add a copy to a concurrent logical volume on all nodes in a cluster by executing a C-SPOC command on any node.

1. Enter the C-SPOC SMIT fastpath:

   ```
   smit cl_admin
   ```

2. Select these options: **Cluster Logical Volume Manager** > **Concurrent Logical Volumes > Set Characteristics of A Concurrent Logical Volume > Add a Copy to a Concurrent Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.

3. Choose the desired logical volume from the pick list and press Enter. SMIT displays a list of physical volumes.
4. Choose a physical volume and press Enter. SMIT displays the **Add a Copy to a Concurrent Logical Volume** screen with the Resource Group, Logical Volume, Reference Node and default fields filled.
5. Enter the new number of mirrors in the **NEW TOTAL number of logical partitions** field and press Enter. The C-SPOC utility changes the number of copies of this logical volume on all cluster nodes.

Note: The current version of the AIX **mklvcopy** command ignores the “synchronize copies” option for concurrent logical volumes, Therefore, to synchronize the new copies you must run the **Synchronizing LVM Mirrors by Logical Volume** command after the **mklvcopy** command completes.

To check the status of the C-SPOC command execution on all nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Removing a Copy from a Concurrent Logical Volume Using C-SPOC

Using the following procedure, you can remove a copy of a concurrent logical volume on all nodes in a cluster by executing a C-SPOC command on any node.

1. Enter the C-SPOC SMIT fastpath:

   ```
   smit cl_admin
   ```

2. Select these options: **Cluster Logical Volume Manager** >**Concurrent Logical Volumes >Set Characteristics of A Concurrent Logical Volume > Remove a Copy from a Concurrent Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.

3. Choose the desired logical volume from the pick list and press Enter. SMIT displays a list of physical volumes.
4. Choose the physical volumes from which you want to remove copies and press Enter. SMIT displays the **Remove a Copy from a Concurrent Logical Volume** screen with the Resource Group, Logical Volume name, Reference Node and Physical Volume names fields filled in.
5. Enter the new number of mirrors in the **NEW maximum number of logical partitions copies** field and check the **PHYSICAL VOLUME name(s) to remove copies from** field to make sure it is correct and press Enter. The C-SPOC utility changes the number of copies of this logical volume on all cluster nodes.

   To check the status of the C-SPOC command execution on all nodes, view the C-SPOC log file in **/tmp/cspoc.log**.

Show Characteristics of a Concurrent Logical Volume Using C-SPOC

To show the current characteristics of a concurrent logical volume, take the following steps.

1. On the source node, vary on the volume group in concurrent mode, using the SMIT **varyonvg** fastpath.
2. Enter the C-SPOC SMIT fastpath:

   ```
   smit cl_admin
   ```
3. Select these options: **Cluster Logical Volume Manager** >**Concurrent Logical Volumes > Show Characteristics of A Concurrent Logical Volume** and press Enter.

   SMIT displays a list of logical volumes arranged by resource group.
4. Choose the desired logical volume from the pick list and press Enter. SMIT displays the **Show Characteristics of A Concurrent Logical Volume** screen with the **Resource Group**, and **Logical Volume name** fields filled in. The **Option list** field offers three choices for the display. Toggle between **Status**, **Physical Volume map**, and **Logical Partition map**.

Synchronizing LVM Mirrors by Logical Volume

Take the following steps to synchronize concurrent LVM mirrors by logical volume using the C-SPOC utility.

1. The physical volumes (**hdisks**) in the volume group must be installed, configured, and available.
2. On any cluster node that can own the concurrent volume group (is in the participating nodes list for the resource group), vary on the volume group, using the SMIT **varyonvg** fastpath (if it is not varied on already).
3. On the source node, enter **smit hacmp**.
4. From the main HACMP menu, choose **Cluster System Management > Cluster Logical Volume manager > Synchronize Concurrent LVM Mirrors > Synchronize By Logical Volume**.
5. SMIT displays a list of logical volumes. Select the desired logical volume and press Enter.
6. SMIT displays a list of physical volumes. Pick one and Press Enter.

7. SMIT displays the **Synchronize LVM Mirrors by Volume Group** screen, with the chosen entries filled in. For other fields, use the defaults or the appropriate entries for your operation:

   Resource Group Name The name of the resource group to which this logical volume belongs is displayed.

   LOGICAL VOLUME name The name of the logical volume that you want to synchronize is displayed.

   Reference node Node from which the name of the physical disk was retrieved is displayed.

   Number of Partitions to Sync in Parallel .Set the range from 1 to 32.

   Synchronize All Partitions The default is **no.**

   Delay Writes to VG from other cluster nodes during this Sync The default is **no**.

8. If this screen reflects the correct information, press Enter to synchronize LVM mirrors by the concurrent logical volume. All nodes in the cluster receive this updated information.
9. If you did this task from a cluster node that does not need the concurrent volume group varied on, vary off the volume group on that node.

Chapter 24 Changing the Cluster Configuration

This chapter describes how to reconfigure the topology and resources in your cluster. It describes changing application server resources in detail. It also covers the DARE Resource Migration utility that allows you to change the status and location of resource groups dynamically using the **cldare** command or the SMIT interface.

Overview

When you configure an HACMP/ES cluster, configuration data is stored in HACMP-specific object classes in the ODM. (To view these object classes, use the **odmget** command, specifying as an argument one of the HACMP/ES object classes, such as HACMPcluster.) The AIX ODM object classes are stored in the default system configuration directory (DCD), **/etc/objrepos**.

You can make changes to both the cluster topology and to the cluster resources while the cluster is running (dynamic reconfiguration).

Note: The directory **/usr/sbin/cluster** and subdirectories have symbolic links to the **/usr/es/sbin/cluster** directory and subdirectories. However, files in these directories are *not* linked as they were in HACMP/ES releases prior to version 4.3.1.

Reconfiguring a Cluster Dynamically

While a cluster is running, the HACMP/ES daemons, scripts, and utilities reference the ODM data stored in the active configuration directory (ACD), not the ODM data stored in the DCD. At cluster start-up, HACMP/ES copies HACMP-specific ODM classes into this directory. If you synchronize the cluster topology or cluster resources definition while the Cluster Manager is running on the local node, this action triggers a dynamic reconfiguration (DARE) event. In a dynamic reconfiguration event, the ODM data in the DCDs on all cluster nodes is updated and the ODM data in the ACD is overwritten with the new configuration data. The HACMP/ES daemons are refreshed so that the new configuration becomes the currently active configuration.

DARE Resource Migration

Additionally, you can use the DARE Resource Migration utility described in this chapter to move resource groups to other cluster nodes to perform system maintenance on a particular cluster node. See Migrating Resources Dynamically—Overview on page 24-31 for complete information and instructions for performing DARE Resource migrations from the command line and through SMIT.

Synchronizing Configuration Changes

When you change the topology or the resources of a cluster, you update the data stored in the ODM in the DCD. For example, when you add an additional network adapter to a cluster node, you must add the adapter to the cluster definition so that the cluster nodes can recognize and use the adapter. When you change the cluster definition on one cluster node, you must also update the ODMs on the other cluster nodes, a process called synchronization. Synchronization causes the information stored in the DCD on the local cluster node to be copied to the ODM object classes in the DCD on the other cluster nodes.

When synchronizing the cluster triggers a dynamic reconfiguration event, HACMP/ES verifies that both cluster topology and cluster resources are correctly configured, even though you may have only changed an element of one of these. Since a change in topology may invalidate the resource configuration, and vice versa, the software checks both.

To save time when performing this process, you can skip cluster verification during synchronization. For more information on this option, see the section Synchronizing the Cluster Topology on page 24-13.

Dynamic Cluster Topology Changes

You can make the following changes to the cluster topology in an active cluster, dynamically:

- Adding or removing one or more nodes
- Adding or removing one or more network adapters
- Adding, removing, or modifying one or more network modules.

When reconfiguring a cluster, remove any dependencies on a resource group or on resources within a resource group before modifying the cluster topology. For example, if you want to remove a node from the cluster, you must remove the node from all resource groups in which it participates. Similarly, if you want to remove a service adapter, you must remove the resource group which is using the associated service address or remove the service address from the resource group.

Dynamic Cluster Resource Changes

You can make the following changes to cluster resources in an active cluster, dynamically:

- Adding, removing, or changing an application server
- Adding, removing, or changing application monitoring
- Adding or removing one or more resource groups
- Adding or removing the contents of one or more resource groups
- Adding, removing, or changing the order of participating nodes in a resource group
- Changing the node relationship of the resource group
- Toggling the state of the Inactive Takeover attribute.

Other configuration changes, such as changing the *name* of an application server or resource group, require that you stop cluster services before they become active. You can include such a change in a dynamic reconfiguration; however, HACMP/ES interprets these changes, particularly name change, as defining a new cluster component rather than as changing an existing component. Such a change causes HACMP/ES to stop the active component before starting the new component, causing an interruption in service.

Note: You can make changes dynamically to either the topology OR the resource configuration, but not both at the same time.

Requirements before Reconfiguring

Before making changes to a cluster definition, ensure that:

- HACMP/ES is installed on all nodes
- All nodes are up and able to communicate with each other.

Viewing the Cluster Topology

Note: When you view the cluster topology, you are viewing the ODM data stored in the DCD, not the data stored in the ACD.

Before making changes to a cluster topology, view the current configuration.

To view the cluster topology, perform the following procedure.

1. Enter the SMIT fastpath `smit hacmp` and select the following options: **Cluster Configuration > Cluster Topology > Show Cluster Topology**.
2. When you press Enter, SMIT displays the screen with the following options.

   Each option provides a different view of the cluster. Select the appropriate option for the task at hand.

- Select the **Show Cluster Topology** option when you want to obtain complete information about the cluster topology including the nodes in the cluster, their adapters, and the networks that connect them.
- Select the **Show Cluster Definitions** option when you want to obtain the name and ID of all clusters accessible from this node.
- Select the **Show Topology Information by Node** option to obtain information about cluster nodes and their adapters.
- Select the **Show Topology Information by Network Name** option to obtain information about the networks that connect the cluster nodes.
- Select the **Show Topology Information by Network Adapter** option to obtain information about the network adapters defined in the cluster.

Changing a Cluster Name or ID

You must stop cluster services, make the change, and then restart cluster services to apply it to the active configuration. You cannot make these changes dynamically.

To change a cluster's name or ID:

1. Enter the fastpath `smit hacmp` and select the following options: **Cluster Configuration > Cluster Topology > Configure Cluster > Change/Show Cluster Definition**.
2. When you press Enter, SMIT displays the fields that define the cluster definition with their current values filled in.

3. Enter any changes to the values displayed and press Enter. A cluster ID can be any integer less than 99,999. A cluster name can include alphabetic and numeric characters and underscores. Use no more than 31 characters.
4. After the command completes, press the F3 key to return to the HACMP/ES SMIT menus to perform further topology reconfiguration or to synchronize the changes you made. To synchronize the cluster topology, return to the Cluster Topology menu and select the **Synchronize Cluster Topology** option. See page 24-10 for more information.

Changing the Configuration of Cluster Nodes

As the system administrator of an HACMP/ES cluster, you may need to perform any of the following tasks relating to cluster nodes:

- Adding one or more nodes to the cluster
- Removing a node from the cluster
- Changing the attributes of a cluster node

Adding a Cluster Node

You can add a node to an active cluster dynamically. You do not need to stop and restart cluster services for the node to become part of the cluster.

Take the following steps to add a node to the cluster.

1. On any cluster node (called the local node from here on), add the new node to the cluster topology definition.

   Enter the fastpath `smit hacmp`. Select the following options: **Cluster Configuration > Cluster Topology > Configure Nodes > Add Cluster Nodes**.
2. When you press Enter, SMIT displays the **Add a Cluster Node** screen.
3. Enter the name of the node (or nodes) that you want to add to the cluster. A node name can include alphabetic and numeric characters and underscores. Use no more than 31 characters. Separate multiple names with spaces. If you specify a duplicate node name, the operation fails. Press Enter to add the node or nodes to the cluster definition.
4. You must also add a service network adapter for the new node to the cluster definition. A node can also have multiple standby adapters, depending on the specific configuration. For information about adding network adapters, see the section Adding a Network Adapter on page 24-6.
5. On the local node, if you have completed the topology changes you want to make, synchronize the cluster topology definition. Return to the SMIT **Cluster Topology** menu and select the **Synchronize Cluster Topology** option. When you press Enter, the cluster topology is synchronized across the cluster.
6. If desired, on the local node, add the newly added node (or nodes) to the list of participating nodes in a resource group. For information about adding a node to a resource group, see the section Reconfiguring Cluster Resources on page 24-14.

   In a cascading resource group, if you give the new node the highest priority, by specifying it first in the list of participating nodes, the newly added node will acquire control of the resource group when you synchronize cluster resources. This can be useful when you want the new node to take over a specific resource. For example, you may be adding a high-powered node to a cluster that runs a heavily used database application and you want this application to run on the newly added node.

Warning: When adding a node to a cluster with a resource group that has *SSA disk fencing enabled*, add the node to the concurrent resource group immediately. All nodes in an SSA concurrent access cluster must participate in the concurrent access resource group. Include the new node in this resource group immediately to avoid the possibility of unrecoverable data loss.

7. On the local node, synchronize cluster resources. Return to the SMIT **Cluster Resources** menu and select the **Synchronize Cluster Resources** option. When you press Enter, the cluster resources are dynamically reconfigured.
8. On the newly added node, start cluster services to integrate the new node into the cluster.

Removing a Cluster Node

You can remove a node from an active cluster dynamically. Before removing a node from the cluster, however, you must remove the node from any resource groups it participates in, and synchronize resources.

Take the following steps to remove a cluster node:

1. Stop cluster services on the node to be removed.
2. On any other cluster node (hereafter called the local node), remove the node from the resource groups in which it participates. See the section Reconfiguring Cluster Resources on page 24-14 if you need more information.
3. On the local node, once you have made all the changes, synchronize cluster resources. Return to the SMIT **Cluster Resources** menu and select the **Synchronize Cluster Resources** option.
4. On the local node, remove the node from the cluster topology definition.
5. From the main HACMP/ES SMIT screen, select the following options: **Cluster Configuration > Cluster Topology > Configure Nodes > Remove a Cluster Node**. SMIT displays a list of all cluster nodes.
6. Select the node you want to remove and press Enter.

   SMIT displays a popup message asking if you are sure you want to proceed. Press Enter again to remove the node from the cluster topology definition.

   Note: When you remove a node, all adapter information associated with the node is also removed.

7. On the local node, Return to the SMIT **Cluster Topology** menu and select the **Synchronize Cluster Topology** option to synchronize the cluster topology. When the synchronization completes, the node is removed from the cluster topology definition.

Changing the Name of a Cluster Node

You must stop cluster services, make the change, and then restart cluster services to apply it to the active configuration.

To change the name of a cluster node, perform the following procedure.

1. From the Cluster Topology menu, select the **Configure Nodes** option and press Enter.
2. Select the **Change/Show Cluster Node Name** option and press Enter. SMIT displays a pick list of cluster nodes. Make your selection and press Enter.
3. SMIT displays the current node name. Enter the new name for the node in the **New Node Name** field. A node name can include alphabetic and numeric characters and underscores. Use no more than 31 characters. When you finish entering data, press the Enter key. SMIT makes the changes you specified.
4. After the command completes, press the F3 key to return to the HACMP/ES SMIT menus to perform further topology reconfiguration or to synchronize the changes you made. To synchronize the cluster topology, return to the Cluster Topology menu and select the **Synchronize Cluster Topology** option. For more information, see page 24-10.

Manual Updates Required After Node Name Change

Be aware that after changing a node name, the change is propagated through the cluster topology only. Cluster *resources* are not modified, and so will still contain references to the old name. The Participating Node Names field must be updated manually (in the Change/Show a Resource Group SMIT screen) for all resource groups associated with the changed node.

Changing the Configuration of Network Adapters

As a system administrator, you may need to perform any of the following tasks relating to cluster network adapters:

- Adding a network adapter
- Swapping a network adapter dynamically
- Removing a network adapter
- Changing a network adapter

Adding a Network Adapter

You can add a network adapter to an active cluster dynamically. You do not need to stop and restart cluster services for the adapter to become part of the cluster.

1. On the node getting the new adapter, install and configure the new adapter. For information about configuring an adapter, see Chapter 18, Configuring an HACMP/ES Cluster.
2. On all cluster nodes, update the **/etc/hosts** file on all the cluster nodes to include the IP address of the new adapter.
3. On any cluster node (hereafter called the local node), add the network adapter to the cluster topology definition.

   From the main HACMP/ES SMIT screen, select the following options: **Cluster Configuration > Cluster Topology > Configure Adapters > Add an Adapter**.

4. When you press Enter, SMIT displays the **Add an Adapter** screen.

   Enter field values as follows:

Adapter IP Label

Enter the IP label (the name) of the adapter as defined in the **/etc/hosts** file.

If the cluster uses IP address takeover or rotating resources, 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.

The label for an RS232 line must end in the characters *ttyn*, where *n* is the number of the tty device associated with the serial network (for example, *clam_tty1*).

The label for a target mode SCSI-2 bus must end in the characters *tmscsin*, where *n* is the SCSI device number (for example, *clam_tmscsi2*).

The label for a target mode SSA loop must end in the characters *tmssan*, where *n* is the SSA device number (for example, *clam_tmssa2*)

Network Type

Indicate the type of network to which this adapter is connected. Press F4 for a popup pick list of pre-installed network types. The network type for an SP switch must be *hps*.

Network Name

Enter an ASCII text string that identifies the network connected to this adapter. The network name can include alphabetic and numeric characters and underscores. Use no more than 31 characters.

The network name is arbitrary, but must be used consistently. If several adapters share the same logical network or subnet, make sure that 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 can be public networks. SP Switch networks can only be private networks. ATM is a private network. RS232 lines, target mode SSA loops, and target mode SCSI-2 busses are serial networks.

Adapter Function

Indicate whether the adapter's function is service, standby, or boot. Press the TAB key to toggle the values.

A node has a single service adapter for each network. A node can have none, one, or more standby adapters for each public network; however, one or more standby adapters is highly recommended to avoid a single point of failure in your cluster environment. Serial networks and the SP Switch do not have standby adapters.

Adapter Identifier Enter the IP address in dotted decimal format or as 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 local **/etc/hosts** file (using the adapter IP label given).

You must enter device file names for serial network adapters. RS232 serial adapters should have the device file name **/dev/*ttyn***.

Target mode SCSI serial adapters should have the device file name **/dev/tmscsi*n***, where ***n*** should match the adapter label number.

Target mode SSA serial adapters should have the device file name **/dev/tmssa*n***, where ***n*** should match the adapter label number.

Adapter Hardware Address (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

- *And*, the service adapter has a boot address

- *And*, you want to use hardware address swapping.

See Chapter 4, Planning Cluster Network Connectivity, for more information on hardware address swapping.

The hardware address is 12 digits for Ethernet, Token-Ring and FDDI; and 14 digits for ATM.

Node Name (Optional) Only adapters which have addresses which are *always* associated with a particular node need a node name defined. These are service addresses which are always associated with only a single node, boot addresses, and standby addresses.

For adapters which have addresses that are part of a rotating resource group, no node name should be defined. Instead, the event scripts use the user-defined resource group configuration to associate these service addresses with the proper node.

5. Press Enter. The system adds these values to the HACMP/ES ODM.
6. On the local node, synchronize the cluster topology. Return to the SMIT **Cluster Topology** menu and select the **Synchronize Cluster Topology** option. When the synchronization completes, the network adapter is added to the cluster topology definition.

Swapping a Network Adapter Dynamically

As a systems administrator, you may at some point experience a problem with a network adapter card on one of the HACMP/ES cluster nodes. If this occurs, you can use the dynamic adapter swap feature to swap the IP address of an active service or boot adapter with the IP address of an active, available standby adapter on the same node and network. Cluster services do not have to be stopped to perform the swap.

You can use this feature to move an IP address off of an adapter that is behaving erratically, to another standby adapter, without shutting down the node. It can also be used if a hot pluggable adapter device is being replaced on the node. Hot pluggable adapters can be physically removed and replaced without powering off the node.

If hardware address swapping is enabled, the hardware address will be swapped along with the IP address.

Note: The dynamic adapter swap feature is not supported on the SP switch network.

Note: The dynamic adapter swap feature can only be performed within a single node. You cannot swap the IP address of a service or boot address with the IP address of a standby adapter on a different node. To move an IP address to another node, move its resource group using the DARE Resource Migration utility.

To dynamically swap a network adapter, perform the following procedure:

1. Make sure that no other HACMP/ES events are running before swapping an adapter.
2. From the main HACMP/ES SMIT screen, select **Cluster Systems Management > Swap Network Adapter.**

   SMIT displays a list of available service/boot adapters.
3. Select the service/boot adapter you want to remove from cluster use, and press Enter.

   SMIT displays a list of available standby adapters.
4. Select the standby adapter you want, and press Enter.
5. The Swap Network Adapter menu appears. Verify the service/boot IP label, and the standby IP label you have chosen. If this is correct, press Enter.

   A pop-up message asks if you are sure you want to do this operation. Press Enter again *only* if you are sure you want to swap the network adapter.

After the adapter swap, the service/boot address becomes an available standby adapter. At this point, you can take action to repair the faulty adapter. If you have a hot pluggable adapter, you can replace the adapter while the node and cluster services are up. Otherwise, you will have to stop cluster services and power down the node to replace the adapter.

If you have a hot pluggable adapter, HACMP/ES makes the adapter unavailable as a standby when you pull it from the node. When the new adapter card is placed in the node, the adapter is incorporated into the cluster as an available standby again. You can then use the dynamic adapter swap feature again to swap the IP address from the standby back to the original adapter.

If you need to power down the node to replace the faulty adapter, HACMP/ES will configure the service/boot and standby addresses on their original adapters when cluster services are restarted. You do not need to use the dynamic adapter swap feature again to swap the adapters. HACMP/ES does not record the swapped adapter information in the AIX ODM. Therefore, the changes are not persistent across system reboots or cluster restarts.

Removing a Network Adapter from a Cluster Node

You can remove a network adapter from an active cluster dynamically. You do not need to stop and restart cluster services.

Take the following steps to remove a network adapter from a cluster node:

1. On any cluster node, remove the network adapter from the cluster topology definition.

   From the main HACMP/ES SMIT screen, select the options: **Cluster Configuration > Cluster Topology > Configure Adapters > Remove an Adapter**.

2. When you press Enter, SMIT displays the list of defined adapter names.

   Select the adapter you want to remove and press Enter. When you remove an adapter, all information associated with the adapter is removed from the ODM. A pop-up message asks if you are sure you want to do this operation. Press Enter again *only* if you are sure you want to remove the adapter and its associated information.

3. On the same node, synchronize the cluster topology. Return to the SMIT **Cluster Topology** menu and select the **Synchronize Cluster Topology** option. When the synchronization completes, the network adapter is removed from the cluster topology definition.

Changing Network Adapter Attributes

You cannot change the attributes of a network adapter dynamically. You must stop and restart cluster services to make the changed configuration the active configuration.

To change the attributes of a network adapter:

1. On any cluster node, change the attributes of the network adapter.

   From the main HACMP/ES SMIT screen, select the options: **Cluster Configuration > Cluster Topology > Configure Adapters > Change/Show an Adapter**.

   SMIT displays a list of adapter names.

2. Select the adapter to change and press Enter.

   SMIT displays a screen listing all the attributes of the adapter, with their current values filled in.

3. Make any desired changes and press Enter.

4. On the local node, synchronize the cluster topology.

   When the synchronization completes, the network adapter is removed from the cluster topology definition.

   Return to the SMIT **Cluster Topology** menu and select the **Synchronize Cluster Topology** option.

Manual Updates Required After Adapter Name Change

Be aware that after changing an adapter name, the change is propagated through the cluster topology only. Cluster *resources* are not modified, and so will still contain references to the old name. The Service IP Label field (in the Change/Show Resources/Attributes SMIT screen) must be updated manually with the new name for any resource group associated with the changed adapter.

Changing the Configuration of a Global Network

You can group multiple HACMP/ES networks *of the same type* under one logical 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.

To change the definition of a global network, you add or remove existing HACMP/ES networks to or from the global network.

Adding an HACMP/ES Network to a Global Network

To add a network to the definition, 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 any new HACMP/ES networks to be included in each global network.

Removing an HACMP/ES Network from a Global Network

To remove a network from the global network definition, 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 the network to remove 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, along with the name of the global network it currently belongs to. Remove the name of the global network and press Enter.
4. Repeat these steps to remove any other HACMP/ES networks from a global network.

Changing the Configuration of a Network Module

The HACMP/ES SMIT interface allows you to add, remove or change an HACMP/ES network module. You rarely need to add or remove an HACMP/ES network module. However, you may want to tune the topology services by changing the failure detection rate of a network module.

The normal heartbeat rate is usually optimal. 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.

Be sure you have tuned the AIX performance parameters for I/O pacing and **syncd** frequency before changing a network module. See the section Planning for Cluster Performance on page 4-17, in Volume I of this Guide for detailed information.

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

- Failure detections rate is dependent on the *fastest* network linking two nodes.
- 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 changing the failure detection speed on the network modules to **slow** to avoid this problem.
- The failure rate of networks varies, depending on their characteristics. “Fast,” “normal,” and “slow” do not necessarily indicate the same actual rate from network to network.

Changing the Attributes of a Network Module

To change the attributes of a network module:

1. Stop cluster services on all cluster nodes.
2. Select **Configure Network Modules** from the Cluster Topology menu and press Enter.
3. Select the **Change/Show Network Modules** option 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 their current values.

Network Module Name Name of network type, for example, ether.

Description 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 Toggle **Normal, Fast, Slow, Custom**. This tunes the interval between heartbeats for the selected network module.

Failure Cycle 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. Default (Normal) for ether is 2. If you choose **Custom 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 any 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 Topology

Whenever you modify the cluster topology definition in the ODM on one node, you must synchronize the change with the ODM data on all cluster nodes. You perform a synchronization by choosing the **Synchronize Cluster Topology** option from the **Cluster Topology** SMIT screen.

Note: Before synchronizing a cluster configuration, ensure that all nodes are powered on, that the HACMP/ES software is installed, that if you are using the **Standard** cluster security mode, the **/etc/hosts** and **/.rhosts** files on all nodes include all HACMP/ES boot and service IP labels, and that if you are using the **Enhanced** cluster security mode, the **/.klogin** files on all nodes participating in the service have an entry for each service principal configured for Kerberos.

To synchronize changes to the cluster topology across all cluster nodes:

1. From the Cluster Configuration menu, select the option: **Cluster Topology > Synchronize Cluster Topology**. When you press Enter, SMIT displays the following screen.

   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.

2. Specify **yes** or **no** in the **Ignore Cluster Verification Errors.** Set the **Emulate or Actual** field to **Actual**. Set the desired setting in the **Skip Cluster Verification** field.
3. Press enter.

   The cluster definition (including all node, adapter, and network method information) is copied to the other cluster nodes.

Press F10 to exit SMIT or F3 to return to the previous SMIT screen.

Skipping Cluster Verification During Synchronization

Note that to save time during cluster synchronization, you can choose to skip cluster verification. However, cluster verification is optional only when a cluster is *inactive*. If any node is active, **clverify** will run.

You can skip cluster verification using SMIT or at the command line. To skip verification of cluster topology using the SMIT interface, choose **Yes** at the **Skip Cluster Verification** field as explained above.

Releasing a Dynamic Reconfiguration Lock

As described in Chapter 3 in the *HACMP for AIX Concepts and Facilities* manual, during a dynamic reconfiguration, HACMP/ES creates a temporary copy of the HACMP/ES-specific ODM classes and stores them in the Staging Configuration Directory (SCD). This allows you to modify the cluster configuration while a dynamic reconfiguration is in progress. You cannot, however, synchronize the new reconfiguration until the first is finished. The presence of an SCD on any cluster node prevents dynamic reconfiguration. If, because of a node failure or other reason, an SCD remains on a node after a dynamic reconfiguration is finished, it will prevent any further dynamic reconfiguration. Before you can perform further reconfiguration, you must remove this lock.

To remove a dynamic reconfiguration lock, perform the following procedure.

1. Enter the SMIT fastpath:

   ```
   smit hacmp
   ```
2. Select **Cluster Recovery Aids** from the HACMP menu and press Enter.
3. Select the **Release Locks Set By Dynamic Automatic Reconfiguration Event** option and press Enter. SMIT displays a screen asking if you want to proceed. If you want to remove the SCD, press Enter.

Processing ODM Data During Dynamic Reconfiguration

When you synchronize the cluster topology, the processing performed by HACMP/ES varies depending on the status of the Cluster Manager.

The following describe the variations that may occur:

The Cluster Manager is not running on any cluster node

If the Cluster Manager is not running on any cluster node (typically the case when a cluster is first configured), synchronizing the topology causes the configuration data stored on each node reachable from the local node to be updated.

The Cluster Manager is running on the local node

If the Cluster Manager is running on the local node, synchronizing the topology triggers a dynamic reconfiguration event. In processing this event, HACMP/ES updates the configuration data stored on each cluster node that is reachable. Further processing makes the new configuration the currently active configuration.

The Cluster Manager is running on some cluster nodes but not on the local node

If the Cluster Manager is running on some cluster nodes but not on the local node, synchronizing the topology causes the configuration data stored on each node that is reachable from the local node to be updated. However, the processing performed during a dynamic reconfiguration to make the new configuration the active configuration is *not* performed.

Undoing a Dynamic Reconfiguration

Before HACMP/ES overwrites the configuration defined in the ACD, it saves a record of the configuration in a cluster snapshot. Only the**.odm** portion of a cluster snapshot is created; the**.info** file is not created. (For more information about cluster snapshots, see Chapter 26, Saving and Restoring Cluster Configurations.) If you want to undo the dynamic reconfiguration, you can use this cluster snapshot to restore the previous configuration.

HACMP/ES saves snapshots of the last ten configurations in the default cluster snapshot directory, **/usr/es/sbin/cluster/snapshots**, with the name **active.*x*.odm**, where *x* is a digit between 0 and 9, with 0 being the most recent.

Restoring the ODM Data in the DCD

If a dynamic reconfiguration operation fails or is interrupted, you may want to restore the configuration in the DCD with the current active configuration, which is stored in the ACD. HACMP/ES allows you to save in a snapshot the changes you made to the configuration in the DCD before you overwrite it.

To replace the ODM data stored in the DCD with the ODM data in the ACD, perform the following procedure.

1. Enter the SMIT fastpath:

   ```
   smit hacmp
   ```

2. From the main HACMP menu, Select **Cluster Configuration > Restore System Default Configuration from Active Configuration** and press Enter.

3. Enter field values as follows:

   Cluster Snapshot Name of System Default HACMP ODMs
   In this field, you specify the name you want assigned to the cluster snapshot HACMP/ES creates before it overwrites the ODM data stored in the DCD with the ODM data from the ACD. You can use this snapshot to save the configuration changes you made.

   Cluster Snapshot Description of System Default HACMP ODMs
   Enter any text string you want stored at the beginning of the snapshot.

Reconfiguring CS/AIX Communication Links

A CS/AIX communications link contains CS/AIX configuration information for a specific node and network adapter. This information enables an RS/6000 computer to participate in an SNA network that includes mainframes, PCs and other workstations.

Changes to a CS/AIX communications link may involve: changing the DLC name; changing the ports or link stations; or changing the service application start script. You can reconfigure CS/AIX communication links using the SMIT interface.

Adding a CS/AIX Communications Link

You can define a CS/AIX communications link dynamically. You can define a CS/AIX communications link and add it to a resource group in a single dynamic reconfiguration operation.

Complete the following steps to create a highly available CS/AIX communications link on any cluster node:

1. Type:

   ```
   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. 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. Select **Make Communications Links Highly Available > Add a Highly Available Communications Link**
4. 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. This field is optional.

Link Station Enter ASCII text string for the names of any CS/AIX link stations. This field is optional.

Service Enter the full pathname of any application start script. This start script starts any application layer processes that use the communications link. This field is optional.

5. Press Enter to add this information to the ODM on the local node.
6. Once you have defined a highly available CS/AIX communications link, you must configure it as a resource in a resource group. See the section, Adding or Removing Individual Resources in this chapter for information on adding a CS/AIX communications link to a resource group.
7. Press F3 key to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes to the cluster you just made. To synchronize the cluster definition, go to the Cluster Resources SMIT screen and select the Synchronize Cluster Resources option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Configuration Changes on page 24-1.

Removing a CS/AIX Communications Link

You can remove a CS/AIX communications link from an active cluster dynamically. Before removing a CS/AIX communications link, you must remove it from any resource group where it is included as a resource. For more information, see the section Reconfiguring Cluster Resources and Resource Groups on page 24-23.

To remove a CS/AIX communications link, complete the following steps:

1. In SMIT, go to the **Make Communications Links Highly Available** screen. From this menu, the select the **Remove a Highly Available Communications Link** option and press Enter.
2. Select the communications link you want to remove and press Enter.

   A message appears asking if you are sure you want to remove the communications link.
3. Press Enter again to confirm the removal.

   The server is removed from the ODM object classes stored in the DCD on the local node.
4. Press F10 to exit SMIT or press F3 to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Configuration Changes on page 24-1.

Changing a CS/AIX Communications Link

To change (or view) a CS/AIX communications link, complete the following steps:

1. In SMIT, go to the **Make Communications Links Highly Available** screen. From this menu, the select the **Change/Show a Highly Available Communications Link** option and press Enter.

   SMIT displays a list of communications links.

2. Select the communications link you want to change and press Enter. The **Change Communications Link** screen appears, with the server name filled in.
3. Make the desired changes on this screen.

   DLC Name Make any changes to the CS/AIX communications link name.

   Port Make any changes to the CS/AIX ports to be started automatically.

   Link Station Make any changes to the CS/AIX link stations.

   Service Make an changes to the application start script. This is the full pathname of the start script that starts any application layer processes that use the communications link.

4. Press Enter to add this information to the ODM stored in the DCD on the local node.
5. Press F10 to exit SMIT or press F3 to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Configuration Changes on page 24-1.

Reconfiguring 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 or rotating resource.
- Points the cluster event scripts to the scripts that they call to start and stop the server application.
- Makes it possible to configure an application monitoring method for the application.

Note that this section does not discuss how to write 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 or to synchronize the changes to the cluster you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Cluster Resources on page 24-30.

Removing an Application Server

You can remove an application server from an active cluster dynamically. Before removing an application server, you must remove it from any resource group where it is included as a resource. For more information, see the section Adding Resources to or Removing Them from a Resource Group on page 24-16.

To remove an application server, complete the following steps:

1. In SMIT, go to the **Define Application Servers** screen, as described in the previous section. From this menu, select the **Remove an Application Server** option and press Enter. SMIT displays the list of application servers.
2. Select the server you want to remove and press Enter. A message appears asking if you are sure you want to remove the server. Press Enter again to confirm the removal. The server is removed from the ODM object classes stored in the DCD on the local node.
3. Press F10 to exit SMIT or press F3 to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Cluster Resources on page 24-40.

Changing an Application Server

When you specify new start or stop scripts to be associated with an application server, the ODM data is updated but the application server is not configured or unconfigured dynamically; thus the application controlled by the application server is not stopped and restarted. The next time the application is stopped, HACMP/ES will call the new stop script—not the stop script that was defined when the application server was originally started.

Note: If you make a change to an application server that has an application monitor defined, you must make the change *separately* to the application monitor as well. Changes to application server information are not automatically communicated to the application monitor configuration.

To change (or view) an application server, complete the following steps:

1. In SMIT, go to the **Define Application Server** screen. From this menu, select the **Change/Show an Application Server** option and press Enter. SMIT displays the application servers.
2. Select the application server you want to change and press Enter. The **Change/Show an Application Server** screen appears, with the server name filled in.
3. Make the desired changes on this screen.

   New Server Name Enter the new name that will identify the server you want to change. This field initially contains the name of the application server you selected. You may use alphabetic or numeric characters and underscores in the name. Use no more than 31 characters.

   Note: Attempting to change the name dynamically will cause the application server to be stopped and restarted.

Start Script Enter the pathname of the script 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. Note that this field initially contains the path to the start script for the application server you selected.

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. Note that this field initially contains the path to the stop script for the application server you selected.

4. Press Enter to add this information to the ODM stored in the DCD on the local node.
5. Press F10 to exit SMIT or press F3 to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see page 24-30.

Changing or Removing Application Monitors

If you have configured application monitoring, you may wish to suspend or remove the monitor at some point, or to change some aspect of the monitoring you have set up, for instance the processes to be monitored, the scripts to run, or the notify, cleanup, or restart methods.

Note: This section discusses changing an existing application monitor. For information about adding a new application monitor, see Chapter 18, Configuring an HACMP/ES Cluster, in Volume 1 of this guide.

Suspending and Resuming Application Monitoring

You can suspend the monitoring of a specified application while the cluster is running. The suspension of monitoring is temporary. If a cluster event occurs that results in the affected resource group moving to a different node, application monitoring resumes automatically on the new node. Similarly, if a node is brought down gracefully without takeover and then restarted, monitoring will resume automatically.

Note: To permanently stop monitoring of an application, you must perform the steps in the section Removing an Application Monitor below.

To temporarily suspend application monitoring, choose **Cluster System Management > Suspend/Resume Application Monitoring > Suspend Application Monitoring**. You are prompted to choose which application monitor you want to suspend. The monitor will be suspended until you choose to resume it, or until a cluster event occurs to resume it automatically, as explained above.

To resume monitoring after suspending it, take the same path in SMIT: **Cluster System Management > Suspend/Resume Application Monitoring > Resume Application Monitoring**. You are prompted to choose which suspended application monitor you want to resume.The monitor resumes, configured as it was prior to suspending it.

Note: Do not make changes to the application monitor configuration while it is in a suspended state.

Changing the Configuration of an Application Monitor

You can change the configuration details of an application monitor by simply editing the SMIT fields you defined when you configured the monitor initially.

Note: When you configured application monitors originally, the Restart Method and Cleanup Method fields had default values. If you changed those fields, and now want to change back to the defaults, you must enter the information manually (by copying the scripts from the Change/Show an Application Server SMIT screen).

To alter any facet of an application monitor you have defined, take the following steps.

1. Type `smit hacmp` to reach the SMIT main HACMP menu.
2. Choose **Cluster Configuration > Cluster Resources > Configure Application Monitoring.**
3. Depending on which type of monitor you are altering, choose either:

   Define Process Application Monitor > Change/Show Process Application Monitor

   or

   Define Custom Application Monitor > Change/Show Custom Application Monitor.
4. From the list of monitors, choose the previously defined application monitor you want to change.
5. Make any desired changes in the SMIT screen fields and press Enter.

   Remember that default values are not restored automatically.

   For details on how to fill in the information for each field, see the instructions in Configuring a Process Application Monitor on page 18-19 in Volume 1 of this guide. Also press F1 for online help.

   The changes you enter take effect the next time the resource group containing the application is restarted.

Removing an Application Monitor

To permanently remove an application monitor:

1. In SMIT, choose **Cluster Configuration > Cluster Resources > Configure Application Monitoring.**
2. Depending on which type of monitor you are altering, choose either:

   Define Process Application Monitor > Remove a Process Application Monitor

   or

   Define Custom Application Monitor > Remove a Custom Application Monitor.

3. Select the monitor to remove.
4. Press enter.

   The selected monitor is deleted.

Note: If the monitor is currently running, it will not be stopped until the next DARE activity or synchronization occurs.

Reconfiguring Cluster Resources and Resource Groups

When you initially configured your HACMP/ES system, you defined each resource as part of a resource group. This allows you to combine related resources into a single logical entity for easier configuration and management. You then configured each resource group to have a particular kind of relationship with a set of nodes. Depending on this relationship, resources were defined as either cascading or rotating. You also assigned a *priority* to each participating node in a cascading resource group chain.

To change the nodes associated with a given resource group or to change the priorities assigned to the nodes in a resource group chain, you must redefine the resource group. You must also redefine the resource group if you add or change a resource assigned to the group.

This section describes how to view, change, add, and delete a resource group. For more information about the initial configuration of cluster resources, see Chapter 18, Configuring an HACMP/ES Cluster.

Adding a Resource Group

You can add a resource group to an active cluster. You do not need to stop and then restart cluster services for the resource group to become part of the current cluster configuration.

To add a resource group, perform the following procedure:

1. From the Cluster Resources menu, select **Define Resource Groups** and press Enter.
2. Select Add a Resource Group and press Enter. SMIT displays the **Add a Resource Group** screen.
3. Enter field values as follows:

   Resource Group Name Enter an ASCII text string that identifies the resource group. You can use alphabetic or numeric characters and underscores. Use no more than 31 characters. If you specify the name of an existing resource group, the command fails.

   Node Relationship Toggle the entry field to select between Cascading, Rotating, or Concurrent.

   Participating Node Names Enter the names of the nodes that you want to be members of the resource chain for this resource group. Enter the node names in order from highest to lowest priority (left to right). Leave a space between node names.

4. Press Enter to add the resource group information to the ODM on the local node.
5. Press F10 to exit SMIT or press F3 to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Cluster Resources on page 24-30.

Removing a Resource Group

You can remove a resource group from an active cluster. You do not need to stop and then restart cluster services for the resource group to be removed from the current cluster configuration.

To remove a resource group, perform the following procedure.

1. From the Cluster Resources menu, select **Define Resource Groups** and then **Remove a Resource Group** and press Enter. SMIT displays a screen listing the defined resource groups.
2. Select the resource group you want to remove and press Enter.

   SMIT displays a popup warning, reminding you that all information about the resource group will be lost.
3. Press Enter again to confirm your action.
4. Press F10 to exit SMIT or press F3 to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Cluster Resources on page 24-30.

Changing a Resource Group

Using the SMIT **Change/Show Cluster Resource** screen, you can perform any of the following tasks:

- Change the name of the resource group
- Add or remove nodes from the list of participating nodes
- Change the priority of participating nodes (by changing their position in the list of participating nodes)
- Change the relationship of the nodes to the resource group.

You can change many of the attributes of a resource group in an active cluster without having to stop and then restart cluster services. However, to change the name of a resource group, you must stop and then restart the cluster to make the change part of the current cluster configuration.

1. Enter the following SMIT fastpath `smit hacmp` and select the following options: **Cluster Configuration > Cluster Resources > Define Resource Groups > Change/Show a Resource Group**.

   SMIT displays a list of the currently defined resource groups. Select the one you want to change and press the Enter key.

2. SMIT displays a screen containing the attributes of the resource group with their values.

3. Enter field values as follows:

   Resource Group Name This field is filled in from the previous menu selection.

   New Resource Group Name Enter the new name you want assigned to this resource group. Use no more than 31 characters. You can use alphabetic or numeric characters and underscores. Duplicate entries are not allowed.

   Note: HACMP/ES interprets a name change as the inclusion of a new component, rather than as a change to an existing component, which will cause the resource group to be unconfigured.

   Node Relationship Toggle the entry field to select between Cascading, Concurrent, or Rotating.

   Note: Do not change a resource group from cascading to rotating in an active cluster.

   Participating Node Names Use this field to enter the names of the nodes that you want to add the resource group or remove from the resource group. You can also change the priority of the nodes in the resource group chain by changing the order in which they are specified. You enter node names in order from highest to lowest priority (left to right). Leave a space between node names.

4. Press Enter to change the resource group information stored in the ODM.

5. Press F10 to exit SMIT or press F3 to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

   If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Cluster Resources on page 24-30.

Adding or Removing Individual Resources

You can add a resource to or remove a resource from a resource group in an active cluster without having to stop and restart cluster services to apply the change to the current configuration. You can add or remove resources from a resource groups even if a node is powered down; however, when a node is powered up, you can obtain a list of possible shared resources for each field by pressing the F4 key.

Resource groups can contain many different types of cluster resources, including IP labels, filesystems, volume groups and application servers. You can change the mix of resources in a resource group and the settings of other cluster resource attributes by using the SMIT **Change/Show Resources for a Resource Group** screen.

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 in the resource group *before* configuring Fast Connect resources.

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:

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 the AIX Connections Resources field and remove all specified realm/service pairs.
5. Select Fast Connect Services and specify the resources you wish to configure in the resource group. If you are specifying Fast Connect fileshares, make sure 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. You’ll find instructions for synchronization at the end of this chapter, beginning on page 24-40.

Using SMIT to Reconfigure Resources

To change the resources in a resource group, perform the following procedure.

1. From the Cluster Resources menu, select the **Change/Show Resources/Attributes for a Resource Group** option and press Enter. SMIT displays a pick list of resource groups from which you must choose.
2. Select the resource group you want to change and press Enter. SMIT displays a screen that lists all the types of resources that can be added to a resource group, with their current values.

Note: If you specify filesystems to NFS mount in a cascading resource group, 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.

3. Enter the field values as follows:

Resource Group Name, Node Relationship, Participating Node Names These fields are not editable.

Service IP Label If IP address takeover is being used, list the IP label 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.

When you enter a file system in this field, the HACMP/ES system determines the correct values for the **Volume Groups** and **Disks** fields.

Filesystems Consistency Check Identify the method of checking consistency of file systems, fsck (default) or logredo (for fast recovery).

Filesystems Recovery Method Identify the recovery method for the file systems, parallel (for fast recovery) or sequential (default).

Filesystems/Directories to Export Identify the filesystems or directories to be exported. These should be a subset of the filesystems listed above. Press F4 for a list.

Filesystems/Directories to NFS Mount 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.

Network for NFS Mount

(This field is optional.)

Choose a previously defined IP network where you want to NFS mount the filesystems. The F4 key lists valid networks.

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.

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.

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.

Volume Groups

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.

If you have previously entered values in the Filesystems field, the appropriate volume groups are already known to the HACMP/ES system.

If you are 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.

Concurrent Volume Groups

Identify the shared volume groups that can be accessed by nodes in concurrent mode. Press F4 to see a list of concurrent volume groups.

Raw Disk PVIDs

Press F4 for a listing of the PVIDs and associated hdisk device names.

If you have previously entered values in the **Filesystems** or **Volume groups** fields, the appropriate disks are already known to the HACMP/ES system.

If you are using an application that directly accesses raw disks, list the raw disks here.

AIX Connections Services

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. *Note that you cannot configure both AIX Connections and AIX Fast Connect in the same resource group.*

AIX Fast Connect Resources

Press F4 to choose from a list of Fast Connect resources that are common to all nodes in the resource group, as 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. *Note that you cannot configure both AIX Connections and AIX Fast Connect in the same resource group.*

Application Servers

Indicate the application servers to include in the resource group. Press F4 to see a list of application servers. See Chapter 18, Configuring an HACMP/ES Cluster, for information on defining application servers.

Highly Available Communications Links

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.

Miscellaneous Data

A text string you want to place into the environment along with the resource group information. This string is accessible by the HACMP/ES scripts.

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 *fallover* and *fallback* in the section, Defining the Node Relationship of Your Cluster on page 3-6.

When the CWOF variable is set to **false**, a cascading resource group will fallback as a node of higher priority joins or reintegrates into the cluster.

When CWOF 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 it reintegrates into the cluster.

The default for CWOF is **false**.

SSA Disk Fencing Activated

Set according to your configuration (concurrent only).

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. After entering data in the fields you want to change, press the Enter key.
5. Press F10 to exit SMIT or press F3 to return to previous SMIT screens to perform other configuration tasks or to synchronize the changes you just made. To synchronize the cluster definition, go to the **Cluster Resources** SMIT screen and select the **Synchronize Cluster Resources** option.

If the Cluster Manager is running on the local node, synchronizing cluster resources triggers a dynamic reconfiguration event. For more information, see the section Synchronizing Cluster Resources on page 24-40.

Migrating Resources Dynamically—Overview

The Dynamic Reconfiguration (DARE) Resource Migration utility allows you to change the status or location of a resource group without stopping cluster services. This utility provides improved cluster management by allowing you to:

- Bring a resource group up/online (this option also restores a resource group to its default setting, removing any sticky markers. Sticky and non-sticky designations are discussed later in this chapter.)
- Bring a resource group down/offline
- Move a resource group to a new location
- Perform maintenance on a node without losing access to the node's resources
- Relocate resource groups for enhanced performance.

Using **cldare** from the command line you can change the status or location of multiple resource groups on multiple nodes at the same time; with the SMIT interface, you can perform one migration at a time. You can also disable resource groups dynamically, preventing them from being acquired during a reintegration. This disabling option allows a "swap" of resources in certain situations. (See page 24-35 for an example of such a situation.)

The rest of this chapter covers:

- Performing two types of migrations: sticky and non-sticky
- Using the **cldare** command and its associated keywords and arguments to perform dynamic migrations from the command line
- Using the SMIT interface to perform dynamic migrations
- Using **clfindres** to check resource group state and presence of sticky markers
- Removing sticky markers with the cluster up or down
- Synchronizing cluster resources.

Resource Migration Types: Sticky and Non-sticky

Before performing a resource migration, you must choose to declare the migration *sticky* or *non-sticky*.

Sticky Resource Migration

The sticky designation supersedes the defined node priority of a resource group. A sticky migration to a specified node ("assigning the sticky attribute to a node") moves the resource group to that node and gives it highest ownership priority of the resource group. Only one sticky attribute can be set in each resource group. The sticky attribute persists until a new resource group migration is done. Thus, it persists even after node failures and a cluster stop. In a subsequent migration request, the user can reassign the sticky designation to another node, or remove it by choosing the non-sticky option. When the sticky attribute is removed, the default node hierarchy for that resource group is reinstated.

A sticky attribute in a resource group affects the fallover and fallback behavior, as well as the initial acquisition of a resource group during the start of the cluster, by changing the default node hierarchy. Despite this new node hierarchy, the resource group behavior follows the rules specified by CWOF and Inactive Takeover flags. For example, in a cascading resource group with CWOF set to **false**, after the reintegration of a node with a sticky attribute, the resource group will fallback onto that node. Furthermore, when the cluster starts and the Inactive Takeover flag is set to **false**, the first node up will acquire the resource group only if it is the sticky node.

As always in a rotating resource group, a node which has the highest ownership priority for the resource group will show preference for that group by trying to acquire it. The resource group will be acquired by a sticky node if the node which owns it fails.

If the sticky node fails, the resource group will fall over to the next highest priority node.

Non-Sticky Resource Migration

A non-sticky resource group migration is the default case when no sticky attribute is specified, and is allowed for rotating and Cascading without Fallback resource groups. Non-sticky resource groups are temporarily placed on the specified node until the next fallover or reintegration occurs and all resource group location policies are reevaluated.

A cascading resource group with the CWOF flag set to **false** cannot be moved by a non-sticky migration since it always resides on the highest priority node which is active. Thus, non-sticky migrations are supported only for cascading resource groups with CWOF = **true**. A rotating resource group immediately resumes a normal rotating resource group fallover policy after a non-sticky migration, but from the new location.

Note: You may find it helpful to perform a non-sticky DARE migration when doing maintenance on a CWOF owner node. Should the default node fail, the CWOF resource group will return to the owner node if it is available.

Note: You cannot mix *node locations*, *default* and *stop* keywords within a single DARE resource migration.

Migrating Resources Dynamically From the Command Line

This section explains how to perform DARE migrations at the command line using **cldare**. Instructions on using the SMIT interface to do this begin on page 24-36. Before performing either method of DARE migration, you should read the preceding overview sections.

The general syntax for migrating, starting, or stopping resource groups dynamically from the command line is as follows:

```
cldare -M <resource group name>:[location|keyword][:sticky] ...
```

-M	Specifies a migration action.
Resource Group Name	The name of the resource group to move or bring online.

Location (Node Name)

To move a resource group or to bring a resource group online on a specific node, you enter a node name in the *location* field to specify the target node where the resource group will be relocated or started.

The node name must be in the resource group's participating node list and the node must be available.

Default Keyword

If you use the *default* keyword to move or start a resource group, the DARE Resource Migration utility removes all previous stickiness for the resource group and returns the resource group to its default fallover behavior where normal configured node priorities apply.

- A cascading resource group will migrate to the highest priority node currently up.
- A rotating resource group will be released from wherever it resides and the highest priority node with a boot address will reacquire the resources.

If you include neither of the location keywords (*location, default)* in the field after the first colon, the DARE Resource Migration utility performs a *default* migration, again making the resources available for reacquisition. You can use this action to start a cascading resource group that has INACTIVE_TAKEOVER set to **false** and that has not yet started because its primary node is down.

Stop Keyword

Using the keyword *stop* deactivates the resource group. If you also use the *sticky* keyword, the resources remain unavailable for reacquisition even after a fallover or reintegration. See Stopping a Resource Group on page 24-34 for more information about using the *stop* keyword.

Sticky

When moving or stopping a resource group, you can make the group *sticky* or non-sticky. Non-sticky is the default. You must specify *sticky* if you want this type.

Repeat this syntax on the command line for each resource group you want to migrate. Do not include spaces between arguments.

For full information on the **cldare** command and all of its associated flags, see the **cldare** man page.

Note: The **cldare** command attempts to perform all requested migrations simultaneously. If, for some reason, the command cannot simultaneously cause all specified resources to be released and reacquired at the new locations, it fails and no migrations occur.

Moving or Starting a Resource Group

Use the **cldare** command with the **-M** flag and a node location specified after the first colon to move a resource group from one node to another or start the resource group if it was previously inactive. The syntax for moving or starting a resource group is as follows:

```
cldare -M <resgroup name>:<target node> [:sticky]
```

You can specify multiple move commands on the command line. The **cldare** utility first parses and checks the command line and verifies the cluster configuration; then the resources are released and reacquired by the specified cluster node. The resource migration process first releases all specified resource groups (wherever they reside in the cluster), then causes the newly specified node to acquire the resource groups.

Attaching the *sticky* designation causes the resource group to attempt to return to the specified node after fallover and reintegration, superseding the normal resource policy.

Stopping a Resource Group

The DARE facility provides a command to stop a resource group; that is, to bring that resource group offline. You can stop a resource group using a sticky or non-sticky option. The former is referred to as a DARE sticky stop, the latter as a DARE stop. A resource group which has been brought offline via a DARE stop remains offline for as long as no cluster event occurs. When a cluster event does occur, the state of the resource group is undefined. It may be brought online or it may stay offline. If you want to bring a resource group permanently offline, use a DARE sticky stop. A sticky-stopped resource group will stay offline until a new DARE migration request for that resource group is issued, which brings it online.

To stop a resource group, use the keyword *stop* by itself, with no location argument, as follows:

```
cldare -M <resgroup name>:stop[:sticky]
```

The DARE Resource Migration utility brings the resource group offline, which includes taking down any service label, unmounting filesystems, etc.

As with sticky *location*s, sticky *stop* requests are in effect until they are:

- superseded by new sticky migration requests for the same resource group
- removed by *default*, non-sticky migration requests for the same resource group.

Note: The sticky designation is optional, but normally you will want to use it with the *stop* keyword. Be careful when using a non-sticky *stop* request, since the resource group may be restarted at the next cluster event. As a result, all non-sticky requests produce warning messages. A non-sticky *stop* can be used to halt a cascading resource group that has INACTIVE_TAKEOVER set to **false** during periods in which its primary node is down.

Special Considerations when Stopping a Resource Group

After DARE stopping a resource group, you should not assume that a joining or rejoining node will bring that resource group online. The following are instances when a resource group must be brought back online with a DARE command.

1. If you use **cldare** *stop* to bring down a cascading resource group (with CWOF = **false**) which is assigned an Inactive Takeover value of **false**, and which resides on the highest priority node, it will remain in an inactive state. You must manually bring the resource group online through DARE migration.
2. 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 **cldare** *stop* command. If Inactive Takeover is set to **false** in such a situation, then the resource group will not be acquired by another node. Unless you bring the resource group up manually, it will remain in an inactive state.
3. 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.

Starting a Resource Group with the Default Keyword

To start a resource group using the **cldare** command with the *default* keyword after the first colon, enter the command as follows:

```
cldare -M <resgroup name>:default [:sticky]
```

You need not specify a target node; the resource group is activated on the node that has been designated as its highest priority node.

You can use the *default* keyword when you want to restore a resource group to its original state, for example to remove an earlier sticky designation.

You can attach the *sticky* keyword to this start command if you want the resource group to stay on that node after a failure or reintegration of another cluster node.

Note: When using the SMIT interface to bring a resource group online, *sticky* is not an option. The resource group will fallover to other nodes just as the resource policy dictates.

Example: Using cldare to Swap Resource Groups

In the three-node cluster indicated here, each node—Node1, Node2, and Node3—has a service adapter and a standby adapter.

There are three cascading resource groups with node priority lists as follows:

RG1 - Node1, Node3	**CrucialRG - Node2, Node3**	**RG3 - Node3, Node1**

Each node is up and possesses a resource group as follows:

Node1 - UP (RG1)	**Node2 - UP (CrucialRG)**	**Node3 - UP (RG3)**

Suppose Node2's resources—contained in CrucialRG—are of particular importance to your operation. A situation occurs in which two cluster nodes fail. Node1 fails first; its resources fall over to Node3, since Node3 is in RG1's priority list. Then Node2 fails. In this case, Node2's crucial resources remain down; they have nowhere to go, since Node3's only standby adapter has already been taken. The cluster now looks like this:

Node1 - DOWN	**Node2 - DOWN**	**Node3 - UP (RG3, RG1)**

The crucial resource group is unavailable. HACMP/ES is able to take care of only one failure, because there are no more standby adapters, so it handles the first failure, Node1, but not the second. However, if you need CrucialRG's resources more than you need RG1's, you can use the DARE resource migration utility to "swap" the resource groups so you can access CrucialRG instead of RG1.

You do this by issuing a *stop* command to bring RG1 offline, then issue a second command to bring CrucialRG online on Node3, as follows:

```
cldare -M rg1:stop:sticky -M crucialrg:node3
```

Note that both commands can be typed on the same line with no space between them.

After these DARE migration commands are completed, access to CrucialRG is restored, and the cluster looks like this:

Node1 - DOWN	**Node2 - DOWN**	**Node3 - UP (RG3, CrucialRG)**

Migrating Resources Dynamically Using SMIT

As of the HACMP/ES 4.3.1 release, you access the DARE migration functions using the SMIT interface instead of the command line.

From the main SMIT hacmp menu, choose **Cluster System Management > Cluster Resource Group Management**.

Here SMIT presents the following options for dynamic resource migration.

Bring a Resource Group Online
This is equivalent to using the **cldare** *default* keyword to start a resource group. Choose this to activate all resources in a specified resource group on the node designated as highest priority for that resource group. You can also use this option to return a resource group to its default setting, removing any previously set sticky designation.

Bring a Resource Group Offline
This is equivalent to the **cldare** *stop* keyword. Use this option to deactivate all resources in a specified resource group. You can choose sticky or normal migration.

Move a Resource Group
This option allows you to deactivate a resource group on one node and then activate it on another node. You can choose sticky or normal migration.

For each migration option, SMIT offers the choice to emulate the migration rather than actually perform it.

Bringing a Resource Group Online with SMIT

Bringing a resource group online is the equivalent of using the *default* keyword with the **cldare** command. This command activates (starts) the specified resource group on the node identified as its highest priority node, or moves the resource group back to its initially designated node, removing any previously assigned stickiness.

1. From the main haes menu, choose **Cluster System Management > Cluster Resource Group Management > Bring a Resource Group Online**.
2. Select the resource group from the picklist and press Enter.
3. The **Bring a Resource Group Online** screen appears. Press the Tab key to toggle through valid entries, or press F4 to view and choose from a picklist of entries. You cannot add entries manually in any of the fields. Note that when using SMIT, you do not have the option of declaring the migration sticky when you bring a resource group online, as you do when using the command line.

Enter field values as follows:

Resource Group to Bring Online
Shows the name of the resource group that you have chosen to activate. You cannot edit this field here.

Emulate or Actual?
Choose **Actual** (the default) to bring the resource group online. Choosing **Emulate** to run the command in a mode that does not make any changes to your cluster, but still displays the results/messages as if the action was being carried out.

Perform Verification First?
Specify whether you want to run the **clverify** utility to verify configuration before performing the migration. The default is **Yes**. It is recommended that any prospective configuration changes be verified before execution. However, if you are certain that the new configuration is appropriate, you can skip the verification step to quicken the process.

Ignore Cluster Verification Errors?
You can choose to have the DARE action continue even if configuration errors are detected during verification. The default is **No**. It is recommended that any prospective configuration changes be verified before execution. However, there may be instances when it is appropriate to proceed despite verification failure. If this is the case, choose **Yes**.

To return to the Cluster Resource Group Management menu, press F3 at any time.

Note: When using SMIT to bring a resource group online, *sticky* is not an option. The resource group will fallover to other nodes just as the resource policy dictates.

Bringing a Resource Group Offline with SMIT

Bringing a resource group offline is equivalent to using the **cldare** command and the *stop* keyword in the location field. This action deactivates, or stops, a resource group.

Take the following steps to bring a resource group offline.

1. From the main haes menu, choose **Cluster System Management > Cluster Resource Group Management > Bring a Resource Group Offline**.
2. Select a resource group from the picklist.

   The **Bring a Resource Group Offline** screen appears.
3. To bring the resource group offline, enter field values as follows:

Resource Group to Bring Offline
Indicates the Resource Group you have selected. This field cannot be edited here.

Use Sticky Migration?
Specify sticky or normal (nonsticky) migration. The default is **No**, causing a nonsticky migration. If you choose **yes**, the DARE special migration location *stop* becomes the highest-priority location for that resource group (until the sticky migration is turned off). The resource group attempts to remain offline during a cluster fallover or reintegration. If you choose **No**, the resource group will be stopped, but its highest-priority location will remain unchanged, and the resource group may be restarted at the time of the next cluster event.

Emulate or Actual?
Choose **Actual** to bring the resource group online. Choosing **Emulate** to run the command in a mode that does not make any changes to your cluster, but still displays the results/messages as if the action was being carried out.

Perform Verification First?
Specify whether you want to run the **clverify** utility before performing the migration. The default is **Yes**. It is recommended that any prospective configuration changes be verified before execution. However, if you are certain that the new configuration is appropriate, the verification step may be skipped to quicken the process.

Ignore Cluster Verification Errors?
You can choose to have the DARE action continue even if errors are detected during verification. The default is **No**. It is recommended that any prospective configuration changes should be verified before execution. However, there may be instances when it is appropriate to proceed despite verification failure. If this is the case, choose **Yes**.

Press F3 at any time to go back to the Cluster Resource Group Management menu.

Moving a Resource Group

Take the following steps to move a resource group.

1. From the main haes menu, choose **Cluster System Management > Cluster Resource Group Management > Move a Resource Group**.
2. Select a resource group from the picklist. The **Move a Resource Group** screen appears.
3. Enter field values as follows:

Resource Group To Be Moved
Indicates the Resource Group you have selected. You cannot edit this field here.

Move Resource Group To Which Node?
Enter the name of the target node to which you are moving the resource group; press F4 for a picklist of the cluster nodes participating in that resource group to which the resource group can migrate—the list does not include the node the resource group is presently on. The target node you specify acquires the resource group.

If you leave this field blank or enter an invalid node name in this field, an error message is displayed.

Use Sticky Migration?
Specify sticky or normal (non-sticky) migration. The default is **No**, causing a nonsticky migration. If you choose **Yes**, the specified target node becomes the highest-priority node for that resource group (until the sticky migration is turned off). The resource group will attempt to remain on that target node—regardless of that node's original priority for the resource group—during a cluster fallover or reintegration.

Emulate or Actual?
Choose **Actual** to bring the resource group online. Choosing **Emulate** to run the command in a mode that does not make any changes to your cluster, but still displays the results/messages as if the action was being carried out.

Perform Verification First?
Specify whether you want to run the **clverify** utility before performing the migration. The default is **Yes**. It is recommended that any prospective configuration changes be verified before execution. However, if you are certain that the new configuration is appropriate, the verification step may be skipped to quicken the process.

Ignore Cluster Verification Errors?
You can choose to have the DARE action continue even if errors are detected during verification. The default is **No**. It is recommended that any prospective configuration changes should be verified before execution. However, there may be instances when it is appropriate to proceed despite verification failure. If this is the case, choose **Yes**.

Press F3 at any time to go back to the Cluster Resource Group Management menu.

Checking Resource Group State and Sticky Markers

The following information will be of use when you wish to check the state of your cluster resource groups at times, and identify if they have sticky markers persisting from earlier resource migrations.

The clfindres Command

To help you check the location and state of resources placed on a specific node, the DARE Resource Migration utility includes a command, **clfindres**, that displays the state and location of specified resource groups. It also indicates whether a resource group has a sticky location, and it identifies that location.

See Appendix A, HACMP for AIX Commands in the *HACMP for AIX Administration Guide* for the complete syntax and typical output of the **clfindres** command.

Removing Sticky Markers When the Cluster is Down

Sticky location markers are stored in the HACMPresource class in the HACMP/ES ODM and are a persistent cluster attribute. While the cluster is *up*, you can remove sticky designations only by performing a subsequent non-sticky migration on the same resource group, using the *default* keyword or specifying no location.

Be aware that persistent sticky location markers are saved and restored in cluster snapshots (discussed in Chapter 26, Saving and Restoring Cluster Configurations), and can cause confusion if the user does not realize they are there. You can use the **clfindres** command described above to find out if sticky markers are present in a resource group.

If you want to remove sticky location markers while the cluster is *down*, do not use the *default* keyword, since it implies activating the resource. Instead, when the cluster is down, use a non-sticky **stop** request, as in this example:

```
cldare -v -M <resgroup name>:stop
```

(The optional -v flag indicates to skip verification.)

Synchronizing Cluster Resources

Whenever you modify the configuration of cluster resources in the ODM on one node, you must synchronize the change across all cluster nodes. You perform a synchronization by choosing the **Synchronize Cluster Resources** option from the Cluster Resources SMIT screen.

The processing performed in synchronization varies depending on whether the cluster manager is active on the local node:

- If the cluster manager is *not* active on the local node when you select this option, the ODM data in the DCD on the local node is copied to the ODMs stored in the DCDs on all cluster nodes.
- If the cluster manager is active on the local node, synchronization triggers a cluster-wide, dynamic reconfiguration event. In dynamic reconfiguration, the configuration data stored in the DCD is updated on each cluster node and, in addition, the new ODM data replaces the ODM data stored in the ACD on each cluster node. The cluster daemons are refreshed

and the new configuration becomes the active configuration. In the HACMP/ES log file, **reconfig_resource_release**, **reconfig_resource_acquire**, and **reconfig_resource_complete** events mark the progress of the dynamic reconfiguration.

- If the cluster manager is active on some cluster nodes but not on the local node, the synchronization is aborted.

To synchronize changes to the cluster resources across all cluster nodes, perform the following procedure.

1. From the Cluster Configuration menu, select **Cluster Resources**, then **Synchronize Cluster Resources**.
2. When you press Enter, SMIT displays the **Synchronize Cluster Resources** screen.
3. Enter the following values:

Ignore Cluster Verification Errors?
If this field is set to **no**, synchronization aborts if verification of the new configuration fails. As part of dynamic reconfiguration processing, the new configuration is verified before it is made the active configuration. If you want synchronization to proceed even if verification fails, set this value to **yes**.

The default is **no**.

Un/Configure Cluster Resources?
If you set this field to **yes**, HACMP for AIX changes the definition of the resource in the ODM and it performs any configuration triggered by the resource change. For example, if you remove a filesystem, HACMP for AIX removes the filesystem from the ODM and also unmounts the filesystem.

If you set this field to **no**, HACMP for AIX changes the definition of the resource in the ODM but does not perform any configuration processing that the change may require. For example, a filesystem would be removed from the HACMP for AIX cluster definition but would not be unmounted. This processing is left to be performed by HACMP for AIX during a fallover.

The default is **yes**. HACMP for AIX attempts to limit the impact on the resource group when a component resource is changed. For example, if you add a filesystem to the resource group that already includes the underlying volume group as a resource, HACMP for AIX does not require any processing of the volume group. Other modifications made to the contents of a resource group may cause the entire resource group to be unconfigured and reconfigured during the dynamic reconfiguration. Cluster clients will experience an interruption in related services while the dynamic reconfiguration is in progress.

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 verification of cluster resources program is run. To save time in the synchronization process, you can toggle this entry field to **yes**. By doing 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.

Note: In some cases, the verification uncovers errors that do not cause the synchronization to fail. HACMP/ES reports the errors in the SMIT command status window so that you are aware of an area of the configuration that may be a problem. You should investigate any error reports, even when they don’t interfere with the synchronization.

Note: Log files that are no longer stored in a default directory, but a user-specified directory instead, 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.

Customizing 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.

See Chapter 8, Cluster Events: Tailoring and Creating for information on customizing cluster events. Also, be sure to consult your planning worksheets and to document any changes you make to your system.

Configuring Custom Cluster Events

To add, change, or remove customized cluster events, 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.

Adding Customized Cluster Events

To add a customized event:

1. Select **Add a Custom Cluster Event** from the menu.
2. 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 script to execute.

3. Press Enter to add the information to the local ODM.
4. 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.

Changing/Showing Custom Cluster Events

To change or show a customized event:

1. Select **Change/Show a Custom Cluster Event** from the menu.

   SMIT displays a picklist of available Cluster Event Methods.

2. Select one entry by highlighting it and press Enter.

   After you select an event, SMIT displays the **Change/Show a Custom Cluster Event** screen.

Enter field values as follows:

Cluster Event Name The current event name is displayed.

New Cluster Event Name Enter the new name for the event.

Cluster Event Description Enter a short description of the event.

Cluster Event Script Filename Enter the pathname of the script to execute.

Removing Custom Cluster Events

To remove a custom cluster event, take the following steps:

1. Select **Remove a Custom Cluster Event** from the **Define Custom Cluster Events** menu.

   When you choose this command, SMIT displays a picklist of customized events that exist in HACMPcustom.

2. Choose a single value. Pressing Enter gives an “Are you sure?” pop-up message. After removal, SMIT returns an OK and control returns to the **Define Custom Cluster Events** SMIT screen

Changing Pre- or Post-Event Processing

Complete the following steps to change the processing for an event. The changes you can make include pointing the Cluster Manager to a different script to process the event, or using the event customization facility to specify pre- or post- processing event scripts. You only need to complete these steps on a single node and then synchronize the ODM data on the other cluster nodes. The HACMP/ES system propagates the information to the other nodes. However, if you have customized the event scripts or written new ones, these must be propagated manually.

1. To start system management for HACMP/ES, enter:

   ```
   smit hacmp
   ```

2. From the main menu, select **Cluster Configuration** > **Cluster Resources** >**Cluster Events** >**Change/Show Cluster Events**. When you press Enter, SMIT displays the list of events.
3. Select a specific event or subevent that you want to configure and press Enter.

   SMIT displays the node name, event name, description, and default event command.
4. Enter field values as follows:

Event Command Enter the name of the command that processes the event. HACMP/ES provides a default script. If additional functionality is required, make any changes by adding pre- or post-event processing scripts or commands of your own design, rather than modifying the default scripts or writing new ones.

Notify Command *This field is 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 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 arguments passed to the event command.

Pre-Event Command *This field is optional.* The field has a picklist of pre-defined custom cluster event names; you can enter more than one custom event name. Use the F7 key to get an alphabetized list of custom event names, or enter the custom event names in the desired order, separated by commas. The names must correspond to a custom event name already defined in the HACMPcustom ODM.

This command is run before the cluster event command executes. This command provides pre-processing for a cluster event. The arguments passed to this command are the event name, event exit status, and the arguments passed to the event command.

Post-Event Command .*This field is optional.* The field has a picklist of pre-defined custom cluster event names; you can enter more than one custom event name. Use the F7 key to get an alphabetized list of custom event names, or enter the custom event names in the desired order, separated by commas. The names must correspond to a custom event name already defined in the HACMPcustom ODM.

This command is run after the cluster event command executes. This command provides post-processing after a cluster event. The arguments passed to this command are the event name, event exit status, and the arguments passed to the event command.

Recovery Command *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 *This field is optional.* 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 if a recovery command is specified.

5. Press Enter to add this information to the ODM on the local node.
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.

Redirecting Cluster Log Files

During normal operation HACMP/ES produces several output log files that you can use to monitor and debug your systems. You can store cluster logs in a location other than its default directory if you so choose.

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**.

For more information on redirecting cluster log files, and instructions on using SMIT to do so, see Chapter 18, Configuring an HACMP/ES Cluster.

Chapter 25 Verifying a Cluster Configuration

This chapter describes how to verify a cluster configuration. Verifying the cluster configuration assures you 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. You should verify the cluster configuration after making changes to a cluster or node.

Note: The directory **/usr/sbin/cluster** and subdirectories have symbolic links to the **/usr/es/sbin/cluster** directory and subdirectories. However, files in these directories are *not* linked as they were in releases prior to 4.3.1.

Overview

After you reconfigure or update a cluster, you should run the cluster verification procedure on one node to check that all nodes agree on the cluster topology, network configuration, and the ownership and takeover of HACMP/ES resources.

Three Ways to Run Cluster Verification

You can run cluster verification in three ways:

Interactively (using a command menu)

To run clverify “interactively,” you start with a simple command, **/usr/es/sbin/cluster/diag/clverify**, and select options from a series of menus until you get to the specific option you want to run.

This option includes a help facility.To use the help facility, type **help** followed by one of the listed options you need help with. For example, if you type `help config`, you see a screen that contains a brief message explaining the purpose of the option, and returns you to the previous prompt to make your next selection.

Directly from the command line

If you already know the option you want to run, and know the complete syntax, you can type the full command and the appropriate options and arguments to run clverify immediately.

Using the SMIT interface

You can use the SMIT interface to run all clverify options except software.

Verifying a Cluster Configuration with the clverify Utility

After reconfiguring or updating a cluster, run the cluster verification procedure on one node to check that all resources used by HACMP/ES are correctly configured and that ownership and takeover of those resources are defined and in agreement across nodes. You can use the **/usr/es/sbin/cluster/diag/clverify** utility to do this.

If you have configured Kerberos on your system, the **clverify** utility also verifies your security setup. In addition, **clverify** checks the existence and consistency of custom cluster snapshot methods and AIX Connections services, if you have configured them.

The **clverify** utility supports several options. Choose the **cluster** option to verify a cluster configuration.

Using the **clverify cluster** option, you can verify either the cluster topology (**topology**) or the node configuration (**config**). The following describes each of these options:

- The **topology** option verifies that all nodes agree on the topology of the cluster. Using this option you can check whether the nodes are in agreement on cluster, node, network, and adapter information. For example, the topology program checks for invalid characters in cluster names, node names, network names, adapter names and resource group names
  - If you select the **sync** option, you can synchronize the topology of the cluster if necessary, forcing agreement with the local node’s definitions.
- The **config** option verifies that the networks are configured correctly and that all nodes agree on the ownership of all defined resources. Using this option you can check the following:
  - Configuration of network information, such as addresses on all nodes in the cluster or whether multiple RS232 serial networks exist on the same tty device. Also checks to ensure that no more than two non-IP networks of one type exist per node.
  - Agreement among all nodes on the ownership of defined resources (filesystems, volume groups, disks, application servers). The **clverify** utility checks for the existence and defined ownership of the filesystems to be taken over, and then checks the volume group and disks where the filesystems reside.
  - Distribution of resources in case of a takeover (node priorities). The **clverify** utility checks that the takeover information matches the owned resources information.
  - Event customization.
  - Configuration of application servers’ start and stop scripts.
  - Agreement among nodes on the major and minor device numbers for NFS-exported file systems.

  If you have configured Kerberos on your system, the **clverify** utility also verifies that:

  - All IP labels listed in the configuration have the appropriate service principals in the **.klogin** file on each node in the cluster.
  - All nodes have the proper service principals.
  - Kerberos is installed on all nodes in the cluster.
  - All nodes have the same security mode setting.

The following sections describe how to run the **clverify** utility in both interactive and command line modes.

Running the clverify Utility Interactively

You can use the **clverify** utility in either of two modes: interactive (using command menus) or directly from the command line. There are "levels" of options for the command. Using the interactive mode, you step through the list of valid options until you get to the specific option you want to run.

Environmental Options for Verifying a Cluster

Use either or both of the following arguments to qualify the environment for the **clverify cluster** program options:

`-e #` Specifies the maximum number of errors clverify can find before aborting the check. By default, the checks performed under the cluster option run until they complete, no matter how many errors are encountered. Specify this argument on the command line after the argument that specifies the type of check. For example:

```
clverify cluster config resources -e 3
```

This option is not available for the topology programs.

`-R` *`file`* Specifies the name of a file into which the warnings output by the check should be saved. Specify this argument as the last argument on the command line. For example:

```
clverify cluster config resources -R warnings.out
```

Using the clverify Help Option

If you want information about the options listed at any point, you can use the **help** option. For example, after typing the command **clverify**, if you type:

```
help cluster
```

you receive a message about the **clverify cluster** command, a list of available options, and the clverify prompt:

```
Verifies that your cluster is configured properly

Valid options are:
topology
config
clverify>
```

Quitting the clverify Interactive Mode

When you type **quit**, the program exits and the system prompt returns.

You can also exit the program from any level by typing CTRL-C at the command line.

Steps for Verifying a Cluster

Follow the steps to run each of the **clverify cluster** commands in order, using the interactive mode.

1. Enter:

   ```
   /usr/es/sbin/cluster/diag/clverify
   ```

   The command returns information on menu options, a list of command options, and the clverify prompt:

   ```
   ----------------------------------------------------------
   To get help on a specific option, type: help <option>
   To return to previous menu, type: back
   To quit the program, type: quit
   ----------------------------------------------------------
   Valid options:
   software
   cluster


   clverify>
   ```

2. Enter the **cluster** subcommand:

   ```
   cluster
   ```

   The command returns the menu information, a list of options, and the clverify.cluster prompt:

   ```
    ...
   Valid options:
   topology
   config


   clverify.cluster>
   ```

3. To see the suboptions for the **topology** option, enter:

   ```
   topology
   ```

   The command returns the menu information, a list of options and the clverify.cluster.topology prompt:

   ```
    ...
   Valid options:
   check
   sync


   clverify.cluster.topology>
   ```

4. To verify that all nodes agree on the topology of the cluster, enter:

   ```
   check
   ```

   When the program finishes, review the output. If a problem with cluster topology exists, you receive a message like the following:

   ```
   ERROR: Could not read local configuration
   ERROR: Local Cluster ID XXX different from Remote Cluster ID XXX.
   ERROR: Nodes have different numbers of networks
   ```

5. (Optional) To synchronize the cluster topology on all cluster nodes as it is defined on the node where you are running this program, enter the following command at the clverify.cluster.topology prompt:

```
sync
```

You should run this program if the **topology check** program reveals disagreement among nodes on the cluster topology, and you are sure you want the configuration as it is defined on the local node.

6. To verify that all nodes agree on the ownership of all defined resources, and that networks and interfaces are properly configured, type **back** until you return to the clverify.cluster prompt, then enter:

   ```
   config
   ```

   The system lists three suboptions to the **config** command. If you choose **all**, the system runs the **networks** program followed by the **resources** program followed by custom-defined verification methods. The **networks** and **resources** programs can be run separately, as explained below.

7. To verify that all networks and interfaces are validly configured and log the results in a file called **verify_nw**, enter:

   ```
   networks -R verify_nw
   ```

   When the program finishes, review the **verify_nw** file. If there are no problems, no messages are logged. If there is a problem with any node's adapters or tty lines, you receive a message like the following:

   ```
   ERROR: The serial device XXX does not exist on node XXX
   ERROR: Service adapter XXX is improperly configured on node XXX
   ```

8. To verify that all nodes agree on the ownership of all defined resources and that all takeover resources are assigned and in agreement, and log the results in a file called **verify_own**, enter:

   ```
   resources -R verify_own
   ```

   The program checks information on the cluster configured resources (file systems, volume groups, disks, and application servers) stored in the global ODM. When the program finishes, review the **verify_own** file. If there are no problems, no messages are logged. If any resource is improperly configured, you receive a message informing you of the error. See the following example.

   ```
   Checking:
         Owned Resources Verification.
   clver: file system /bogus is improperly configured on node seaweed
   clver: file system /home is configured to auto-mount on node seaweed
   clver: volume group bogusvg is improperly configured on node seaweed
   clver: volume group homevg is configured to auto-varyon on node
   seaweed
   clver: volume group bogus2vg is improperly configured on node
   seashell
   clver: application clmarket's start-file
   (/usr/es/sbin/cluster/demos/clmarket] does not exist on node seaweed
   clver: application clmarket's stop-file
   (/usr/es/sbin/cluster/demos/clmarket) does not exist on node seaweed
   clver: physical disk hdisk3, used in owned volume group homevg is not
   owned on node seaweed
   ```

Running the clverify Utility from the Command Line

Once you are familiar with the options for the **clverify cluster** utility, you can enter the complete syntax from the command line instead of using interactive mode.

The complete syntax for the **clverify cluster** command is shown below.

```
clverify software {lpp} [-e num] [-R file]
```

```
clverify cluster {topology check | topology sync | config networks |
      config resources | config both}[-e num] [-R <file>]
include [-e num] to abort the program after num errors
include [-R file] to redirect output to a file
```

For example, to verify the cluster networks configuration and log the results in a file called **verify_nw**, enter;

```
/usr/sbin/cluster/diag/clverify cluster config networks -R verify_nw
```

If you enter an incomplete command, the utility lists the options for the level entered, and puts you in interactive mode to complete the command. For example, if you enter:

```
clverify cluster
```

you get the list of valid options:

```
topology
config


clverify.cluster>
```

To quit the program at any time, type CTRL-C or **quit**.

Verifying Cluster Configuration Using SMIT

After reconfiguring or updating a cluster, run the cluster verification procedure on one node to check that all resources used by HACMP/ES are validly configured, that ownership and takeover of those resources are defined and agreed upon by all nodes, and that Kerberos security, if configured, has been correctly set up.

Cluster verification also lets you add new custom verification methods that check specific components within your cluster configuration. You can change or remove these methods from the verification process depending on the level of cluster verification you want.

Verifying Networks and Resources

Complete the following steps to verify the cluster networks and resources configuration.

1. Enter the following command:

   ```
   smit hacmp
   ```
   SMIT displays the HACMP/ES menu.

2. Select **Cluster Configuration > Cluster Verification > Verify Cluster** and press Enter. Fill in the fields as follows:

   Base HACMP Verification Methods
   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 method in the **Define Custom Verification Method** field.

Custom-Defined Verification Methods

Enter the name of a custom-defined verification method. You can also press F4 for a list of previously defined verification methods. By default, if no methods are selected, the clverify utility also will not check the base verification methods, and it generates an error message.

The order in which verification methods are listed determines the sequence in which selected methods are run. This sequence remains the same for subsequent verifications until different methods 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.

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.

Adding a Custom Verification Method

To add a custom verification method:

1. From the Cluster Configuration menu, select **Cluster Verification** > **Define Custom Verification Method** > **Add a Custom Verification Method** and press Enter**.**
2. Fill in the fields as follows:

   Verification Method Name

   Enter a name for the verification method. Method names can be 16 bytes long and can contain alphanumeric characters. Do not use the word “all,” as this is a keyword indicating that all custom verification methods are to be run.

   Verification Method Description

   Enter a description of the verification method. Method descriptions can be 1024 bytes.

   Verification Method

   Enter a filename for the verification method. The method name can be different from the filename and can be 1024 bytes.

3. Press Enter.

   SMIT runs the **clverify** utility. The output from the verification is displayed in the SMIT Command Status window.

Changing or Showing a Custom Verification Method

To change or show a custom verification method:

1. From the Cluster Configuration menu, select **Cluster Verification** > **Define Custom Verification Method** > **Change/Show a Custom Verification Method** and press Enter.

   SMIT displays a popup list of custom verification methods.

2. Select the verification method you want to change or show and press Enter.
3. Enter a new name, new verification method description, and/or new filename as desired for the verification method.
4. Press Enter.

   SMIT runs the **clverify** utility. The output from the verification is displayed in the SMIT Command Status window.

Removing a Custom Verification Method

To remove a custom verification method:

1. From the Cluster Configuration menu, select **Cluster Verification** > **Define Custom Verification Method** > **Remove a Custom Verification Method** and press Enter.

   SMIT displays a popup list of custom verification methods.

2. Select the verification method you want to remove and press Enter.

   SMIT prompts you to confirm that you want to remove the specified verification method.

3. Press Enter to remove the verification method, or press F3 to cancel.

Chapter 26 Saving and Restoring Cluster Configurations

This chapter explains how to use the cluster snapshot utility to save and restore cluster configurations.

Note: The directory **/usr/sbin/cluster** and subdirectories have symbolic links to the **/usr/es/sbin /cluster** directory and subdirectories. However, files in these directories are *not* linked as they were in releases prior to 4.3.1.

Overview

The cluster snapshot utility allows you to save in a file a record of all the data that defines a particular cluster configuration. This facility gives you the ability to recreate a particular cluster configuration, a process called applying a snapshot, provided the cluster is configured with the requisite hardware and software to support the configuration.

In addition, a snapshot can provide useful information for troubleshooting cluster problems. Because the snapshots are simple ASCII files that can be sent via e-mail, they can make remote problem determination easier. For information about this cluster snapshot utility, see the *HACMP for AIX Troubleshooting Guide.*

You can also add your own custom snapshot methods to store additional user-specified cluster and system information in your snapshots. The output from these user-defined custom methods is reported along with the conventional snapshot information.

Note: You cannot use the cluster snapshot facility in a cluster concurrently running different versions of HACMP/ES.

Information Saved in a Cluster Snapshot

The primary information saved in a cluster snapshot is the data stored in the HACMP/ES ODM classes. This is the information used to recreate the cluster configuration when a cluster snapshot is applied.

The cluster snapshot utility saves the following HACMP/ES ODM classes in the cluster snapshot.

HACMPcluster	Cluster configuration information including cluster ID number, cluster name, names of participating nodes, and information about their IDs.
HACMPnode	Node information including name and ID number.
HACMPnetwork	Network information including the name, attribute, and cluster ID.

HACMPnim Network interface information including name, description, and pathname of the module.

HACMPadapter Adapter information including the type, IP label, and function.

HACMPgroup Resource group information including name, type and participating nodes.

HACMPresource Resource group information including the values of the Inactive takeover attribute and the disk fencing attribute.

HACMPserver Information about application servers.

HACMPevent Event information including the name, description, and names of pre- and post-processing scripts.

HACMPcommand Data needed by certain HACMP/ES commands.

HACMPfence Settings of the IBM SSA fence registers. This data, which is device-dependent, is saved but not restored. The objects in this class are regenerated as part of applying a snapshot.

HACMPdaemons HACMP/ES daemon start-up and stop parameters which may or may not be node-specific. This node-specific information is preserved during restoration; however, the object class itself (including the object data) is the same on all nodes. This class also contains the Cluster Lock Manager Resource Allocation parameters.

HACMPsp2 Data HACMP/ES requires to support the IBM Scalable POWERparallel (SP) system. This information is device-specific and is not used during a snapshot restoration.

HACMPcustom Custom verification, custom event, and custom snapshot method information: method name, full pathname to the method, method description and type definition.

The cluster snapshot does not save any user-customized scripts, applications, or other non-HACMP/ES configuration parameters. For example, the name of an application server and the location of its start and stop scripts are stored in the HACMPserver ODM object class. However, the scripts themselves as well as any applications they may call are not saved.

The cluster snapshot also does not save any device- or configuration-specific data which is outside the scope of HACMP/ES. For instance, the facility saves the names of shared file systems and volume groups; however, other details, such as NFS options or LVM mirroring configuration are not saved.

Format of a Cluster Snapshot

The cluster snapshot utility stores the data it saves in two separate files:

ODM Data File (.odm) This file contains all the data stored in the HACMP/ES ODM object classes for the cluster. This file is given a user-defined basename with the .odm file extension. Because the ODM information must be largely the same on every cluster node, the cluster snapshot saves the values from only one node.

Cluster State Information File (.info) This file contains the output from standard AIX and HACMP/ES system management commands. This file is given the same user-defined basename with the .info file extension. Output from any custom snapshot methods is appended to this file.

The following section describes the contents of the **.odm** file. For more information about the **.info** file, see the *HACMP for AIX Troubleshooting Guide*.

Cluster Snapshot ODM Data File

The cluster snapshot ODM data file is an ASCII text file divided into three delimited sections:

Version section This section identifies the version of the cluster snapshot. The characters <VER identify the start of this section; the characters </VER identify the end of this section. The version number is set by the cluster snapshot software.

Description section This section contains user-defined text that describes the cluster snapshot. You can specify up to 255 characters of descriptive text. The characters <DSC identify the start of this section; the characters </DSC identify the end of this section.

ODM data section This section contains the HACMP/ES ODM object classes in generic AIX ODM stanza format. The characters <ODM identify the start of this section; the characters </ODM identify the end of this section.

The following is an excerpt from a sample cluster snapshot ODM data file showing the various ODM stanzas that are saved.

```
<VER
1.0
</VER

<DSC
My Cluster Snapshot
</DSC
<ODM

HACMPcluster:
        id = 97531
        name = "Breeze1"
        nodename = "mynode"

HACMPnode:
        name = "mynode"
        object = "VERBOSE_LOGGING"
        value = "high"
.
.
.
</ODM
```

clconvert_snapshot Utility

You can run **clconvert_snapshot** to upgrade cluster snapshots from previous versions of HACMP/ES to the most recent version. The **clconvert_snapshot** is not run automatically during installation, and must always be run from the command line. Each time you run the **clconvert_snapshot** command, conversion progress is logged to the **/tmp/clconvert.log** file.

Note: Root user privilege is required to run **clconvert_snapshot**. You must know the HACMP version from which you are converting in order to run this utility.

For more information on the **clconvert_snapshot** utility, refer to the **clconvert_snapshot** man page or to the *HACMP for AIX Administration Guide*, Appendix A, clconvert_snapshot Utility on page 26-4.

Defining a Custom Snapshot Method

If you want additional, customized system and cluster information to be appended to the **.info** file, you should define custom snapshot methods to be executed when you create your cluster snapshot.

To define a custom snapshot method, perform the following steps. Notice that there are two paths to the correct screen.

1. To reach the main HACMP/ES SMIT menu, type:

   ```
   smit hacmp
   ```

2. Select **Cluster Configuration.**
3. From there, select either **Cluster Custom Modification** or **Cluster Snapshot.**

4. Select **Define Custom Snapshot Method > Add a Custom Snapshot Method.**
5. Enter information as follows:

Custom Snapshot Method Name	Enter a name for the custom snapshot method you would like to create.
Custom Snapshot Method Description	Add any descriptive information about the custom method.
Custom Snapshot Script Filename	Add the full pathname to the custom snapshot scriptfile.

Once you have defined one or more custom snapshot methods, when you create a cluster snapshot you are asked to specify which custom method(s) you wish to be executed in addition to the conventional snapshot.

Changing or Removing a Custom Snapshot Method

After you have defined a custom snapshot method, you can change or delete it using the other menu items in the **Define Custom Snapshot Method** SMIT screen: **Change/Show a Custom Snapshot Method** and **Remove a Custom Snapshot Method**.

When you select one of these menus, a picklist of existing custom snapshot methods appears. Choose the one you wish to change or remove and fill in the appropriate fields, or answer the prompt to confirm deletion.

Creating a Cluster Snapshot

You can initiate cluster snapshot creation from any cluster node. You can create a cluster snapshot on a running cluster, and you can create multiple snapshots. The cluster snapshot facility retrieves information from each node in the cluster. Accessibility to all nodes is required. Because of the large amount of data which must be retrieved when creating the cluster snapshot, the time and memory consumed may be substantial, especially when the number of cluster nodes is high. Cluster snapshot files typically require approximately 10 Kb per node.

Note: To get an accurate snapshot of a system that has been configured with Kerberos security, you must set up *all* Kerberos service principals before taking the snapshot. See Chapter 18, Configuring an HACMP/ES Cluster, for details about configuring cluster security.

To create a cluster snapshot:

1. Go to the main HACMP/ES SMIT menu by entering:

   ```
   smit hacmp
   ```

2. Select **Cluster Configuration > Cluster Snapshots > Add a Cluster Snapshot**.

3. Enter field values as follows:

Cluster Snapshot Name
Enter the name you want for the basename for the cluster snapshot files. The default directory path for storage and retrieval of the snapshot is **/usr/es/sbin/cluster/snapshots**. You can specify an alternate path using the SNAPSHOTPATH environment variable.

Custom Defined Snapshot Methods
Specify one or more custom snapshot methods to be executed if desired; press F4 for a picklist of custom methods on this node. If you select **All**, the custom methods will be executed in alphabetical order on each node.

Cluster Snapshot Description?
Enter any descriptive text you want inserted into the cluster snapshot. You can specify any text string up to 255 characters in length.

Applying a Cluster Snapshot

Applying a cluster snapshot overwrites the data in the existing HACMP/ES ODM classes on all nodes in the cluster with the new ODM data contained in the snapshot. You can apply a cluster snapshot from any cluster node.

Applying a cluster snapshot may affect both AIX and HACMP/ES ODM objects and system files as well as user-defined files.

If cluster services are inactive on all cluster nodes, applying the snapshot changes the ODM data stored in the system default configuration directory (DCD). If cluster services are active on the local node, applying a snapshot triggers a cluster-wide dynamic reconfiguration event. In dynamic reconfiguration, in addition to synchronizing the ODM data stored in the DCDs on each node, HACMP/ES replaces the current configuration data stored in the active configuration directory (ACD) with the changed configuration data in the DCD. The snapshot becomes the currently active configuration. For more information about dynamic reconfiguration of a cluster, see Chapter 24, Changing the Cluster Configuration.

Note: A cluster snapshot used for dynamic reconfiguration must only contain changes to either the cluster topology or to cluster resources but not both. You cannot change both the cluster topology and cluster resources in a single dynamic reconfiguration event.

To apply a cluster snapshot using SMIT, perform the following steps.

1. To view the main HACMP/ES SMIT menu, enter:

   ```
   smit hacmp
   ```

2. Select **Cluster Configuration > Cluster Snapshots > Apply a Cluster Snapshot**.

   SMIT displays the **Cluster Snapshot to Apply** screen containing a list of all the cluster snapshots that exist in the directory specified by the SNAPSHOTPATH environment variable.

3. Select the cluster snapshot that you want to apply and press Enter.

   SMIT displays the **Apply a Cluster Snapshot** screen.

4. Enter field values as follows:

   Cluster Snapshot Name
   Displays the current basename of the cluster snapshot.

   Cluster Snapshot Description?
   Displays the text stored in the description section of the snapshot files.

   Un/Configure Cluster Resources?
   If you set this field to **yes**, HACMP/ES changes the definition of the resource in the ODM and it performs any configuration triggered by the resource change. For example, if you remove a file system, HACMP/ES removes the file system from the ODM and also unmounts the file system.

   If you set this field to **no**, HACMP/ES changes the definition of the resource in the ODM but does not perform any configuration processing that the change may require. For example, a file system would be removed from the HACMP/ES cluster definition but would not be unmounted. This processing is left to be performed by HACMP/ES during a fallover. By default, this field is set to **yes.**

   HACMP/ES attempts to limit the impact on the resource group when a component resource is changed. For example, if you add a file system to the resource group that already includes the underlying volume group as a resource, HACMP/ES does not require any processing of the volume group. Other modifications made to the contents of a resource group may cause the entire resource group to be unconfigured and reconfigured during the dynamic reconfiguration. Cluster clients will experience an interruption in related services while the dynamic reconfiguration is in progress.

   Force apply if verify fails?
   If this field is set to **no**, synchronization aborts if verification of the new configuration fails. As part of dynamic reconfiguration processing, the new configuration is verified before it is made the active configuration.

   If you want synchronization to proceed even if verification fails, set this value to **yes**. By default, this field is set to **no**.

Undoing an Applied Snapshot

Before the new configuration is applied, the cluster snapshot facility saves the current configuration in a snapshot called **~snapshot.n.odm**, where **n** is either 1, 2, or 3. The saved snapshots are cycled so that only three generations of snapshots exist. If the apply process fails, you can re-apply the previous configuration. These saved snapshot are stored in the directory specified by the SNAPSHOTPATH environment variable.

Changing a Cluster Snapshot

After creating a cluster snapshot, you can change the basename assigned to cluster snapshot files and the description contained in these files. Note that you must use the SMIT interface to perform this task.

To change a cluster snapshot, perform the following steps.

1. To view the main HACMP/ES SMIT menu, enter:

   ```
   smit hacmp
   ```

2. Select **Cluster Configuration > Cluster Snapshots** > **Change/Show a Cluster Snapshot**.

   SMIT displays the **Cluster Snapshot to Change/Show** screen with a list of all the cluster snapshots that exist in the directory specified by SNAPSHOTPATH.

3. Select the cluster snapshot that you want to change and press Enter.

4. Enter field values as follows:

   Cluster Snapshot Name Displays the current basename of the cluster snapshot.

   New Cluster Snapshot Name Enter the new name you want assigned as the basename of the cluster snapshot files.

   Cluster Snapshot Description? SMIT displays the current description. You can edit the text using up to 255 characters.

Removing a Cluster Snapshot

Removing a cluster snapshot deletes both of the ASCII files that define the snapshot from the snapshots directory. (The directory in which the snapshots are stored is defined in the SNAPSHOTPATH environment variable.) You must use SMIT to remove a cluster snapshot.

To remove a cluster snapshot using the SMIT interface, perform the following steps.

1. To view the main HACMP/ES SMIT menu, enter:

   ```
   smit hacmp
   ```

2. Select **Cluster Configuration > Cluster Snapshots > Remove a Cluster Snapshot**.

   SMIT generates and displays a list of all the cluster snapshots that exist in the directory specified by the SNAPSHOTPATH environment variable.

3. Select the cluster snapshot that you want to remove and press Enter.

   The cluster snapshot facility deletes the files in the snapshot directory that are associated with that snapshot.

Chapter 27 Managing Users and Groups in a Cluster

This chapter explains how to use the C-SPOC utility to manage user accounts and groups on all nodes in a cluster by executing a C-SPOC command on a single node.

Overview

One of the basic tasks any system administrator must perform is setting up user accounts and groups. All users require accounts to gain access to the system. Every user account must belong to a group. Groups provide an additional level of security and allow system administrators to manipulate a group of users as a single entity.

For users of an HACMP/ES cluster, system administrators must create duplicate accounts on each cluster node. The user account information stored in the **/etc/passwd** file and in other files stored in the **/etc/security** directory should be consistent on all cluster nodes. For example, if a cluster 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.

System administrators typically keep user accounts synchronized across cluster nodes by copying the key system account and security files to all cluster nodes whenever a new account is created or an existing account is changed. For C-SPOC clusters, the C-SPOC utility simplifies the cluster-wide synchronization of user accounts by propagating the new account or changes to an existing account across all cluster nodes automatically.

The following sections describe how to perform many common administrative tasks involved with managing user accounts and groups in a cluster using the C-SPOC utility.

Managing User Accounts

As system administrator of an HACMP/ES cluster, you may be called upon to perform the following tasks:

- Listing all user accounts on all cluster nodes
- Adding users to all cluster nodes
- Changing characteristics of a user account on all cluster nodes
- Removing a user account from all cluster nodes.

The following sections describe how to accomplish these tasks on all nodes in a cluster using the C-SPOC utility.

Listing Users On All Cluster Nodes

To obtain information about all user accounts on cluster nodes (or about a particular user account), use the C-SPOC **cl_lsuser** command or the C-SPOC SMIT **List all the Users on the Cluster** screen. The **cl_lsuser** command executes the AIX **lsuser** command on each node. To obtain a listing of all user accounts in the cluster, you must specify the **ALL** argument.

If you specify a user name that does not exist on one of the cluster nodes, the **cl_lsuser** command outputs a warning message but continues execution of the command on other cluster nodes. For more information about the **cl_lsuser** command, see its man page.

Warning: If you have a Network Information Service (NIS) database installed on any cluster node, some user information may not appear when you use the **cl_lsuser** command.

To list all user accounts on all cluster nodes using the C-SPOC utility:

1. Enter the fastpath `smit cl_admin` and select the following options: **Cluster Users & Groups > Users > List All Users in the Cluster**.
2. When you press Enter, SMIT executes the **cl_lsuser** command and displays a listing of user accounts similar to the following:

```
                      COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

[TOP]
sigmund  root   0       /
sigmund  daemon 1       /etc
sigmund  bin    2       /bin
sigmund  sys    3       /usr/sys
sigmund  adm    4       /var/adm
sigmund  uucp   5       /usr/lib/uucp
sigmund  guest  100     /home/guest
sigmund  nobody -2      /
sigmund  lpd    9       /
sigmund  nuucp  6       /var/spool/uucppublic
orion    root   0       /
orion    daemon 1       /etc
orion    bin    2       /bin
[MORE...18]
```

If the command fails, check the C-SPOC log file, **/tmp/cspoc.log**, to obtain the status of the command on each cluster node.

Note: **/tmp/cspoc.log** is the default location of this log file. If you redirected this log, check the appropriate location.

Adding User Accounts on all Cluster Nodes

Adding a user to the cluster involves three steps:

- Add an entry for the new user to the **/etc/passwd** file and other system security files.
- Create a home directory for the new user.
- Add the user to a group file.

On AIX systems, you use the **mkuser** command to perform these tasks. This command adds entries for the new user to various system security files, including **/etc/passwd** and **/etc/security/passwd**, adds the new user to a group, and creates a home directory for the new user. Every user account has a number of attributes associated with it. When you create a user, the **mkuser** command fills in values for these attributes from the system default file **/usr/es/lib/security/mkuser.default** file. You can override these default values by specifying an attribute and a value on the **mkuser** command line.

To add a user on one or more nodes in a C-SPOC cluster, use the C-SPOC **cl_mkuser** command or the **Add a User to the Cluster** SMIT screen. The **cl_mkuser** command calls the AIX **mkuser** command to create the user account on each cluster node you specify. The **cl_mkuser** command creates a home directory for the new account on each cluster node.

If a user with the same name already exists on one of the cluster nodes, the operation fails, returning the message "*user-name* already exists on node *node-name*." Optionally, you can specify that the **cl_mkuser** command continue processing even if the user name already exists on one of the cluster nodes by specifying the C-SPOC **-f** flag. For more information, see the **cl_mkuser** command man page.

Note: User accounts do not become active until the root user assigns a password for the account on each node, using the **/etc/passwd** command. The **cl_mkuser** command, like the AIX **mkuser** command, does not create passwords. Until the initial password is set with the **passwd** command or the **pwdadm** command, the password field in the **/etc/passwd** file is set to * (an asterisk) to indicate that there is no valid password. The new user account is disabled until authentication information is added to the **/etc/security/passwd** file.

To add a user to all nodes in a cluster using the C-SPOC utility, perform the following procedure on any cluster node.

1. Enter the SMIT fastpath **smit cl_mkuser** and follow the options to **Add a User to the Cluster**:
2. Enter data in the entry fields to set up the account. AIX provides help screens that describe each attribute. The **User Name** field is the only required field.

   Note: You should specify a value in the **User ID** field so that the account's user ID will be the same on all cluster nodes. If you do not specify this value, AIX could assign different user IDs on each node. A mismatch of user IDs for an account could prevent a user from logging on to another cluster node in the event of a fallover.

3. After entering in user data, press Enter. The **cl_mkuser** command executes, creating the user account specified on all cluster nodes.

If the command fails, check the C-SPOC log file, **/tmp/cspoc.log**. This file contains the status of the **cl_mkuser** command on each cluster node.

Note: **/tmp/cspoc.log** is the default location of this log file. If you redirected this log, check the appropriate location.

Changing Attributes of Users in a Cluster

On AIX systems, you can change any of the attributes associated with an existing user account by using the **chuser** command. Using the **chuser** command you specify the name of the user account you want to change and then specify the attributes with their new values. If you use the SMIT **Change User Attributes** screen, the complete list of user attributes is displayed and you can supply new values for any attributes. The **chuser** command modifies the user information stored in the **/etc/passwd** file and the files in the **/etc/security** directory.

To change the attributes of a user account on one or more cluster nodes, use the C-SPOC **cl_chuser** command or the C-SPOC **Change User Attributes** SMIT screen. The **cl_chuser** command executes the AIX **chuser** command on each cluster node.

Note: Do not use the **cl_chuser** command if you have a Network Information Service (NIS) database installed on any node in your cluster.

All cluster nodes must be active and a user with the specified name must exist on all the nodes for the change operation to proceed. Optionally, you can specify that **cl_chuser** command continue processing if the specified user name exists on any of the cluster nodes. See the **cl_chuser** command man page for more information.

To change the characteristics of a user account on all cluster nodes using the C-SPOC utility:

1. Enter the SMIT fastpath:

   ```
   smit cl_chuser
   ```

   SMIT displays the **Change/Show Characteristics of a User in the Cluster** screen:

```
        Change / Show Characteristics of a User in the Cluster

        Type or select a value for the entry field.
        Press Enter AFTER making all desired changes.

        [Entry Fields]
        Select nodes by Resource Group               []
+
                *** No selection means all nodes! ***
```

2. Specify the name of the user account you want to change and press Enter. Press F4 to obtain a listing of users from which to choose. SMIT displays a complete listing of the user account attributes with their current values filled in.
3. Enter the new values for attributes you want to change and press Enter. AIX provides help screens that explain each attribute. SMIT executes the C-SPOC **cl_chuser** command to change the attributes of the user account on all cluster nodes.

If the command fails, check the C-SPOC log file, **/tmp/cspoc.log**. This file contains the status of the command on each cluster node.

Note: **/tmp/cspoc.log** is the default location of this log file. If you redirected this log, check the appropriate location.

Removing Users from a Cluster

On AIX systems, you remove a user account by using the **rmuser** command or the SMIT **Remove a User From the System** screen. Using the **rmuser** command you specify the name of the user account you want to remove and specify whether you want the user password and other authentication information removed from the **/etc/security/passwd** file.

To remove a user account from one or more cluster nodes, use the C-SPOC **cl_rmuser** command or the C-SPOC **Remove a User from the Cluster** SMIT screen. The **cl_rmuser** command executes the AIX **rmuser** command on all cluster nodes.

Note: The system removes the user account but does not remove the home directory or any files owned by the user. These files are only accessible to users with root authority or by the group in which the user was a member.

To remove a user from all cluster nodes using the C-SPOC utility:

1. Enter the following SMIT fastpath:

   ```
   smit cl_rmuser
   ```

   SMIT displays the Remove a User screen:

2. Enter field data as follows:

User Name	This is the only required field. You must enter a user name for the new account. The user name can be up to 8 characters in length.
Remove Authentication information?	If you specify Yes, password and other authentication information is deleted from system security files.

3. After entering in user data, press Enter. SMIT removes the user account specified from both nodes.

If the command fails, check the C-SPOC log file, **/tmp/cspoc.log**. This file contains the status of the command's execution on each cluster node.

Note: **/tmp/cspoc.log** is the default location of this log file. If you redirected this log, check the appropriate location.

Managing Group Accounts

All users must belong to a group. Groups add a level of security. As the system administrator of an HACMP/ES cluster, you may be called upon to perform the following tasks:

- Listing all groups defined on all cluster nodes.
- Adding groups to all cluster nodes.
- Changing the characteristics of a group on all cluster nodes.
- Removing a group from all cluster nodes.

Using the C-SPOC facility, you can perform these tasks cluster-wide from any node in a cluster.

Listing Groups on All Cluster Nodes

To obtain information about all the groups defined on an AIX system, or about a particular group, you use the **lsgroup** command. Each group has associated attributes that include the names of the users in the group, the user name of the administrator of the group, and the group ID.

To obtain information about the groups defined on all cluster nodes, use the C-SPOC **cl_lsgroup** command, specifying the **ALL** argument, or by using the C-SPOC SMIT **List all the Groups on the Cluster** screen. The **cl_lsgroup** command executes the **lsgroup** command on each cluster node. The output from the **lsgroup** command for all nodes is displayed on the node on which the command was executed.

If you specify a group name that does not exist on a cluster node, the **cl_lsgroup** command outputs a warning message but continues execution of the command on all other cluster nodes. For more information about the **cl_lsgroup** command, see its man page.

Note: If you have a Network Information Service (NIS) database installed on any cluster node, some user information may not appear when you use the **cl_lsgroup** command.

To list all the groups defined on each cluster node using the C-SPOC utility's SMIT interface:

1. Enter:

   ```
   smit cl_lsgroup
   ```

2. SMIT displays the following command status window.

```
                       COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

[TOP]
cav      system 0       true    root
cav      staff  1       false   daemon
cav      bin    2       true    root,bin
cav      sys    3       true    root,bin,sys
cav      adm    4       true    bin,adm
cav      uucp   5       true    nuucp,uucp
cav      mail   6       true
cav      security       7       true    root
cav      cron   8       true    root
cav      printq 9       true
cav      audit  10      true    root
cav      ecs    28      true
cav      nobody -2      false   nobody,lpd
[MORE...56]
```

If the command fails, check the C-SPOC log file, **/tmp/cspoc.log**. This file contains the status of the command execution on each cluster node.

Note: **/tmp/cspoc.log** is the default location of this log file. If you redirected this log, check the appropriate location.

Adding Groups on Cluster Nodes

To define a new group on AIX systems, you use the **mkgroup** command. This command adds an entry for the new group to various system security files, including **/etc/group** and **/etc/security/group.** Every group has a number of attributes associated with it. When you create a group, you must specify the name of the group. You can optionally specify values for other group attributes.

To define a new group on all cluster nodes, use the C-SPOC **cl_mkgroup** command or the C-SPOC **Add a Group to the Cluster** SMIT screen. The **cl_mkgroup** command performs some verification and then calls the AIX **mkgroup** command on each cluster node to create the group you specify.

If a group with the same name already exists on a cluster node, the operation is aborted. By default, the **cl_mkgroup** command requires that both nodes in the HACMP/ES cluster must be powered up and accessible over the network; otherwise, the **cl_mkgroup** command fails with an error. Optionally, if you specify the C-SPOC **-f** flag, the **cl_mkgroup** command continues processing even if it encounters these errors on one of the cluster nodes. See the **cl_mkgroup** command man page for more information.

To define a new group on cluster nodes using the C-SPOC utility:

1. Enter the SMIT fastpath:

   ```
   smit cl_mkgroup
   ```

   SMIT displays the Add a Group screen.

   Enter data in entry fields to create the group account. The **Group Name** is the only required field. Note, however, that you should specify the **Group ID** field as well.
2. After you finish filling in the SMIT fields, press Enter. The C-SPOC **cl_mkgroup** command executes, creating the new group on all cluster nodes.

If the command fails, check the C-SPOC log file, **/tmp/cspoc.log**. This file provides more information about the execution of the command on each cluster node.

Note: **/tmp/cspoc.log** is the default location of this log file. If you redirected this log, check the appropriate location.

Changing Characteristics of Groups in a Cluster

To change the attributes of a group on an AIX system, use the **chgroup** command. Using the **chgroup** command you specify the name of the group you want to change and the attributes with their new values. If you use the SMIT **Change Group Attributes** screen, SMIT displays the complete list of attributes associated with a group which you can modify. The **chgroup** command modifies the user information stored in the **/etc/group** and the **/etc/security/group** files.

To change the attributes of a group on all cluster nodes, use the C-SPOC **cl_chgroup** command or the C-SPOC **Change Group Attributes** SMIT screen. The **cl_chgroup** command executes the AIX **chgroup** command on each cluster node.

Warning: Do not use the **cl_chgroup** command if you have a Network Information Service (NIS) database installed on either node in your cluster. Using the command in this environment could cause serious system database inconsistencies.

For the C-SPOC command to succeed, all cluster nodes must be accessible and a user with the name specified must exist on all cluster nodes. Optionally, if you specify the C-SPOC **-f** flag, the **cl_chgroup** command continues processing even if it encounters an error on one of the cluster nodes. See the **cl_chgroup** command man page for more information.

To change the attributes of a group on all cluster nodes using the C-SPOC utility:

1. Enter the SMIT fastpath `smit cl_chgroup`

   SMIT displays the Change a Group screen.
2. Specify the name of the group you want to change and press Enter. Press F4 to obtain a listing of groups from which to choose. SMIT displays a complete listing of the attributes of the group specified, with their current values filled in.
3. Change the value of any group attribute and press Enter. The **cl_chgroup** command executes, writing the new attribute value in the appropriate system security files on all cluster nodes.

If the command fails, check the C-SPOC log file, **/tmp/cspoc.log**. This file contains the status of the command execution on each cluster node.

Note: **/tmp/cspoc.log** is the default location of this log file. If you redirected this log, check the appropriate location.

Removing Groups from the Cluster

To delete a group on an AIX system, you use the **rmgroup** command. This command removes the entry for the group from the **/etc/group** and **/etc/security/group** files**.** Users that are members of the group are not deleted.

If the group is the primary group for any user, the remove operation fails unless you redefine the user's primary group with the **chuser** command. (For more information about using the **chuser** command, see the section Changing Attributes of Users in a Cluster on page 27-4.) Only the root user can remove an administrative group or a group with administrative users as members.

To remove a group from all cluster nodes, use the C-SPOC **cl_rmgroup** command or the C-SPOC **Remove a Group from the Cluster** SMIT screen. The **cl_rmgroup** command performs some cluster-wide verification checks and then calls the AIX **rmgroup** command to remove the group on each cluster.

If a group with the name specified does not exist on one of the cluster nodes, the **cl_rmgroup** command reports a warning message but continues the operation on the other cluster node. By default, the **cl_rmgroup** command requires that all cluster nodes be powered up and accessible over the network; otherwise, the **cl_rmgroup** command fails with an error. Optionally, if you specify the C-SPOC **-f** flag, the **cl_rmgroup** command continues processing even if it encounters an error on one of the cluster nodes. See the **cl_rmgroup** command man page for more information.

To remove a group from cluster nodes using the C-SPOC utility:

1. Enter `smit cl_rmgroup`:

   SMIT displays the Remove a Group screen.

2. Enter the name of the group you want to remove. Press the F4 key to get a listing of available groups from which to choose. After specifying the group name, press Enter. The **cl_rmgroup** command executes, removing the group from all cluster nodes.

If the command fails, check the C-SPOC log file, **/tmp/cspoc.log**. This file contains the status of the command execution on each cluster node.

Note: **/tmp/cspoc.log** is the default location of this log file. If you redirected this log, check the appropriate location.

Chapter 28 Additional Tasks: NFS and Run-Time Parameters

This chapter describes how to ensure that NFS works properly on an HACMP/ES cluster, and how to change a node's run-time parameters.

Maintaining NFS

In order for NFS to work as expected on an HACMP/ES cluster, you must be aware of certain configuration tasks and issues:

- Creating shared volume groups
- Exporting NFS filesystems
- Mounting issues
- Fallover/takeover issues.

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

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 all nodes in your cluster are not identically configured, you will have problems using NFS in an HACMP/ES environment. The reason is that the system uses the major number as part of the file handle to uniquely identify a Network File System.

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 file system and re-exports the file system, the client application will not recover, since the file system exported by the node will appear to be different from the one exported by the failed node.

Useful Commands for Handling Volume Groups

To see the major number of existing volume groups, from the **/dev** directory, use the command **ls -l *vg***. Each line of output looks like the following example. The first number on each line after the owner and group names ("system" is the group name in the example) is the major number of that device.

```
crw-rw---- 1 root system  10, 0 Jun 23 1995 IPL_rootvg
crw-rw---- 1 root system  28, 0 Apr 12 15:15 chowvg
```

To find a free common major number, you can do the command shown above on each node, and compare the numbers. Or, use the command **lvlstmajor** on each node. The output is the next free major number. Compare the numbers returned, then select a common free major number.

To create the volume group *sharedvg* on *hdiskx* and *hdisky* with a major number 24:

```
mkvg -f -n -y sharedvg -s 4 -V 24 hdiskx hdisky
```

To import the volume group *sharedvg* with the major number 24 from *hdiskx*:

```
importvg -y sharedvg -V24 hdiskx
```

To complete this process you must set the autovaryon = NO. Do this by running the following command:

```
chvg -a n sharedvg
```

NFS Exporting Filesystems and Directories

The default scripts provided with the HACMP for AIX software do not use the export options specified in the standard AIX **/etc/exports** file. Instead, they call the **/usr/sbin/cluster/events/utils/cl_export_fs** utility, which either uses default export options, or uses an alternate **/usr/sbin/cluster/etc/exports** file to determine export options. Similar to the **/etc/exports** file, the **/usr/sbin/cluster/etc/exports** file can contain special options, such as read-only permission, to use when NFS-exporting particular directories and (or) filesystems.

The **cl_export_fs** utility determines whether to use the default export options or the options in the alternate exports file by determining whether the filesystem to export is listed in the alternate exports file. If the filesystem is not found in the alternate exports file, or the alternate exports file does not exist, the filesystem or directory will be NFS exported with the option of root access for all cluster nodes.

For more information, see the section NFS Exporting Filesystems and Directories on page 18-29 in Volume 1 of this manual.

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.

Creating NFS Mount Points on Clients

A mount point is required in order to mount a filesystem via NFS. Mount points are required for NFS clients, not servers; however, you should be aware that a server can also be a client.

You must create a mount point on all nodes which will NFS mount the filesystem, but which are not members of the resource group from which the NFS filesystem is exported. 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.

NFS Mount Points vs. Local Mount Points

An NFS mount point is different from the local mount point, and requires that the filesystem be NFS mounted on the node which also has the filesystem locally mounted. All applications must reference the filesystem strictly through the NFS mount. With the NFS mount point different from the local mount point, you do not need to unmount the NFS mount on takeover or reintegration.

Note: In SMIT, you must manually add the NFS and local mount points into the **Filesystems to NFS Mount** field. Use a semicolon to separate the names, as follows:

```
<NFS mount point>;<local mount point>
```

For example, if the local mount point is **/afs** (as in the example below), and the NFS mount point is **/afs_NFS**, you would enter the names like this:

```
/afs_NFS;/afs
```

See Setting Up NFS Mount Point Different from Local Mount Point on page 6-11 for more information about how HACMP/ES handles NFS mounting in cascading resource groups.

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:

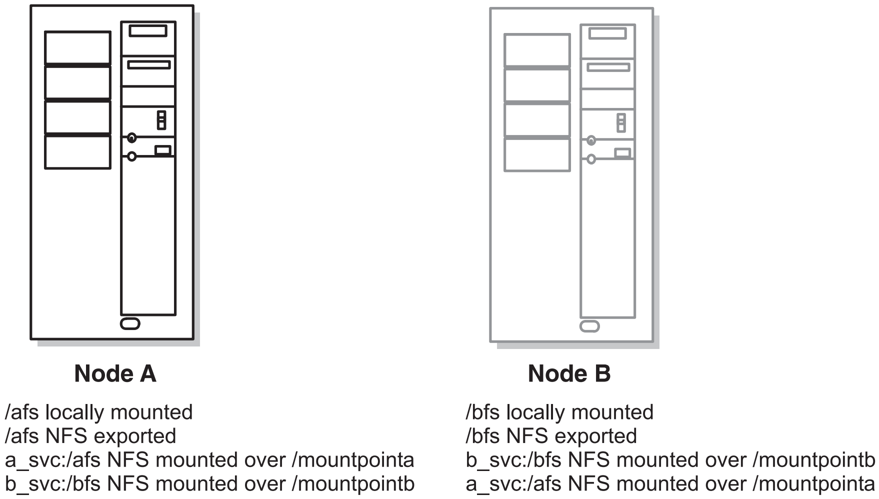
Resource Group	*NodeA_rg*
Participating node names	Node A Node B
Filesystems	**/afs** (filesystems to be locally mounted by node currently owning the resource group)
Filesystems to export	**/afs (**Filesystem to NFS-export by node currently owning resource group. Filesystem is subset of filesystem listed above.)
Filesystems to NFS mount	**/mountpointa;/afs** (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	*NodeB_rg*
Participating node names	Node B Node A
Filesystems	**/bfs**

Filesystems to export	**/bfs**
Filesystems to NFS mount	**/mountpointb;/bfs**

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/utils** 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.

Reliable NFS Server Capability

An HACMP/ES two-node cluster now takes 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**.

Note on Reliable NFS Server Functionality

While exporting or unexporting filesystems, the NFS daemons stop temporarily during event processing. This affects all NFS mounts on a node, causing the NFS clients to be suspended during the stop, as they cannot reach the server. Any clients holding locks are notified that they must reclaim these once the NFS daemons are restarted.

Changing a Node's Run-Time Parameters

Each cluster node supports two run-time parameters. These allow you to:

- Set the level of debug information output by the HACMP/ES scripts. By default, HACMP/ES sets the debug information parameter to high, which produces detailed output from script execution.
- Identify if the node is running Network Information Services (NIS) or name serving. If the node is running NIS, HACMP/ES must disable these services before resolving addresses and re-enable them after name resolution completes.

To change the run-time parameters for a node:

1. Enter the fastpath `smit hacmp`. Then select the following options: **Cluster Configuration > Cluster Resources > Change/Show Run Time Parameters**.

   SMIT displays the **Change/Show Run Time Parameters** screen. Select a node from the list.

2. Enter field values as follows:

   Debug Level Cluster event scripts have two levels of logging. The **low** level only logs events and errors encountered while the script executes. The **high** (default) level logs all commands performed by the script and is strongly recommended. The **high** level provides the level of script tracing needed to resolve many cluster problems.

Host uses NIS or Name Server If the cluster uses Network Information Services (NIS) or name serving, set this field to **true**. HACMP/ES then disables these services before resolving addresses and reenables them after name resolution completes. The default is **false**.

3. Press Enter to add the values into the HACMP/ES for AIX ODM.
4. Press F3 to return to the **Cluster Resources** menu.
5. Select **Synchronize Cluster Resources** from the Cluster Resources menu.
6. 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.

If the cluster is inactive, you can toggle this entry field to **yes** to save time in the synchronization process.

7. Press Enter to synchronize the resource group configuration and node environment across the cluster.
8. Press F3 until you return to the HACMP/ES menu, or F10 to exit SMIT.

Chapter 29 Troubleshooting HACMP/ES Clusters

This chapter describes how to diagnose a problem with an HACMP/ES cluster, and lists some common problems and possible solutions. It shows how to view cluster log files and obtain trace information on HACMP/ES daemons.

For information specific to RSCT daemons, see chapters 30-32 and also Appendix F, RSCT Commands and Utilities, and Appendix G, RSCT Messages. For additional information on diagnosing RSCT problems on the RS/6000 SP, also see the *PSSP Diagnosis Guide.*

Note: The directory **/usr/sbin/cluster** and subdirectories have symbolic links to the **/usr/es/sbin/cluster** directory and subdirectories. However, individual files in these directories are *not* linked as they were in releases prior to version 4.3.1.

Note: The default locations of log files are used in this chapter. If you have redirected any logs, check the appropriate location.

Viewing HACMP/ES Cluster Log Files

Your first approach to diagnosing a problem affecting your cluster should be to examine the cluster log files for messages output by the HACMP/ES subsystems. These messages can provide invaluable information toward understanding the current state of the cluster. The following sections describe the types of messages output by the HACMP/ES software and the log files into which the system writes these messages.

Types of Cluster Messages

The HACMP/ES software generates several types of messages:

Event notification messages

Cluster events cause HACMP/ES scripts to execute. When scripts start, complete, or encounter error conditions, the HACMP/ES software generates a message. For example, the following fragment from a cluster log file illustrates the start and completion messages for several HACMP/ES scripts. The messages include any parameters passed to the script.

```
Feb 25 11:02:46 EVENT START: node_up 2
Feb 25 11:02:46 EVENT START: node_up_local
Feb 25 11:02:47 EVENT START: acquire_service_addr
Feb 25 11:02:56 EVENT COMPLETED: acquire_service_addr
```

Verbose script output

In addition to the start, completion, and error messages generated by scripts, the HACMP/ES software can also generate a detailed report of each step of script processing. In verbose mode, which is the default, the shell generates a message for each command executed in the script, including the values of all arguments to these commands. Verbose mode is recommended. The following fragment from a cluster log file illustrates the verbose output of the **node_up** script.

The verbose messages are prefixed with a plus (+) sign.

```
Feb 25 11:02:46 EVENT START: node_up 2
+ set -u
+ [ 2 = 2 ]
+ /usr/sbin/cluster/events/cmd/clcallev node_up_local
Feb 25 11:02:46 EVENT START: node_up_local
+ set -u
+ rm -f /usr/sbin/cluster/server.status
+ /usr/sbin/cluster/events/cmd/clcallev acquire_service_addr

Feb 25 11:02:47 EVENT START: acquire_service_addr
+ set -u
+ +grep : boot + cut -d: -f1
/usr/sbin/cluster/utilities/cllsif -cSi 2
```

Cluster state messages

When an HACMP/ES cluster starts, stops, or goes through other state changes, it generates messages. These messages may be informational, such as a warning message, or they may report a fatal error condition that causes an HACMP/ES subsystem to terminate. In addition to the **clstart** and **clstop** commands, the following HACMP/ES subsystems generate status messages:

- The Cluster Manager daemon (**clstrmgr**)
- The Cluster Information Program daemon (**clinfo**)
- The Cluster SMUX Peer daemon (**clsmuxpd**).
- The Cluster Lock Manager daemon **(cllockd)**.

The following example illustrates cluster state messages output by the Cluster Manager, the Clinfo daemon, and several HACMP/ES scripts.

```
Feb 25 11:02:30 limpet HACMP/ES: Starting execution of
/etc/rc.cluster with parameters: --
Feb 25 11:02:32 limpet HACMP/ES: clstart: called with flags -sm
Feb 25 11:02:36 limpet clstrmgr[18363]: CLUSTER MANAGER STARTED
Feb 25 11:02:40 limpet HACMP/ES: Completed execution of
/etc/rc.cluster with parameters: --. Exit status = 0
Feb 25 11:02:46 limpet HACMP/ES: EVENT START: node_up 2
Feb 25 11:02:47 limpet HACMP/ES: EVENT START: node_up_local
Feb 25 11:02:47 limpet HACMP/ES: EVENT START: acquire_service_addr
Feb 25 11:02:53 limpet HACMP/ES: EVENT COMPLETED:
acquire_service_addr
Feb 25 11:02:54 limpet HACMP/ES: EVENT START: get_disk_vg_fs
Feb 25 11:02:55 limpet HACMP/ES: EVENT COMPLETED: get_disk_vg_fs
Feb 25 11:03:35 limpet clinfo[6543]: read_config: node address too
long, ignoring.
```

Cluster Message Log Files

The HACMP/ES software writes the messages it generates to the system console and to several log files. Each log file contains a different subset of messages generated by the HACMP/ES software. When viewed as a group, the log files provide a detailed view of all cluster activity.

The following list describes the log files into which the HACMP/ES software writes messages and the types of cluster messages they contain. The list also provides recommendations for using the different log files.

/usr/es/adm/cluster.log

Contains time-stamped, formatted messages generated by HACMP/ES scripts and daemons. For information about viewing this log file and interpreting its messages, see the following Understanding the cluster.log File on page 29-5.

Recommended Use: Because this log file provides a high-level view of current cluster status, it is a good place to look first when diagnosing a cluster problem.

/tmp/hacmp.out

Contains time-stamped, formatted messages generated by HACMP/ES scripts on the current day. The **/tmp/hacmp.out** log file does not contain state messages.

In verbose mode (recommended), this log file contains a line-by-line record of every command executed by scripts, including the values of all arguments to each command. For information about viewing this log file and interpreting its messages, see the section Understanding the hacmp.out Log File on page 29-8.

Recommended Use: Because the information in this log file supplements and expands upon the information in the **/usr/es/adm/cluster.log** file, it is the primary source of information when investigating a problem.

system error log

Contains time-stamped, formatted messages from all AIX subsystems, including scripts and daemons. For information about viewing this log file and interpreting the messages it contains, see the section Understanding the System Error Log on page 29-12.

Recommended Use: Because the system error log contains time-stamped messages from many other system components, it is a good place to correlate cluster events with system events.

/usr/es/sbin/cluster/history/cluster.*mmdd*

Contains time-stamped, formatted messages generated by HACMP/ES scripts. The system creates a cluster history file every day, identifying each file by its file name extension, where *mm* indicates the month and *dd* indicates the day. For information about viewing this log file and interpreting its messages, see the section Understanding the Cluster History Log File on page 29-15.

Recommended Use: Use the cluster history log files to get an extended view of cluster behavior over time.

/tmp/clstrmgr.debug

Contains time-stamped, formatted messages generated by **clstrmgr** activity. By default, the messages are short. IBM Support personnel may have you turn on **clstrmgr** debug options (for verbose, detailed information) to help them understand a particular problem. With debugging turned on, this file grows quickly. You should clean up the file and turn off debug options as soon as possible.

Recommended Use: Information in this file is for IBM Support personnel.

/tmp/cspoc.log

Contains time-stamped, formatted messages generated by HACMP/ES C-SPOC commands. The **/tmp/cspoc.log** file resides on the node that invokes the C-SPOC command.

Recommended Use: Use the C-SPOC log file when tracing a C-SPOC command's execution on cluster nodes.

/tmp/dms_loads.out

Stores log messages every time HACMP/ES triggers the deadman switch. Over time, this file can grow large enough to cause cluster problems. To avoid this situation, link this file to **/dev/null**.

There may be situations in which you want to turn off the deadman switch. For more information, see Editing the rc.cluster File to Turn Deadman Switch Off on page 20-5. For other information about the DMS, see the *HACMP for AIX Troubleshooting Guide*.

/var/ha/log/grpsvcs
Contains time-stamped messages in ASCII format. These track the execution of internal activities of the **grpsvcs** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

/var/ha/log/topsvcs
Contains time-stamped messages in ASCII format. These track the execution of internal activities of the **topsvcs** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

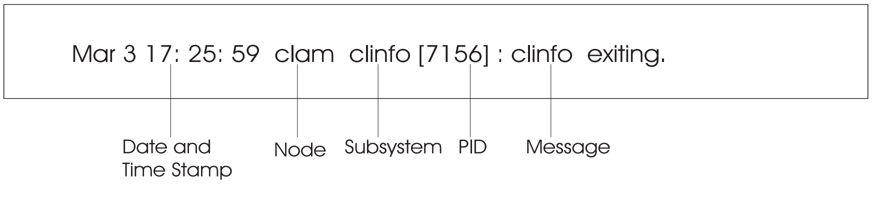
/var/ha/log/grpglsm
Contains time-stamped messages in ASCII format. These track the execution of internal activities of the **grpglsm** daemon. IBM support personnel use this information for troubleshooting. The file gets trimmed regularly. Therefore please save it promptly if there is a chance you may need it.

Understanding the cluster.log File

The **/usr/es/adm/cluster.log** file is a standard text file. When checking this file, first find the most recent error message associated with your problem. Then read back through the log file to the first message relating to that problem. Many error messages cascade from an initial error that usually indicates the problem source.

Format of Messages in the cluster.log File

The entries in the **/usr/es/adm/cluster.log** file use the following format:



```
Mar 3 17: 25: 59  clam  clinfo [7156] : clinfo  exiting.
```

Each entry has the following information:

Date and Time stamp The day and time on which the event occurred.

Node The node on which the event occurred.

Subsystem The HACMP/ES subsystem that generated the event. The subsystems are identified by the following abbreviations:

- **clstrmgr—**The Cluster Manager daemon

- **clinfo—**The Cluster Information Program daemon

- **clsmuxpd—**The Cluster SMUX Peer daemon

- **cllockd—**The Cluster Lock Manager daemon

- **HACMP/ES for AIX—**Startup and reconfiguration scripts.

PID The process ID of the daemon generating the message. (Not included for messages output by scripts.)

Message The message text.

The entry in the previous example indicates that the Cluster Information program (**clinfo**) stopped running on the node named *nodea* at 5:25 P.M. on March 3.

Because the **/usr/es/adm/cluster.log** file is a standard ASCII text file, you can view it using standard AIX file commands, such as the **more** or **tail** commands. However, you can also use the SMIT interface or the HACMP/ES **cldiag** diagnostic utility. The following sections describe each of the options.

Viewing the cluster.log File Using SMIT

To view the **/usr/es/adm/cluster.log** file using SMIT:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **RAS Support** and press Enter.
3. On the RAS Support menu, select **View HACMP Log Files** and press Enter.
4. On the View HACMP Log Files screen, select **Scan the HACMP System Log** and press Enter. This option references the **/usr/es/adm/cluster.log** file.

Note: You can choose to either *scan* the contents of the **/usr/es/adm/cluster.log** file as it exists, or you can *watch* an active log file as new events are appended to it in real time. Typically, you *scan* the file to try to find a problem that has already occurred; you *watch* the file as you test a solution to a problem to determine the results.

Viewing cluster.log File Using the cldiag Utility

To view the **/usr/es/adm/cluster.log** file using the **cldiag** utility, you must include the **/usr/sbin/cluster/diag** directory in your PATH environment variable. Then to run the utility from any directory, perform the following steps.

1. Enter the **cldiag** fastpath.

   The utility returns a list of options and the **cldiag** prompt:

   ```
   -------------------------------------------------------
   To get help on a specific option, type: help <option>
   To return to previous menu, type: back
   To quit the program, type: quit
   -------------------------------------------------------

   valid options:
   debug
   logs
   vgs
   error
   trace

   cldiag>
   ```

   The **cldiag** utility help subcommand provides a brief synopsis of the syntax of the option specified. For more information about the command syntax, see the **cldiag** man page.

2. Enter the **logs** option at the **cldiag** prompt:

   ```
   cldiag> logs
   ```

   The **cldiag** utility displays the following options and prompt. Note that the prompt changes to reflect the last option selection.

   ```
   valid options:
   scripts
   syslog

   cldiag.logs>
   ```

3. To view the **/usr/es/adm/cluster.log** file, enter:

   ```
   cldiag.logs> syslog
   ```

   By default, the **cldiag** utility displays all messages in the log file for every cluster process on the local node. However, you can optionally view only those messages associated with a particular process or with specific processes.

   To view specific messages, quit the **cldiag** utility and use the **lssrc -g cluster** command at the system prompt to obtain the name of cluster processes. Then restart the **cldiag** utility and specify the name of the process whose messages you want to view. If you want to view more than one process, separate multiple names with spaces.

   For example, to view only those messages generated by the Cluster Manager and Clinfo, specify the names as in the following example:

   ```
   cldiag.logs> syslog clstrmgr clinfo
   ```

   Using flags associated with the **syslog** option, you can specify the types of messages you want to view, the time period covered by the messages, and the file in which you want the messages stored.

This table lists the optional command line flags and their function:

Flag	Function
-h ***hostname***	View messages generated by a particular cluster node.
-e	View only error-level messages.
-w	View only warning-level messages.
-d ***days***	View messages logged during a particular time period. Specify the time period in days.
-R ***filename***	Store the messages in the file specified. By default, the **cldiag** utility writes the messages to stdout.

For example, to obtain a listing of all Cluster Manager error-level messages recorded in the last two days and have the listing written to a file named **cm_errors.out**, enter the following:

```
cldiag logs syslog -d 2 -e -Rcm_errors.out clstrmgr
```

This example illustrates how to execute a **cldiag** function directly without traversing the menu hierarchy.

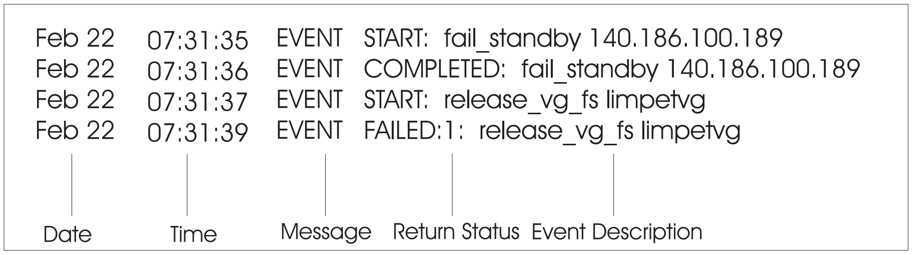
Understanding the hacmp.out Log File

The **/tmp/hacmp.out** file is a standard text file. The system creates a new **hacmp.out** log file every day and retains the last seven copies. Each copy is identified by a number appended to the file name. The most recent log file is named **/tmp/hacmp.out**; the oldest version of the file is named **/tmp/hacmp.out.7**.

When checking the **/tmp/hacmp.out** file, search for EVENT FAILED messages. These messages indicate that a failure has occurred. Then, starting from the failure message, read back through the log file to determine exactly what went wrong. The **/tmp/hacmp.out** log file provides the most important source of information when investigating a problem.

Format of Messages in the hacmp.out Log File

In non-verbose mode, the **/tmp/hacmp.out** log contains the start, completion, and error notification messages output by all HACMP/ES scripts. The following example illustrates the start of the script executed in response to the **node_up** cluster event as it appears in an **/tmp/hacmp.out** file:



```
Feb 22   07:31:35   EVENT   START:  fail_standby 140.186.100.189
Feb 22   07:31:36   EVENT   COMPLETED:  fail_standby 140.186.100.189
Feb 22   07:31:37   EVENT   START:  release_vg_fs limpetvg
Feb 22   07:31:39   EVENT   FAILED:1:  release_vg_fs limpetvg
```

Each entry contains the following information:

Date and Time Stamp The day and time on which the event occurred.

Message Text that describes the cluster activity.

Return Status Messages that report failures include the status returned from the script. This information is not included for scripts that complete successfully.

Event Description The specific action attempted or completed on a node, file system, or volume group.

In verbose mode, the **/tmp/hacmp.out** file also includes the values of arguments and flag settings passed to the scripts and commands. These lines are prefixed with a plus sign (+). The following example illustrates the flags and arguments passed to the **release_vg_fs** script in the previous example.

```
Feb 22 07:31:37 EVENT START: release_vg_fs limpetvg

+ [ -n ]
+ [ -n limpetvg ]
+ echo limpetvg
+ awk -F {for(i=1;i<=NF;i++) a[$i]=$i}
END {for(i in a) print a[i]}
+ sort -r
VG=limpetvg
+ /usr/sbin/cluster/events/utils/cl_deactivate_vgs limpetvg
+ set -u
+ [ 1 -ne 0 ]
+ fgrep -s -x limpetvg
+ lsvg -o
+ [ 0 -ne 0 ]
+ varyoffvg limpetvg
0516-012 lvaryoffvg: Logical volume must be closed. If the logical
      volume contains a filesystem, the umount command will close
      the LV device.
0516-942 varyoffvg: Unable to vary off volume group limpetvg.
+ [ 1 -ne 0 ]
+ cl_log 28 /usr/sbin/cluster/events/utils/cl_deactivate_vgs: Failed
varyoff of limpetvg. /usr/sbin/cluster/events/utils/cl_deactivate_vgs
limpetvg
+ set -u
+ [ 4 -lt 2 ]
MSG_ID=28
DEFAULT_MSG=/usr/sbin/cluster/events/utils/cl_deactivate_vgs: Failed
varyoff of limpetvg.
+ [ 4 -gt 2 ]
+ shift 2
+ dspmsg scripts.cat 28
/usr/sbin/cluster/events/utils/cl_deactivate_vgs: Failed varyoff of
limpetvg. /usr/sbin/cluster/events/utils/cl_deactivate_vgs limpetvg

MSG=/usr/sbin/cluster/events/utils/cl_deactivate_vgs: Failed varyoff of
limpetvg.
+ [ /usr/sbin/cluster/events/utils/cl_deactivate_vgs: Failed varyoff of
limpetvg. = ]
+ logger -t HACMP /usr/sbin/cluster/events/utils/cl_deactivate_vgs:
Failed varyoff of limpetvg.
+ echo /usr/sbin/cluster/events/utils/cl_deactivate_vgs: Failed varyoff
of limpetvg.
```

```
/usr/sbin/cluster/events/utils/cl_deactivate_vgs: Failed varyoff of
limpetvg.
+ exit 0
STATUS=1
+ exit 1
+ [ 1 -ne 0 ]
STATUS=1
+ exit 1
Feb 25 07:31:52 EVENT FAILED:1: release_vg_fs limpetvg
```

Because the **/tmp/hacmp.out** file is a standard ASCII text file, you can view it using standard AIX file commands, such as the **more** or **tail** commands. However, you can also use the SMIT interface or the HACMP/ES **cldiag** diagnostic utility. The following sections describe each of the options.

Viewing the hacmp.out File Using SMIT

To view the **/tmp/hacmp.out** file using SMIT:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **RAS Support** and press Enter.
3. On the RAS Support screen, select **View HACMP Log Files** and press Enter.

On the View HACMP Log Files menu, you can choose to either *scan* the contents of the **/tmp/hacmp.out** file or *watch* as new events are appended to the log file. Typically, you will scan the file to try to find a problem that has already occurred and then watch the file as you test a solution to the problem. In the menu, the **/tmp/hacmp.out** file is referred to as the HACMP Scripts Log File.

4. Select **Scan the HACMP Script Log File** and press Enter. SMIT displays the following screen.
5. Select a script log file and press Enter.
6. Press F10 to exit SMIT.

Viewing hacmp.out File Using the cldiag Utility

To view the **/tmp/hacmp.out** file using the **cldiag** utility, you must include the **/usr/es/sbin/cluster/diag** directory in your PATH environment variable. Then to run the utility from any directory:

1. Enter the **cldiag** fastpath.

   The utility returns a list of options and the **cldiag** prompt:

   ```
   -------------------------------------------------------
   To get help on a specific option, type: help <option>
   To return to previous menu, type: back
   To quit the program, type: quit
   -------------------------------------------------------

   valid options:
   debug
   logs
   vgs
   error
   trace

   cldiag>
   ```

The **cldiag** utility help subcommand provides a brief synopsis of the syntax of the option specified. For more information about the command syntax, see the **cldiag** man page.

2. Enter the **logs** option at the **cldiag** prompt:

   ```
   cldiag> logs
   ```

   The **cldiag** utility displays the following options and prompt. Note that the prompt changes to reflect the current option selection:

   ```
   valid options:
   scripts
   syslog

   cldiag.logs>
   ```

3. To view the **/tmp/hacmp.out** file, enter:

   ```
   cldiag.logs> scripts
   ```

   By default, the **cldiag** utility writes the entire contents of **/tmp/hacmp.out** file to stdout. However, you can view only messages related to one or more specific events, such as **node_up or node_up_local**. Separate multiple events by spaces. The following example views only those messages associated with the **node_up** and **node_up_local** events:

   ```
   cldiag.logs> scripts node_up node_up_local
   ```

   By using flags associated with the **scripts** options, you can specify the types of messages you want to view, the time period covered by the messages, and file in which you want the messages stored. The following table lists the optional command line flags and their function:

Flag	Function
-h *hostname*	View messages generated by a particular cluster node. By default, the **scripts** subcommand only displays messages generated by the local node.
-s	View only start and completion messages.
-f	View only failure messages.
-d *days*	View messages logged during a particular time period. You can specify a time period up to seven days. (The HACMP/ES software only keeps the latest seven copies of the **/tmp/hacmp.out** file.) By default, the current day's log, **/tmp/hacmp.out**, is displayed.
-R *filename*	Store the messages in the file specified. By default, the **cldiag** utility writes the messages to stdout.

For example, to obtain a listing of all failure messages associated with the **node_up** event recorded in the last two days, and have the listing written to a file named **script_errors.out**, enter the following:

```
cldiag logs scripts -d 2 -f -R script_errors.out node_up
```

Setting the Level of Information Recorded in the hacmp.out File

To set the level of information recorded in the **/tmp/hacmp.out** file:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES for AIX menu.
2. On the HACMP for AIX menu, select **Cluster Configuration** and press Enter.
3. On the Cluster Configuration screen, select **Cluster Resources** and press Enter.
4. On the Cluster Resources screen, select **Change/Show Run Time Parameters** and press Enter. SMIT prompts you to specify the node name of the cluster node you want to modify.

   Run time parameters are configured on a per-node basis. After you specify the node, SMIT displays the following screen.
5. To obtain verbose output, make sure the value of the **Debug Level** field is **high**. If necessary, press Enter to record a new value. The Command Status screen appears.
6. Press F10 to exit SMIT.

Changing the Name or Placement of the hacmp.out Log File

If you want to change the name or placement of the **/tmp/hacmp.out** file, you may redirect to a directory of your choice. For more information on redirecting log files, see Customizing Cluster Log Files on page 18-40 in Volume 1 of this guide.

Understanding the System Error Log

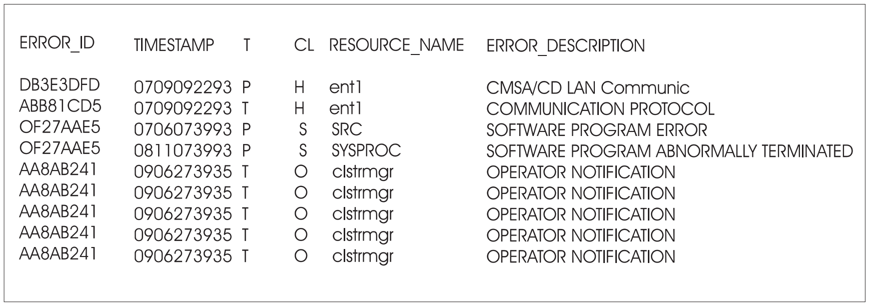
The HACMP/ES software logs script messages to the system error log whenever a script starts, stops, or encounters an error condition, or whenever a daemon generates a state message.

Format of Messages in the System Error Log

The HACMP/ES messages in the system error log follow the same format as that used by other AIX subsystems. You can view the messages in the system error log in short or long format.

In short format, also called summary format, each message in the system error log occupies a single line.

The following figure illustrates the short format of the system error log:



```
ERROR_ID   TIMESTAMP    T    CL  RESOURCE_NAME  ERROR_DESCRIPTION

DB3E3DFD   0709092293  P     H   ent1           CMSA/CD LAN Communic
ABB81CD5   0709092293  T     H   ent1           COMMUNICATION PROTOCOL
OF27AAE5   0706073993  P     S   SRC            SOFTWARE PROGRAM ERROR
OF27AAE5   0811073993  P     S   SYSPROC        SOFTWARE PROGRAM ABNORMALLY TERMINATED
AA8AB241   0906273935  T     O   clstrmgr       OPERATOR NOTIFICATION
AA8AB241   0906273935  T     O   clstrmgr       OPERATOR NOTIFICATION
AA8AB241   0906273935  T     O   clstrmgr       OPERATOR NOTIFICATION
AA8AB241   0906273935  T     O   clstrmgr       OPERATOR NOTIFICATION
AA8AB241   0906273935  T     O   clstrmgr       OPERATOR NOTIFICATION
```

Error_ID A unique error identifier.

Timestamp The day and time on which the event occurred.

T Error type: permanent (P), unresolved (U), or temporary (T).

CL Error class: hardware (H), software (S), or informational (O).

Resource_name A text string that identifies the AIX resource or subsystem that generated the message. HACMP/ES messages are identified by the name of their daemon or script.

Error_description A text string that briefly describes the error.

In long format, a page of formatted information is displayed for each error. See the AIX InfoExplorer facility for a detailed description of this format.

Unlike the HACMP/ES log files, the **system error log** is not a text file.

Using the AIX Error Report Command

The AIX **errpt** command generates an error report from entries in the system error log. See the **errpt** man page for information on using this command.

Viewing the System Error Log Using SMIT

To view the system error log using SMIT:

1. Enter the **smit problem** fastpath to display the main AIX SMIT Problem Determination screen.
2. On the Problem Determination screen, select **Error Log** and press Enter.
3. On the Error Log screen, select **Change / Show Characteristics of the Error Log** and press Enter. SMIT displays the following screen.
4. Press F10 to exit SMIT.

For more information on this log file, refer to your AIX documentation.

Viewing the System Error Log Using the cldiag Utility

To view the system error log using the **cldiag** utility, you must include the **/usr/es/sbin/cluster/diag** directory in your PATH environment variable. Then to run the utility from any directory:

1. Enter the **cldiag** fastpath.

   The utility returns a list of options and the **cldiag** prompt:

   ```
   -------------------------------------------------------
   To get help on a specific option, type: help <option>
   To return to previous menu, type: back
   To quit the program, type: quit
   -------------------------------------------------------

   valid options:
   debug
   logs
   vgs
   error
   trace

   cldiag>
   ```

   The **cldiag** utility help subcommand provides a brief synopsis of the syntax of the option specified. For more information about command syntax, see the **cldiag** man page.

2. To view the system error log, enter at the **cldiag** prompt the **error** option with the type of error display you want. For example, to view a listing of the system error log in short format, enter the following command:

   ```
   cldiag> error short
   ```

   To obtain a listing of system error log messages in long format, enter the **error** option with the **long** type designation. To view only those messages in the system error log generated by the HACMP/ES software, enter the **error cluster** option. When you request a listing of cluster error messages, the **cldiag** utility displays system error log messages in short format.

   By default, the **cldiag** utility displays the system error log from the local node. Using flags associated with the **error** option, however, you can choose to view the messages for any other cluster node. In addition, you can specify a file into which the **cldiag** utility writes the error log. The following list describes the optional command line flags and their functions:

Flag	Function
-h *hostname*	View messages generated by a particular cluster node. By default, only messages on the local node are displayed.
-R *filename*	Store the messages in the file specified. By default, the **cldiag** utility writes the messages to stdout.

   For example, to obtain a listing of all cluster-related messages in the system error log and have the listing written to a file named **system_errors.out**, enter the following:

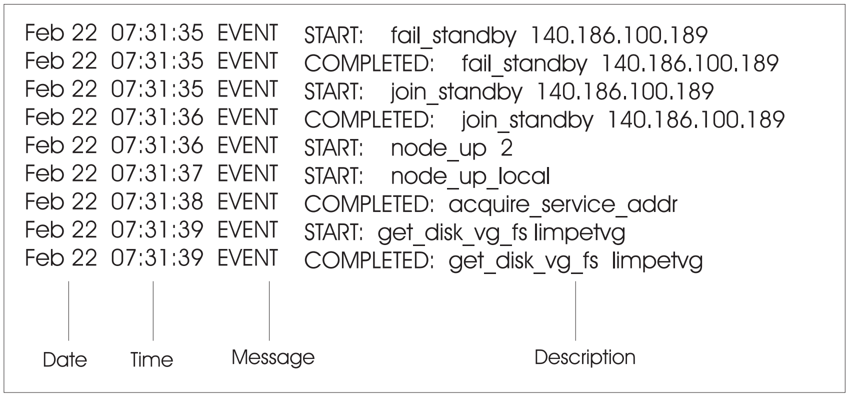
   ```
   cldiag error cluster -R system_errors.out
   ```

Understanding the Cluster History Log File

The cluster history log file is a standard text file with the system-assigned name **/usr/es/sbin/cluster/history/cluster.*mmdd***, where ***mm*** indicates the month and ***dd*** indicates the day in the month. You should decide how many of these log files you want to retain and purge the excess copies on a regular basis to conserve disk storage space. You may also decide to include the cluster history file in your regular system backup procedures.

Format of Messages in the Cluster History Log File

Entries in the cluster history log file use the following format:



```
Feb 22  07:31:35  EVENT   START:    fail_standby  140.186.100.189
Feb 22  07:31:35  EVENT   COMPLETED:    fail_standby  140.186.100.189
Feb 22  07:31:35  EVENT   START:    join_standby  140.186.100.189
Feb 22  07:31:36  EVENT   COMPLETED:    join_standby  140.186.100.189
Feb 22  07:31:36  EVENT   START:    node_up  2
Feb 22  07:31:37  EVENT   START:    node_up_local
Feb 22  07:31:38  EVENT   COMPLETED:  acquire_service_addr
Feb 22  07:31:39  EVENT   START:  get_disk_vg_fs limpetvg
Feb 22  07:31:39  EVENT   COMPLETED:  get_disk_vg_fs  limpetvg
```

Date and Time Stamp The date and time at which the event occurred.

Message Text of the message.

Description Name of the event script.

Viewing the Cluster History Log File

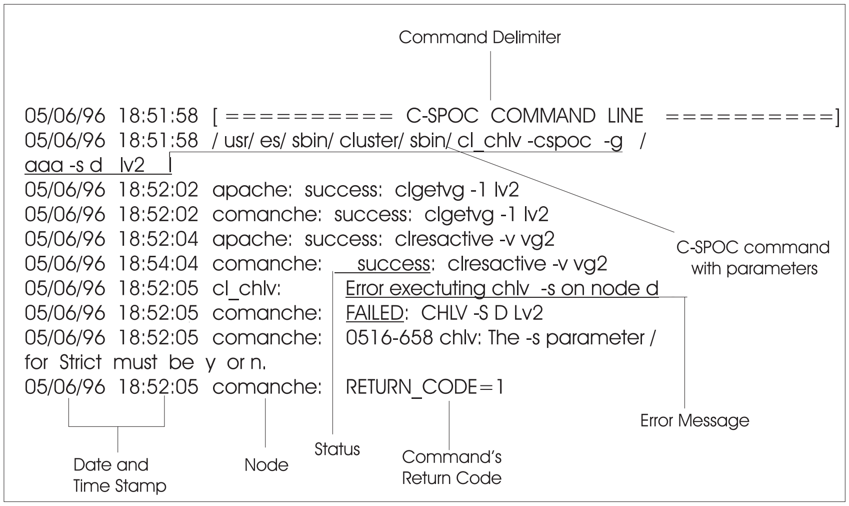
Because the cluster history log file is a standard text file, you can view its contents using standard AIX file commands, such as **cat**, **more**, and **tail**. You cannot view this log file using SMIT or the **cldiag** utility.

Understanding the cspoc.log File

The **/tmp/cspoc.log** file is a standard text file that resides on the source node, the node on which the C-SPOC command is invoked. Many error messages cascade from an underlying AIX error that usually indicates the problem source and success or failure status.

Format of Messages in the cspoc.log File

The entries in the **/tmp/cspoc.log** file use the following format:



```
05/06/96  18:51:58  [ ==========  C-SPOC  COMMAND  LINE   ==========]
05/06/96  18:51:58  / usr/ es/ sbin/ cluster/ sbin/ cl_chlv -cspoc  -g   /
aaa -s d   lv2    l
05/06/96  18:52:02  apache:  success:  clgetvg -1 lv2
05/06/96  18:52:02  comanche:  success:  clgetvg -1 lv2
05/06/96  18:52:04  apache:  success:  clresactive -v vg2
05/06/96  18:54:04  comanche:    success:  clresactive -v vg2
05/06/96  18:52:05  cl_chlv:      Error exectuting chlv  -s on node d
05/06/96  18:52:05  comanche:   FAILED:  CHLV -S D Lv2
05/06/96  18:52:05  comanche:   0516-658 chlv: The -s parameter /
for  Strict  must  be  y  or n.
05/06/96  18:52:05  comanche:   RETURN_CODE=1
```

Each **/tmp/cspoc.log** entry contains a command delimeter to separate C-SPOC command output. This delimeter is followed by the first line of the command's output, which contains arguments (parameters) passed to the command. Additionally, each entry contains the following information:

Date and Time stamp The date and time on which the command was issued.

Node The name of the node on which the command was executed.

Status Text indicating the command's success or failure. Command output that reports a failure also includes the command's return code. No return code is generated for successful command completion. See Appendix A, HACMP for AIX Messages, in the *HACMP for AIX Troubleshooting Guide* for a description of each C-SPOC message.

Error Message Text describing the actual error. The message is recorded in the Error message field. See Appendix A, HACMP for AIX Messages, in the *HACMP for AIX Troubleshooting Guide* for a description of each message.

Note: Error messages generated as a result of standard C-SPOC validation are printed to **stderr** and to the **/tmp/cspoc.log** file.

Viewing the cspoc.log File

The **/tmp/cspoc.log** file is a standard text file that can be viewed in any of the following ways:

- Using standard AIX file commands, such as the **more** or **tail** commands
- Using the SMIT interface

You cannot view this log file using the **cldiag** utility.

Using Standard AIX File Commands

Use standard AIX file commands, such as the **more** or **tail** commands, to view the contents of the **/tmp/cspoc.log** file. See the **more** or **tail** man pages for information on using these commands.

Using the SMIT Interface to View the cspoc.log File

To view the **/tmp/cspoc.log** file using SMIT:

1. Enter `smit hacmp`, then select **RAS Support -> View HACMP/ES Log Files ->Scan the C-SPOC System Log File**.
2. Select **Scan the C-SPOC System Log File** and press Enter. This option references the **/tmp/cspoc.log** file.

Note: Note that you can choose to either *scan* the contents of the **/tmp/cspoc.log** file as it exists, or you can *watch* an active log file as new events are appended to it in real time. Typically, you *scan* the file to try to find a problem that has already occurred; you *watch* the file while duplicating a problem to help determine its cause, or as you test a solution to a problem to determine the results.

Tracing HACMP/ES Daemons

The trace facility helps you isolate a problem within an HACMP/ES environment by allowing you to monitor selected events. Using the trace facility, you can capture a sequential flow of time-stamped system events that provide a fine level of detail on the activity within an HACMP/ES cluster.

The trace facility is a low-level debugging tool that augments the troubleshooting facilities described earlier in this book. While tracing is extremely useful for problem determination and analysis, interpreting a trace report typically requires IBM support.

The trace facility generates large amounts of data. The most practical way to use the trace facility is for short periods of time—from a few seconds to a few minutes. This should be ample time to gather sufficient information about the event you are tracking and to limit use of space on your storage device.

The trace facility has a negligible impact on system performance because of its efficiency.

Use the trace facility to track the operation of the following HACMP/ES daemons:

- The Cluster Manager daemon (**clstrmgr**)
- The Cluster Information Program daemon (**clinfo**)
- The Cluster SMUX Peer daemon (**clsmuxpd**)
- The Cluster Lock Manager daemon (**cllockd**).

The **clstrmgr, clinfo**, and **clsmuxpd** daemons are user-level applications under the control of the SRC. Before you can start a trace on one of these daemons, you must first enable tracing for that daemon. *Enabling* tracing on a daemon adds that daemon to the master list of daemons for which you want to record trace data.

The **cllockd** is a kernel extension. You do not need to enable tracing on a kernel extension.

You can initiate a trace session using either SMIT or the HACMP/ES **cldiag** utility. Using SMIT, you can enable tracing in the HACMP/ES daemons, start and stop a trace session in the daemons, and generate a trace report. Using the **cldiag** utility, you can activate tracing in any HACMP/ES daemon without having to perform the enabling step. The **cldiag** utility performs the enabling procedure, if necessary, and generates the trace report automatically. The following sections describe how to initiate a trace session using either SMIT or the **cldiag** utility.

Using SMIT to Obtain Trace Information on SRC-Controlled Daemons

To initiate a trace session using the SMIT interface:

1. Enable tracing on the daemon or daemons you specify.

   Use the **Enable/Disable Tracing of HACMP Daemons** screen to indicate that the selected daemons should have trace data recorded for them.

2. Start the trace session.

   Use the **Start/Stop/Report Tracing of HACMP Services** screen to trigger the collection of data.

3. Stop the trace session.

   You must stop the trace session before you can generate a report. The tracing session stops when you use either the **Start/Stop/Report Tracing of HACMP Services** screen to stop the tracing session or when the log file becomes full.

4. Generate a trace report.

   Once the trace session is stopped, use the **Start/Stop/Report Tracing of HACMP Services** screen to generate a report.

Each step is described in the following sections.

Enabling Tracing on HACMP/ES Daemons

To enable tracing on the following HACMP daemons (**clstrmgr**, **clinfo**, or **clsmuxpd**):

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP menu, select **RAS Support** and press Enter.
3. On the RAS Support screen, select **Trace Facility** and press Enter.

4. On the Trace Facility screen, select **Enable/Disable Tracing of HACMP Daemons** and press Enter.
5. On the Trace Subsystem screen, select **Start Trace** and press Enter. Note that even though this screen is entitled *Start Trace*, you only use this screen to *enable* tracing. Enabling tracing on this screen does not start a trace session. Rather, it indicates that you want events related to this particular daemon captured the next time you start a trace session. See Starting a Trace Session on page 29-20 for more information.
6. Enter the PID of the daemon whose trace data you want to capture in the **Subsystem PROCESS ID** field.
   a. Press F4 to see a list of all processes and their PIDs.
   b. Select the daemon and press Enter. Note that you can select *only* one daemon at a time. Repeat these steps for each additional daemon that you want to trace.
7. Indicate whether you want a short or long trace event in the **Trace Type** field.

   A *short* trace contains terse information. For the **clstrmgr** daemon, a short trace produces messages only when topology events occur. A *long* trace contains detailed information on time-stamped events.
8. Press Enter to enable the trace.

SMIT displays a screen indicating that tracing for the specified process is enabled.

Disabling Tracing on HACMP/ES Daemons

To disable tracing on the **clstrmgr**, **clinfo**, or **clsmuxpd** daemons:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **RAS Support** and press Enter.
3. On the RAS Support screen, select **Trace Facility** and press Enter.
4. On the Trace Facility screen, select **Enable/Disable Tracing of HACMP Daemons** and press Enter.
5. On the Trace Subsystem screen, select **STOP Trace** and press Enter. Note that even though this screen is entitled *Stop Trace*, you only use this screen to *disable* tracing. Disabling tracing on this screen does not stop the current trace session. Rather, it indicates that you do not want events related to this particular daemon captured the next time you start a trace session.
6. Enter the PID of the process for which you want to disable tracing in the **Subsystem PROCESS ID** field.
   a. Press F4 to see a list of all processes and their PIDs. A menu appears.
   b. Select the process for which you want to disable tracing and press Enter. Note that you can disable only one daemon at a time. To disable more than one daemon, repeat these steps.
7. Press Enter to disable the trace.

SMIT displays a screen indicating that tracing for the specified daemon has been disabled.

Starting a Trace Session

Starting a trace session triggers the actual recording of data on system events into the system trace log from which you can later generate a report.

Remember, you can start a trace on the **clstrmgr**, **clinfo**, and **clsmuxpd** daemons only if you have previously enabled tracing for them. You do not need to enable tracing on the **cllockd** daemon; it is a kernel extension.

To start a trace session:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **RAS Support** and press Enter.
3. On the RAS Support screen, select **Trace Facility** and press Enter.
4. On the Trace Facility screen, select **Start/Stop/Report Tracing of HACMP Services** and press Enter.
5. On the Start/Stop/Report Tracing of HACMP Services screen, select **START Trace** and press Enter. SMIT displays the Start Tracing screen.
6. Enter the trace IDs of the daemons that you want to trace in the **Additional IDs of events that you want to include in trace** field.

   Press F4 to see a list of the trace IDs. (Press Ctrl-v to scroll through the list.) Move the cursor to the first daemon whose events you want to trace and press F7 to select it. Repeat this process for each daemon that you want to trace. When you are done, press Enter. The values that you selected are displayed in the **Additional IDs of events that you want to include in trace** field.

   The HACMP/ES daemons have the following trace IDs:

clstrmgr	910
clinfo	911
clsmuxpd	913

7. Enter values as necessary into the remaining fields and press Enter. See “Understanding SMIT for the Trace Facility” in the AIX InfoExplorer facility for additional information on this screen.

SMIT displays a screen indicating that the trace session has started.

Stopping a Trace Session

You need to stop a trace session before you can generate a trace report. A trace session ends when you actively stop it or when the log file is full.

To stop a trace session:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **RAS Support** and press Enter.
3. On the RAS Support screen, select **Trace Facility** and press Enter.

4. On the Trace Facility screen, select **Start/Stop/Report Tracing of HACMP Services** and press Enter.
5. Select **Stop Trace** and press Enter. SMIT displays a screen indicating that the trace session has stopped.

Generating a Trace Report

A trace report formats the information stored in the trace log file and displays it in a readable form. The report displays text and data for each event according to the rules provided in the trace format file.

When you generate a report, you can specify:

- Events to include (or omit)
- The format of the report.

To generate a trace report:

1. Enter the **smit hacmp** fastpath to display the HACMP/ES menu.
2. On the HACMP/ES menu, select **RAS Support** and press Enter.
3. On the RAS Support screen, select **Trace Facility** and press Enter.
4. On the Trace Facility screen, select **Start/Stop/Report Tracing of HACMP Services** and press Enter.
5. On the Trace screen, select **Generate a Trace Report** and press Enter. A dialog box appears.
6. Indicate the destination and press Enter. SMIT displays the next screen.
7. Enter the trace IDs of the daemons whose events you want to include in the report in the **IDs of events to INCLUDE in Report** field.

   Press F4 to see a list of the trace IDs. (Press Ctrl-v to scroll through the list.) Move the cursor to the first daemon whose events you want to include in the report and press F7 to select it. Repeat this procedure for each daemon that you want to include in the report. When you are done, press Enter. The values that you selected are displayed in the **IDs of events to INCLUDE in Report** field.

   The HACMP/ES daemons have the following trace IDs:

clstrmgr	910
clinfo	911
clsmuxpd	913

8. Enter values as necessary in the remaining fields and press Enter. See “Understanding the Trace Facility” in the AIX InfoExplorer facility for additional information on this screen.
9. When the screen is complete, press Enter to generate the report. The output is sent to the specified destination. For an example of a trace report, see the section Sample Trace Report on page 29-23.

Using the cldiag Utility to Obtain Trace Information

When using the **cldiag** utility, you must include the **/usr/es/sbin/cluster/diag** directory in your PATH environment variable. Then you can run the utility from any directory. You do not need to enable tracing on any of the HACMP/ES daemons before starting a trace session.

To start a trace session using the **cldiag** utility:

1. Enter the **cldiag** fastpath.

   The utility returns a list of options and the **cldiag** prompt:

   ```
   -------------------------------------------------------
   To get help on a specific option, type: help <option>
   To return to previous menu, type: back
   To quit the program, type: quit
   -------------------------------------------------------

   Valid options are:
   debug
   logs
   vgs
   error
   trace

   cldiag>
   ```

   The **cldiag** utility help subcommand provides a brief synopsis of the syntax of the option specified. For more information about the command syntax, see the **cldiag** man page.

2. To activate tracing, enter the **trace** option at the **cldiag** prompt. You must specify (as an argument to the **trace** option) the name of the HACMP/ES daemons for which you want tracing activated. Use spaces to separate the names of the daemons. For example, to activate tracing in the Cluster Manager and Clinfo daemons, enter the following:

   ```
   cldiag> trace clstrmgr clinfo
   ```

   For a complete list of the HACMP/ES daemons, see page 29-18.

   By using flags associated with the trace option, you can specify the duration of the trace session, the level of detail included in the trace (short or long), and the name of a file in which you want the trace report stored. The list on the following page describes the optional command line flags and their function:

Flag	Function
-l	Obtains a long trace. A *long* trace contains detailed information about specific time-stamped events. By default, the **cldiag** utility performs a short trace. A *short* trace contains terse information. For example, a short trace of the **clstrmgr** daemon generates messages only when topology events occur.
-t *time*	Specifies the duration of the trace session. Specify the time period in seconds. By default, the trace session lasts 30 seconds.
-R *filename*	Stores the messages in the file specified. By default, the **cldiag** utility writes the messages to stdout.

For example, to obtain a 15-second trace of the Cluster Manager daemon and have the trace report written to the file **cm_trace.rpt**, enter the following:

```
cldiag trace -t 15 -R cm_trace.rpt clstrmgr
```

For an example of the default trace report, see the Sample Trace Report section below.

Sample Trace Report

The sample trace report shown below was obtained by entering the following command:

```
cldiag trace -R clinfo_trace.rpt clinfo


Wed Nov 15 13:01:37 1997
System: AIX steamer Node: 3
Machine: 000040542000
Internet Address: 00000000 0.0.0.0

trace -j 011 -s -a

ID PROCESS NAME I SYSTEM CALL ELAPSED APPL SYSCALL KERNEL INTERRUPT

001 trace 0.000000 TRACE ON channel 0
Fri Mar 10 13:01:38 1997
011 trace 19.569326 HACMP/ES:clinfo Exiting Function:
broadcast_map_request
011 trace 19.569336 HACMP/ES:clinfo Entering Function: skew_delay
011 trace 19.569351 HACMP/ES:clinfo Exiting Function: skew_delay,
amount: 718650720
011 trace 19.569360 HACMP/ES:clinfo Exiting Function: service_context
011 trace 19.569368 HACMP/ES:clinfo Entering Function: dump_valid_nodes
011 trace 19.569380 HACMP/ES:clinfo Entering Function: dump_valid_nodes
011 trace 19.569387 HACMP/ES:clinfo Entering Function: dump_valid_nodes
011 trace 19.569394 HACMP/ES:clinfo Entering Function: dump_valid_nodes
011 trace 19.569402 HACMP/ES:clinfo Waiting for event
011 trace 22.569933 HACMP/ES:clinfo Entering Function: service_context
011 trace 22.569995 HACMP/ES:clinfo Cluster ID: -1
011 trace 22.570075 HACMP/ES:clinfo Cluster ID: -1
011 trace 22.570087 HACMP/ES:clinfo Cluster ID: -1
011 trace 22.570097 HACMP/ES:clinfo Time Expired: -1
011 trace 22.570106 HACMP/ES:clinfo Entering Function:
broadcast_map_request
002 trace 23.575955 TRACE OFF channel 0
 Wed Nov 15 13:02:01 1997
```

Common Problems and Solutions

This section covers how to diagnose problems that you may encounter as you use HACMP for AIX, and offers possible solutions.

Application Monitor Problems

If you are running application monitors (see Configuring Application Monitoring on page 18-17 for more information), you may encounter occasional problems or situations in which you want to check the state or the configuration of a monitor. Here are some possible problems and ways to diagnose and act on them.

Problem: Checking the State of an Application Monitor

In some circumstances, it may not be clear whether an application monitor is currently running or not. To check on the state of an application monitor, run the following command:

```
ps -ef | grep <application server name> | grep clappmond
```

This command produces a long line of verbose output if the application is being monitored.

If there is no output, the application is not being monitored.

Solution

If the application monitor is not running, there may be a number of reasons, including

- No monitor has been configured for the application server
- The monitor has not started yet because the stabilization interval has not completed
- The monitor is in a suspended state
- The monitor was not configured properly
- An error has occurred

Check to see that a monitor has been configured, the stabilization interval has passed, and the monitor has not been placed in a suspended state, before concluding that something is wrong.

If something is clearly wrong, reexamine the original configuration of the monitor in SMIT (**Cluster Configuration > Cluster Resources > Configure Application Monitoring > Change/Show...**) and reconfigure as needed.

Problem: Application Monitor Does Not Perform Specified Failure Action

The specified failure action does not occur even when an application has clearly failed.

Solution

Check the Restart Interval. If set too short, the Restart Counter may be reset to zero too quickly, resulting in an endless series of restart attempts and no other action taken.

Resource Group Down though Highest Priority Node Up

Problem

You may encounter situations when a cascading resource group is down but the highest priority node is already up. A joining node will not cause the resource group to come online.

This situation could occur in the following scenarios:

- If a Cascading without Fallback resource group is placed on the non-highest priority node, and that node is brought down with a graceful shutdown or a **cldare** *stop.*
- In the **fallover** option of Application Monitoring, if a **rg_move** event moves a resource group from its highest priority node to a lower priority node, and you bring the lower priority node down by a graceful shutdown or a **cldare** *stop.*
- If you use **cldare** *stop* to bring down a cascading resource group which is assigned an Inactive Takeover value of **false** and resides on the highest priority node.

Solution

Unless you bring the resource group up manually, it will remain in an inactive state.

To bring the resource group back up:

1. Enter `smitty hacmp` at the command line
2. Choose **Cluster System Management** > **Cluster Resource Group Management** > **Bring a Resource Group Online**
3. Select the appropriate resource group.

Part 5 RSCT Services

This part contains detailed information on the RS/6000 Cluster Technology Services.

Chapter 30 The Group Services Subsystem

This chapter introduces you to the Group Services subsystem. It includes information about the subsystem's components, its configuration, other components on which it depends, and its operation. It also discusses the relationship of the Group Services subsystem to the other RSCT high availability subsystems. Finally, it describes a procedure you can use to check the status of the subsystem.

After reading this chapter, you will be able to manage the Group Services subsystem and, if necessary, perform problem determination.

For more information, you may want to refer to the *RS/6000 SP High Availability Infrastructure*, SG24-4838 and the *PSSP Diagnosis Guide,* GA22-7350-02.

Note: The default locations of log files are used in this chapter. If you redirected any logs, check the appropriate location.

Introducing Group Services

Group Services is a distributed subsystem of the IBM RS/6000 Cluster Technology (RSCT) on the RS/6000 system. It is one of three RSCT subsystems in that provide a set of high availability services. Other subsystems that provide high availability services include the Event Management and Topology Services (heartbeat) subsystems.

These three distributed subsystems operate within a domain. A domain is a set of RS/6000 machines upon which the RSCT components execute and, exclusively of other machines, provide their services. On the RS/6000 SP, a domain is a system partition. Note that a machine may be in more than one RSCT domain; the control workstation is a member of each system partition and, therefore, a member of each RSCT domain. When a machine is a member of more than one domain, there is an executing copy of each RSCT component per domain.

The function of the Group Services subsystem is to provide other subsystems with a distributed coordination and synchronization service. These other subsystems that depend upon Group Services are called **client subsystems**. Each client subsystem forms one or more **groups** by having its processes connect to the Group Services subsystem and use the various Group Services interfaces. A process of a client subsystem is called a **GS client**. For example, Event Management is a Group Services client subsystem. The Event Manager daemon on each node is a GS client.

A group consists of two pieces of information:

- The list of processes that have joined the group, called the **group membership list**
- A client-specified **group state value**.

Group Services guarantees that all processes that are joined to a group see the same values for the group information, and that they see all changes to the group information in the same order. In addition, the processes may initiate changes to the group information via **protocols** that are controlled by Group Services.

A GS client that has joined a group is called a **provider**. A GS client that wishes only to monitor a group, without being able to initiate changes in the group, is called a **subscriber**.

Once a GS client has initialized its connection to Group Services, it may join a group and become a provider. All other GS clients that have already joined the group (those that have already become providers) are told as part of a join protocol about the new providers that wish to join. The existing providers may either accept new joiners unconditionally (by establishing a one-phase join protocol) or vote on the protocol (by establishing an n-phase protocol). During a vote, they can choose to approve the protocol and accept the new providers into the group, or reject the protocol and refuse to allow the new providers to join.

Group Services monitors the status of all of the processes that are joined to a group. If either the process or the node on which a process is executing fails, Group Services initiates a failure protocol that informs the remaining providers in the group that one or more providers have been lost.

Join and failure protocols are used to modify the group's membership list. Any provider in the group may also propose protocols to modify the group's state value. All protocols may be either unconditional (one-phase) protocols, which are automatically approved and not voted on, or conditional (n-phase) protocols, which are voted on by the providers.

During each phase of an n-phase protocol, each provider may take application-specific action and must vote to approve, reject, or continue the protocol. The protocol completes when it is either approved (the proposed changes become established in the group), or rejected (the proposed changes are dropped).

For more conceptual information about the Group Services subsystem, see *RS/6000 Cluster Technology Group Services Programming Guide and Reference* (order number SA22-7355-00).

Group Services Components

The Group Services subsystem consists of the following components:

Group Services daemon The central component of the Group Services subsystem.

Group Services API (GSAPI) The application programming interface that GS clients use to obtain the services of the Group Services subsystem.

Port numbers TCP/IP port numbers that the Group Services subsystem uses for communications. The Group Services subsystem also uses Unix domain sockets.

Control script A shell script that is used to add, start, stop, and delete the Group Services subsystem, which operates under System Resource Controller (SRC) control.

Files and directories Various files and directories that are used by the Group Services subsystem to maintain run-time data.

The sections that follow contain more details about each of these components.

The Group Services Daemon (hagsd)

The Group Services daemon is contained in the executable file **/usr/sbin/rsct/bin/hagsd**. This daemon runs on each node of an HACMP/ES cluster.

Note that the operational domain of the Group Services subsystem is an HACMP/ES cluster. Unless otherwise stated, a reference to "the Group Services subsystem" is a reference to the Group Services subsystem in a single HACMP/ES cluster.

A GS client communicates with a Group Services daemon that is running on the same node as the GS client. A GS client communicates with the Group Services daemon, through the GSAPI, using a Unix domain socket. Before a GS client registers with Group Services, it must set the HA_DOMAIN_NAME and HA_GS_SUBSYS environment variables.

The Group Services API (GSAPI)

The Group Services Application Programming Interface (GSAPI) is a shared library that a GS client uses to obtain the services of the Group Services subsystem. This shared library is supplied in two versions: one for non-thread safe programs and one for thread-safe programs. These libraries are referenced by the following path names:

- **/usr/lib/libha_gs.a** (non-thread safe version)
- **/usr/lib/libha_gs_r.a** (thread-safe version)

These path names are actually symbolic links to **/usr/sbin/rsct/lib/libha_gs.a** and **/usr/sbin/rsct/lib/libha_gs_r.a**, respectively. The symbolic links are placed in **/usr/lib** for ease of use. For serviceability, the actual libraries are placed in the **/usr/sbin/rsct/lib** directory. These libraries are supplied as shared libraries, also for serviceability.

For details on the GSAPI, see *RSCT Group Services Programming Guide and Reference*.

To allow non-root users to use Group Services:

1. Create a group named hagsuser.
2. Add the desired user IDs to the hagsuser group.
3. Stop and restart **hags** (if it was running before you created the hagsuser group).

Users in the created hagsuser group can use the GSAPI.

Port Numbers and Sockets

The Group Services subsystem uses several types of communications:

- UDP port numbers for intra-domain communications, that is, communications between Group Services daemons within an operational domain.
- Unix domain sockets for communication between GS clients and the local Group Services daemon (via the GSAPI).

Intra-Domain Port Numbers

For communication between Group Services daemons within an operational domain, the Group Services subsystem uses a single UDP port number. This port number is recorded in the Global ODM. This value is set during HACMP/ES installation and configuration.

This intra-domain port number is also set in the **/etc/services** file, using the service name **grpsvcs.*cluster_name***, where *cluster_name* is the name of the HACMP/ES cluster. The **/etc/services** file is updated on all nodes in the cluster. The Group Services daemon obtains the port number from the **/etc/services** file during initialization.

Unix Domain Sockets

Unix domain sockets are used for communication between GS clients and the local Group Services daemon (via the GSAPI). These are connection-oriented sockets. The following socket name is used (`cluster_name` is the name of the cluster):

/var/ha/soc/grpsvcsdsocket.cluster_name	Used by the GSAPI to connect to the Group Services daemon.

The Control Script (grpsvcsctrl)

The Group Services control script is contained in the executable file **/usr/sbin/rsct/bin/grpsvcsctrl.** This script is normally invoked by the HACMP/ES startup process.

If necessary, you can invoke **grpsvcsctrl** directly from the command line.

For more information about **grpsvcsctrl**, see Appendix F, RSCT Commands and Utilities.

The purpose of the **grpsvcsctrl** command is to add (configure) the Group Services subsystem to the HACMP/ES cluster. It can also be used to remove the subsystem from a cluster, start the subsystem, stop it, and clean the subsystem from the cluster.

For more information, see Configuring Group Services on page 30-6.

Files and Directories

The Group Services subsystem uses the following directories:

- **/var/ha/lck**, for lock files
- **/var/ha/log**, for log files
- **/var/ha/run**, for Group Services daemon current working directories
- **/var/ha/soc**, for socket files.

The /var/ha/lck Directory (Lock Files)

In the **/var/ha/lck** directory, the **grpsvcs.tid**.*cluster_name* directory is used to ensure a single running instance of the Group Services daemon, and to establish an instance number for each invocation of the daemon. In the directory name, `cluster_name` is the name of the cluster.

The /var/ha/log Directory (Log Files)

The **/var/ha/log** directory contains trace output from the Group Services daemon.

On the nodes, the file is called **grpsvcs**_*nodenum*_*instnum*.*cluster*_*name*, where:

- *nodenum* is the node number on which the daemon is running
- *instnum* is the instance number of the daemon
- *cluster*_*name* is the name of the cluster to which the node belongs.

The Group Services daemon limits the log size to a pre-established number of lines (by default, 5,000 lines). When the limit is reached, the daemon appends the string .**bak** to the name of the current log file and begins a new log. If a **.bak** version already exists, it is removed before the current log is renamed.

The /var/ha/run Directory (Daemon Working Files)

In the **/var/ha/run directory**, a directory called **grpsvcs**.*cluster*_*name* is created, where *cluster*_*name* is the cluster name. This directory is the current working directory for the Group Services daemon. If the Group Services daemon abnormally terminates, the core dump file is placed in this directory. Whenever the Group Services daemon starts, it renames any core file to **core**_*nodenum.instnum*, where *nodenum* is the node number on which the daemon is running and *instnum* is the instance number of the previous instance of the daemon.

Components on Which Group Services Depends

The Group Services subsystem depends on the following components:

System Resource Controller (SRC) An AIX feature that can be used to define and control subsystems. The Group Services subsystem is called **grpsvcs** on HACMP/ES cluster nodes. The subsystem name is used with the SRC commands (for example, **startsrc** and **lssrc**).

Topology Services An RSCT subsystem that is used to determine which nodes in a system can be reached (that is, are running) at any given time. It is often referred to as **heartbeat**. The Topology Services subsystem is SRC-controlled. It is called **topsvcs** on the HACMP/ES cluster.

Configuring and Operating Group Services

The following sections describe how the components of the Group Services subsystem work together to provide group services. Included are discussions of Group Services:

- Configuration
- Daemon initialization and errors
- Operation.

Configuring Group Services

RSCT is installed as part of the installation of the PSSP or the HACMP/ES product. The Group Services subsystem is contained in the **rsct.basic.rte** and **rsct.basic.sp** fileset.

After the components are installed, the subsystem must be configured for operation. Group Services is configured automatically during cluster configuration.

The **grpsvcsctrl** command provides a number of functions for controlling the operation of the Group Services system. You can use it to:

- Add (configure) the Group Services subsystem
- Start the subsystem
- Stop the subsystem
- Delete (unconfigure) the subsystem
- "Clean" all Group Services subsystems
- Turn tracing of the Group Services daemon on or off.

grpsvcsctrl affects the Group Services subsystem in the current HACMP/ES cluster.

Adding the Subsystem

The definition for the Group Services subsystem is added during HACMP/ES cluster configuration by calling the **grpsvcsctrl** command

Starting and Stopping the Subsystem

The Group Services subsystem is started and stopped when HACMP/ES is started and stopped.

Deleting or Cleaning the Subsystem

If HACMP/ES is removed from a system, Group Services definitions are removed as well.

Tracing the Subsystem

The tracing function of the **grpsvcsctrl** command supplies additional problem determination information when it is requested by the IBM Support Center. Normally, tracing should **not** be turned on, because it may slightly degrade Group Services subsystem performance and can consume large amounts of disk space in the **/var** file system.

Initializing Group Services Daemon

Normally, the Group Services daemon is started by the normal HACMP/ES startup procedure. If necessary, the Group Services daemon can be started using the **grpsvcsctrl** command or the **startsrc** command directly.

During its initialization, the Group Services daemon performs the following steps:

1. It gets the number of the node on which it is running from Topology Services.
2. It tries to connect to the Topology Services subsystem.If the connection cannot be established because the Topology Services subsystem is not running, it is scheduled to be retried every 20 seconds. This continues until the connection to Topology Services is established. Until the connection is established, the Group Services daemon writes an AIX error log entry periodically and no clients may connect to the Group Services subsystem.

3. It performs actions that are necessary to become a daemon.This includes establishing communications with the SRC subsystem so that it can return status in response to SRC commands.
4. It establishes the Group Services domain, which is the set of nodes within the HACMP/ES cluster in which a Group Services daemon is executing.

   At this point, one of the GS daemons establishes itself as the GS nameserver. For details, see Establishing the GS Nameserver on page 30-7.

   Until the domain is established, no GS client requests to join or subscribe to groups are processed.
5. It enters the main control loop.

   In this loop, the Group Services daemon waits for requests from GS clients, messages from other Group Services daemons, messages from the Topology Services subsystem, and requests from the SRC for status.

Establishing the GS Nameserver

The Group Services subsystem must be able to keep track of the groups that its clients want to form. To do this, it establishes a GS nameserver within each domain (a domain is the set of running nodes within each HACMP/ES cluster). The GS nameserver is responsible for keeping track of all client groups that are created in that domain.

To ensure that only one node becomes a GS nameserver, Group Services uses the following protocol:

1. When each daemon is connected to the Topology Services subsystem, it waits for Topology Services to tell it which nodes are currently running in this HACMP/ES cluster.
2. Based on the input from Topology Services, each daemon finds the lowest-numbered running node in the domain. The daemon compares its own node number to the lowest-numbered node and performs one of the following:
   - If the daemon's node is the lowest-numbered node, the daemon waits for all other running nodes to nominate it as the GS nameserver.
   - If the daemon's node is not the lowest-numbered node, it sends nomination messages to the lowest-numbered node periodically, initially every 5 seconds.
3. Once all running nodes have nominated the GS nameserver-to-be and a coronation timer (about 20 seconds) has expired, the nominee sends an insert message to the nodes. All nodes must acknowledge this message. When they do, the nominee becomes the established GS nameserver, and it sends a commit message to all of the nodes.
4. At this point, the Group Services domain is established, and requests by clients to join or subscribe to groups are processed.

Note that this description is in effect when all nodes are being booted simultaneously, such as at initial system power-on. It is often the case, however, that a Group Services daemon is already running on at least one node and is already established as the domain's GS nameserver. In that case, the GS nameserver waits only for Topology Services to identify the newly running nodes. The GS nameserver will then send the newly running nodes proclaim messages that direct the nodes to nominate it as nameserver. Once those nodes then nominate the GS nameserver, the GS nameserver simply executes one or more insert protocols to insert the newly-running nodes into the domain.

Group Services Initialization Errors

The Group Services subsystem creates AIX error log entries to indicate severe internal problems. For most of these, the best response is to contact the IBM Support Center.

However, if you get a message that there has been no heartbeat connection for some time, it indicates that the Topology Services subsystem is not running.

To check the status of the Topology Services subsystem, issue the **lssrc -g topsvcs** command. If the response indicates that the Topology Services subsystem is inoperative, you will need to stop and restart HACMP/ES on the node. If you are unable to restart it, call the IBM Support Center.

Group Services Daemon Operation

Normal operation of the Group Services subsystem requires no administrative intervention. The subsystem normally recovers from temporary failures, such as node failures or failures of Group Services daemons, automatically. However, there are some operational characteristics that may be of interest to administrators.

Due to AIX per-process file descriptor limits, the Group Services subsystem can support a maximum of approximately 2000 GS clients on each node.

The maximum number of nodes that can be contained within a domain is 2048, although Group Services can support node numbers in the range 0 to 65535.

The maximum number of groups to which a GS client may subscribe or that a GS client may join is equivalent to the largest value containable in a signed integer variable.

The maximum number of groups allowed within a domain is 65,535.

These limits are the theoretical maximum limits. In practice, the amount of memory available to the Group Services daemon and its clients will reduce the limits to smaller values

Group Services Procedures

For the most part the Group Services subsystem runs itself without requiring administrator intervention. However, on occasion, you may need to check the status of the subsystem.

Displaying the Status of the Group Services Daemon

You can display the operational status of the Group Services daemon by issuing the **lssrc** command.

On a node, enter:

```
lssrc -l -s grpsvcs
```

In response, the **lssrc** command writes the status information to standard output. The information includes:

- The information provided by the **lssrc -s grpsvcs** command (short form)
- The number of currently connected clients and their process IDs

- The status of the Group Services domain
- The node number on which the GS nameserver is running
- Statistics for client groups with providers or subscribers on this node.

Note that if the **lssrc** command times out, the Group Services daemon is probably unable to connect to the Topology Services subsystem. For more information, see Group Services Initialization Errors on page 30-8.

Here is a sample of the output from the **lssrc -l -s grpsvcs** command on a node for the k21sp2 HACMP/ES cluster:

```
Subsystem         Group            PID     Status
grpsvcs        grpsvcs             11938   active
 4 locally-connected clients.  Their PIDs:
 21344 17000 18852 23486
 HA Group Services domain information:
 Domain established by node 9.
 Number of groups known locally: 3
                    Number of   Number of local
Group name         providers   providers/subscribers
cssMembership          5       1               0
ha_em_peers            7       1               0
CLSTRMGR.1011          5       1               0
```

In this domain, the GS nameserver is on node 9 of the HACMP/ES cluster.

If a GS nameserver has not yet been established, the status indicates that the domain is not established. Similarly, if the GS nameserver fails, the status shows that the domain is recovering. Both of these conditions should clear in a short time. If they do not and the Topology Services subsystem is active, call the IBM Support Center.

Chapter 31 The Event Management Subsystem

This chapter introduces you to the Event Management subsystem. It includes information about the subsystem's components, its configuration, other components on which it depends, and its operation. It also discusses the relationship of the Event Management subsystem to the other RSCT high availability subsystems. Finally, it contains procedures you can use to check the status of the subsystem, or to add or change some of the Event Management configuration data.

For additional information on diagnosing RSCT problems on the RS/6000 SP, see the *PSSP Diagnosis Guide,* GA22-7350-02.

Note: The default locations of log files are used in this chapter. If you redirected any logs, check the appropriate location.

Introducing Event Management

Event Management is a distributed subsystem of the IBM RS/6000 Cluster Technology (RSCT) on the RS/6000 system. The RSCT provides a set of high availability services to HACMP/ES. The other services in RSCT are the Group Services and Topology Services (heartbeat) distributed subsystems. These three distributed subsystems operate within a domain. A domain is a set of RS/6000 machines upon which the RSCT components execute and, exclusively of other machines, provide their services. An HACMP/ES cluster is a domain; the domain name is the cluster name. Note that a machine may be in more than one RSCT domain; a RS/6000 SP node that has HACMP/ES installed is a member of the HACMP domain and a member of the SP domain (system partition). When a machine is a member of more than one domain, there is an executing copy of each RSCT component per domain.

The function of the Event Management subsystem is to match information about the state of system resources with information about resource conditions that are of interest to client programs, which may include applications, subsystems, and other programs. In this chapter, these client programs are referred to as *EM clients*.

Resource states are represented by resource variables. Resource conditions are represented as expressions that have a syntax that is a subset of the expression syntax of the C programming language.

Resource monitors are programs that observe the state of specific system resources and transform this state into several resource variables. The resource monitors periodically pass these variables to the Event Manager daemon. The Event Manager daemon applies expressions, which have been specified by EM clients, to each resource variable. If the expression is true, an event is generated and sent to the appropriate EM client. EM clients may also query the Event Manager daemon for the current values of resource variables.

Resource variables, resource monitors, and other related information are contained in an Event Management Configuration Database (EMCDB) for the HACMP/ES cluster. This EMCDB is included with RSCT.

Event Management Components

The Event Management subsystem consists of the following components:

- Event Manager daemon. The central component of the Event Management subsystem.
- Event Management API (EMAPI). The application programming interface that EM clients use to obtain the services of the Event Management subsystem.
- Resource Monitor API (RMAPI). The application programming interface that resource monitors use to supply resource variables to the Event Manager daemon.
- Resource monitors. Programs that monitor the state of system resources. Several resource monitors are shipped with RSCT.
- Port numbers. TCP/IP port numbers that the Event Management subsystem uses for communications. The Event Management subsystem also uses Unix domain sockets.
- Configuration database. The HACMP/ES Event Management Configuration Database (EMCDB) is included in RSCT.
- Control script. A shell script that is used to add, start, stop, and delete the Event Management subsystem, which operates under System Resource Controller (SRC) control.
- Files and directories. Various files and directories that are used by the Event Management subsystem to maintain run-time data.

The sections that follow contain more details about each of these components.

The Event Manager Daemon (haemd)

The Event Manager daemon is contained in the executable file **/usr/sbin/rsct/bin/haemd**. This daemon runs on each node (machine) of a domain. If a node is a member of more than one domain, there is one executing copy of the daemon per domain.

EM clients communicate with a single Event Manager daemon in the domain that contains the resources in which they are interested. If the domain contains the cluster node on which the EM client is running, the client communicates with the local Event Manager daemon, through the EMAPI, using a Unix domain socket.

Once an EM client has established communications with an Event Manager daemon, the EM client has access to all of the resources in that domain.

Resource monitors communicate with the Event Manager daemon, through the RMAPI, using Unix domain sockets. Resource monitors are always on the same node as the Event Manager daemon. Note that if a node is a member of multiple domains, just as there is one executing copy of the daemon per domain, there is also one executing copy of each resource monitor per domain.

The Event Management API (EMAPI)

The Event Management Application Programming Interface (EMAPI) is a shared library that an EM client uses to obtain the services of the Event Management subsystem. This shared library is supplied in two versions: one for non-thread safe programs and one for thread-safe programs. These libraries are referenced by the following path names:

- **/usr/lib/libha_em.a** (non-thread safe version)
- **/usr/lib/libha_em_r.a** (thread-safe version)

These path names are actually symbolic links to **/usr/sbin/rsct/lib/libha_em.a** and **/usr/sbin/rsct/lib/libha_em_r.a**, respectively. The symbolic links are placed in **/usr/lib** for ease of use. For serviceability, the actual libraries are placed in the **/usr/sbin/rsct/lib** directory. These libraries are supplied as shared libraries, also for serviceability.

For details on the EMAPI, see *Event Management Programming Guide and Reference.*

The Resource Monitor API (RMAPI)

The Resource Monitor Application Programming Interface (RMAPI) is a shared library that a resource monitor uses to supply resource variables to the Event Manager daemon. This shared library is supplied only in a non-thread safe version. This library is referenced through the path name **usr/lib/libha_rr.a** which is a symbolic link to **/usr/sbin/rsct/lib/libha_rr.a**. The symbolic link is placed in **/usr/lib** for ease of use. For serviceability, the actual library is placed in the **/usr/sbin/rsct/lib** directory. This library is supplied as a shared library, also for serviceability.

For details on the RMAPI, see *Event Management Programming Guide and Reference*.

Resource Monitors

A resource monitor conforms to one of the following programming models:

- The resource monitor is a daemon. When necessary, it is started by the Event Manager daemon. The Event Manager daemon connects to the resource monitor to establish communications. The resource monitor has a connection type of *server*.
- The resource monitor logic is incorporated in a subsystem that manages the resources. The Event Manager daemon connects to the resource monitor to establish communications, but the Event Manager daemon does not start the resource monitor (because it is actually another subsystem). The resource monitor has a connection type of *server*.
- The resource monitor logic is implemented in a command. The command can be used by scripts to supply resource variables to the Event Manager daemon. A command-based resource monitor connects to the Event Manager daemon to establish communications. The resource monitor has a connection type of *client*.

Note that a *server* type resource monitor may have multiple executing copies or instances. Each resource monitor instance may supply different instances of the same named resource variable or may supply different resource variables.

In addition, the Event Manager daemon itself performs some resource monitoring function. This function is considered to be a resource monitor with a connection type of *internal.*

RSCT supplies the following resource monitors for use in the HACMP/ES cluster:

IBM.RSCT.harmpd Supplies resource variables that represent the number of processes executing a particular program. These variables can be used to determine whether a particular system daemon is running. This is a daemon with a connection type of **server**.

aixos Supplies resource variables that represent AIX operating system resources. This is a daemon with a connection type of **server**.

There is also one internal resource monitor incorporated in the Event Manager daemon itself:

Membership Supplies resource variables that represent the Host Membership and Adapter Membership states. The Event Manager daemon obtains this information directly from the Group Services subsystem by subscribing to the **HostMembership, enMembership**, and **cssMembership** system groups.

Port Numbers and Sockets

The Event Management subsystem uses several types of communications:

- UDP port numbers for intra-domain communications, that is, communications between Event Manager daemons within a domain.
- Unix domain sockets for communication between EM clients and the local Event Manager daemon (via the EMAPI) and between resource monitors and the Event Manager daemon (via the RMAPI).

Intra-Domain Port Numbers

For communication between Event Manager daemons within a domain, the Event Management subsystem uses a single UDP port number. This intra-domain port number is also set in the **/etc/services** file, using the service name **emsvcs**. The **/etc/services** file is updated on all nodes in the domain when HACMP/ES is installed. The Event Manager daemon obtains the port number from the **/etc/services** file during initialization.

Unix Domain Sockets

Unix domain sockets are used for communication:

- Between EM clients and the local Event Manager daemon (via the EMAPI)
- Between resource monitors and the Event Manager daemon (via the RMAPI).

These are connection-oriented sockets. The following socket names are used (*domain_name* is the name of the domain):

/var/ha/soc/em.clsrv.*domain_name* Used by the EMAPI to connect to the Event Manager daemon.

/var/ha/soc/em.rmsrv.*domain_nam***e** Used by resource monitors to connect to the Event Manager daemon.

/var/ha/soc/haem/em.RM*rmname.rm*inst. *domain_name* Used by the Event Management daemon to connect to the resource monitor that is specified by rmname and rminst, where rmname is the resource monitor name and.rminst is the resource monitor instance number. This resource monitor has a connection type of server.

The Configuration Database (EMCDB)

The Event Management Configuration Database (EMCDB) is installed with RSCT. The format of the EMCDB is designed to permit quick loading of the database by the Event Manager daemon and the RMAPI. It also contains configuration data in an optimized format to minimize the amount of data that must be sent between Event Manager daemons and between an Event Manager daemon and its resource monitors.

When RSCT is installed, the EMCDB is placed in a staging file. When the Event Management subsystem is configured on a node by the **emsvcsctrl** command, the EMCDB is copied from the staging file to a run-time file named **/etc/ha/cfg/em.HACMP.cdb**.

For more information, see Reading the EMCDB on page 31-11.

The Control Script (emsvcsctrl)

The Event Management control script is contained in the executable file **/usr/sbin/rsct/bin/emsvcsctrl.** This script is normally invoked by the HACMP/ES start-up script.

If necessary, you can invoke **emsvcsctrl** directly from the command line.

For more information about **emsvcsctrl**, see Appendix F, RSCT Commands and Utilities.

The purpose of the **emsvcsctrl** command is to add (configure) the Event Management subsystem to a domain. It can also be used to remove the subsystem from a domain, start the subsystem, stop it, and clean the subsystem from all domains.

For more information, see Configuring Event Management on page 31-8.

Files and Directories

The Event Management subsystem uses the following directories:

- **/var/ha/lck**, for lock files
- **/var/ha/log**, for log files
- **/var/ha/run**, for Event Manager daemon current working directories
- **/var/ha/soc**, for socket files
- **/etc/ha/cfg**, for run-time EMCDB and related files

The /var/ha/lck Directory (Lock Files)

In the **/var/ha/lck/haem** directory, the **em.RM***rmname.domain_name* file is used to manage one or more running instances of a resource monitor. In this file name, rmname is the name of the resource monitor and domain_name is the name of the domain.

The **em.RM***rmname.rminstSHMdomain_name* file in the **/var/ha/lck** directory is used by the Event Management daemon to manage a shared memory segment used by the daemon and resource monitor instance specified by rmname and rminst where rmname is the resource monitor name and rminst is the resource monitor instance number. The file contains the shared memory segment ID.

The **em.haemd**.*domain_name* file in the **/var/ha/lck** directory is used to ensure a single running instance of an Event Manager daemon in a domain.

The /var/ha/log Directory (Log Files)

In the **/var/ha/log** directory, the **em.trace**.*domain_name* file contains trace output from the Event Manager daemon. In the file name, *domain_name* is the name of the domain.

The **em.msgtrace**.*domain_name* file contains message trace output from the Event Manager daemon.

The **em.default**.*domain_name* file contains any error messages from the Event Manager daemon that cannot be written to the AIX Error Log. Normally, all daemon error messages are written to the AIX Error Log.

In addition to error messages that cannot be written to the AIX Error Log, the **/var/ha/log/em.default**.*domain_name* file also contains error messages that result from repetitive operational errors. Therefore, both the AIX Error Log and the **/var/ha/log/em.default.***domain_name* file must be examined when performing problem determination on the Event Management subsystem.

The size of the **/var/ha/log/em.default**.*domain_name f*ile is examined every two minutes. If the size of the file exceeds 256K, the file is renamed to **/var/ha/log/em.default.***syspar_name*.**last** and a new default file is created. No more than two copies of this file are kept: the “current” **em.default**.*syspar_name* file and the “last” file.

The **/var/ha/log/emldefault.***domain_name* file is used to record additional error information if the Event Manager daemon cannot start a resource monitor. The error information includes the name of the resource monitor that could not be started.

The /var/ha/run Directory (Daemon Working Files)

In the **/var/ha/run** directory, a directory called **haem**.*domain_name* is created, where *domain_name* is the domain name. This directory is the current working directory for the Event Manager daemon. If the Event Manager daemon abnormally terminates, the core dump file is placed in this directory. Whenever the Event Manager daemon starts, it renames any core file to **core.last**.

If the Event Manager daemon detects an error in the shared memory segment used by the daemon and a resource monitor instance, it creates a dump file containing the first 4096 bytes of the shared memory segment in this directory. The dump file is named rzdump.RMrmname.rminst.time where rmname is the resource monitor name, rminstis the resource monitor instance, and time is the time stamp.

This directory also contains the working directories of any resource monitors that are started by the Event Manager daemon. Each directory has the name of its resource monitor.

Finally, this directory contains the **Rcache_local** and **Rcache_remote** directories. These directories contain the registration cache for local and remote client registration requests, respectively.

For each EM client that establishes communications with the Event Manager daemon, a cache subdirectory is created in the appropriate registration cache directory. This subdirectory has a name of the form p,s,n,q,i, where:

- p is the process ID of the EM client
- s is the seconds portion of the timestamp recorded when the EM client established its session with the Event Manager daemon

- n is the nanoseconds portion of the timestamp
- q is a sequence number for the session
- i is the IP address of the host where the EM client is running. For local clients, the IP address is 0.0.0.0.

Within each subdirectory are several files with numeric names. Each file contains a registration request. The file name is the event command group ID of the group of events within the request.

The /etc/ha/cfg Directory (Run-Time EMCDB and Related Files)

The **/etc/ha/cfg** directory contains the run-time EMCDB file for each system partition. The file is called **/etc/ha/cfg/em**.**hacmp**.**cdb**.

In addition, there is a file called **em**.*domain_name***.cdb_vers**, where *domain_name* is the name of the domain. This file is created by the Event Manager daemon and contains the version string of the EMCDB file used by the daemon. This file is also used by the RMAPI to ensure that it is using the same EMCDB as the Event Manager daemon.

Components on which Event Management Depends

The Event Management subsystem depends on the following components:

SPMI The System Performance Monitor Interface (SPMI) is a library that is included with Performance Aide for AIX (perfagent.tools file set). This library is used by the RMAPI and the **aixos** resource monitor.

Group Services Another RSCT subsystem. It provides high availability membership services for coordinating activities on multiple nodes that are used by the Event Management subsystem.

The sections that follow contain more details about each of these components.

The System Performance Monitor Interface (SPMI) Library

The System Performance Monitor Interface (SPMI) library resides at **/usr/lib/libSpmi.a**. The RMAPI uses the shared memory technology of this library to deliver resource variable values to the Performance Toolbox for AIX product and the Performance Toolbox Parallel Extensions (PTPE) optional feature of PSSP. Thus, a resource monitor supplies data not only for event management, but also for performance monitoring.

Also, the **aixos** resource monitor uses the SPMI to obtain AIX operating system statistics that are the source of the AIX operating system resource variables that are supplied by RSCT.

The Group Services Subsystem

The Group Services subsystem provides another set of high availability services. The Event Management subsystem primarily uses Group Services to monitor the state of each Event Manager daemon in the domain, as follows. Each Event Manager daemon in the domain joins a Group Services group called **ha_em_peers**. Group Services informs each Event Manager daemon, in a synchronized fashion, when a daemon has joined or left the group.

Associated with the **ha_em_peers** group is a group state that contains the version string of the EMCDB. Thus, when an Event Manager daemon joins the **ha_em_peers** group, it can determine the version of the EMCDB that the rest of the group is using.

The Event Manager daemon also subscribes to the **HostMembership, enMembership**, and **cssMembership** system groups. Instances of the IBM.RSCT.Membership.Node.state resource variable are derived from **HostMembership** group information. Instances of the IBM.RSCT.Membership.LANAdapter.state resource variable are derived from **enMembership** and **cssMembership** group information.

Configuring and Operating Event Management

The following sections describe how the components of the Event Management subsystem work together to provide event management services. Included are discussions of:

- Event Management configuration
- Event Manager daemon initialization
- Event Management operation.

Configuring Event Management

RSCT is installed as part of the installation of the HACMP/ES product. The Event Management subsystem is contained in the **rsct.basic.rte** and **rsct.basic.sp** filesets. The EMAPI libraries are contained in the **rsct.clients.rte** and **rsct.clients.sp** filesets.

After the components are installed, the subsystem must be configured for operation. Event Management configuration is performed by the **emsvcsctrl** command, which is invoked by the HACMP/ES start-up script.

The **emsvcsctrl** command provides a number of functions for controlling the operation of the Event Management system. You can use it to:

- Add (configure) the Event Management subsystem
- Start the subsystem
- Stop the subsystem
- Delete the subsystem
- “Clean” all Event Management subsystems
- Turn tracing of the Event Manager daemon on or off

Adding the Subsystem

1. The first step is to add the Event Management startup program to the System Resource Controller (SRC) using the **mkssys** command.
2. The second step is to add the **aixos** resource monitor daemon **harmad** to the SRC using the **mkssys** command.
3. The last step is to copy the EMCDB from the staging file to its run-time location.

Note that if the **emsvcsctrl** add function terminates with an error, the command can be rerun after the problem is fixed. The command takes into account any steps that already completed successfully.

Starting and Stopping the Subsystem

The start and stop functions of the **emsvcsctrl** command simply run the **startsrc** and **stopsrc** commands, respectively. However, emsvcsctrl automatically specifies the subsystem argument to these SRC commands.

Deleting the Subsystem

The delete function of the **emsvcsctrl** command removes the subsystem from the SRC.

Cleaning the Subsystem

The clean function of the **emsvcsctrl** command performs the same function as the delete function. In addition, it removes any files and directories created by the Event Management subsystem.

Tracing the Subsystem

The tracing function of the emsvcsctrl command is provided to supply additional problem determination information when it is requested by the IBM Support Center. Normally, tracing should not be turned on, because it may slightly degrade Event Management subsystem performance and can consume large amounts of disk space in the /var file system.

Initializing Event Manager Daemon

Normally, the Event Manager daemon startup program, **haemd_HACMP**, is started by the HACMP/ES start-up script. If necessary, you can start the startup program using the **emsvcsctrl** command or the **startsrc** command directly. The startup program performs the following steps:

1. It gets the number of the node on which it is running using the **/usr/sbin/cluster/utilities/clhandle** command.
2. It obtains the domain name by calling the **/usr/sbin/cluster/utilities/cldomain** command. Remember that the cluster name is the domain name.
3. Finally, the startup program invokes the Event Manager program **haemd**, passing the information just collected and any arguments passed to the startup program itself. Note that a new process is not started; the process image is just replaced. This permits the Event Manager daemon to be controlled by the SRC. During its initialization, the Event Manager program performs the following steps:

   a. It performs actions that are necessary to become a daemon. This includes establishing communications with the SRC subsystem so that it can return status in response to SRC commands.

   b. It removes from the registration cache all of the subdirectories for local EM clients that no longer exist. That is, if the process ID in the subdirectory name cannot be found, it removes the subdirectory.

c. It tries to connect to the Group Services subsystem. If the connection cannot be established because the Group Services subsystem is not running, it is scheduled to be retried in 5 seconds. This continues until the connection to Group Services is established. Meanwhile, Event Manager daemon initialization continues.

d. It enters the main control loop.

In this loop, the Event Manager daemon waits for requests from EM clients, messages from resource monitors and other Event Manager daemons, messages from the Group Services subsystem, and requests from the SRC for status. It also waits for internal signals that indicate a function that was previously scheduled should now be executed, for example, retrying a connection to Group Services.

However, EM client requests, messages from resource monitors, and messages from other Event Manager daemons (called peers) are refused until the Event Manager daemon has successfully joined the daemon peer group (the **ha_em_peers** group) and has fetched the correct version of the EMCDB.

Joining the Peer Group

After the Event Manager daemon has successfully established a connection with the Group Services subsystem, it tries to join the daemon peer group, a Group Services group called **ha_em_peers**. If this is the first Event Manager daemon to come up in the domain, it establishes the peer group. Otherwise, the other daemons in the peer group either accept or reject the daemon's join request. If an existing peer group member is still recovering from a prior termination of the joining daemon, the join request is rejected. If its join request is rejected, the daemon tries to join again in 15 seconds. This continues until the daemon's join request is accepted.

When it joins the daemon peer group, the Event Manager daemon examines the group state. The group state is the EMCDB version string.

If the group state is null, the joining daemon proposes that the group state be set to the version string that the daemon has fetched from **/etc/ha/cfg/em.HACMP.cdb**. If several daemons try to join the group at about the same time, and the group state is null, then each daemon proposes the group state. When the group is formed, Group Services selects one of the proposals and sets the group state to it. Note that each daemon is proposing the EMCDB version string that it has fetched from **/etc/ha/cfg/em.HACMP.cdb**.

If the group state is not null when it is examined by the joining daemon, a group has already formed and the daemon does not propose a new group state.

After the daemon has successfully joined the peer group, it compares the EMCDB version string contained in the group state to the version string it fetched from **etc/ha/cfg/em.HACMP.cdb**. If they are different, the daemon terminates with an error.

An Event Manager daemon is prevented from joining the peer group as long as any other Event Manager daemon, currently in the peer group, is non-responsive to “pings” from the Group Services subsystem. (When an Event Manager daemon successfully joins the peer group, Group Services requests a response from the Event Manager daemon every two minutes. If the daemon does not respond to the request within two minutes, it is considered to be non-responsive. A daemon is also considered to be non-responsive if it does not reply to the join requests of other daemons within one minute.) The Event Manager daemon status, as displayed by the lssrc command, indicates if a daemon cannot join the peer group. If this is the case, the

em.default.*domain_name* file of any other daemon in the peer group should be examined for errors indicating that an Event Manager daemon is non-responsive. If so, and the non-responsive daemon does not terminate itself within a few minutes, perform the User Response specified for the error.

Reading the EMCDB

Once the daemon has joined the peer group and has determined the EMCDB version, it reads the run-time EMCDB file from the **/etc/ha/cfg** directory.

After the daemon has read and validated the EMCDB, it enables daemon communications. This permits EM clients to send requests to the daemon, resource monitors to connect to the daemon, and peers to send messages to the daemon. At this point, the initialization of the Event Manager daemon is complete.

Operating Event Management Daemon

Normal operation of the Event Management subsystem requires no administrative intervention. The subsystem recovers from temporary failures automatically. However, there are some characteristics that may be of interest to administrators.

For performance reasons, the Event Manager daemon has an internal limit of 256 open file descriptors. In practice, this limits the number of EM client sessions to about 225. This file descriptor limit is per daemon; it does not limit the number of EM clients in the operational domain.

The Event Manager daemon connects to a resource monitor of type server as the resource monitor starts. If a server resource monitor is running prior to the start of the Event Manager daemon, the daemon connects to the resource monitor after enabling daemon communications. If able, the daemon starts a server resource monitor when necessary. However, connection and start attempts are considered under the following circumstances:

1. If it is necessary that the daemon start the resource monitor before each connection attempt, then after three attempts within two hours the resource monitor is “locked” and no further attempts are made.
2. If the resource monitor is not startable by the Event Manager daemon, then after about three successful connections within two hours the resource monitor is “locked” and no further attempts are made.

The rationale for locking the resource monitor is that, if it cannot be started and stay running or successful connections are frequently being lost, then a problem exists with the resource monitor. Once the problem has been determined and corrected, the **haemunlkrm** command can be used to unlock the resource monitor and connect to it, starting it first if necessary. Note that locking does not apply to **client** type resource monitors.

The primary function of the Event Manager daemon is to generate events, by observing resource variable values and applying expressions to those values. However, this function is performed for a resource variable only if an EM client has registered to receive events for that resource variable. The Event Manager daemon also observes a resource variable once to satisfy a query request, if the resource variable is not already being observed. When observations are necessary, the Event Manager daemon commands the appropriate resource monitor (if it has a connection type of server) to supply resource variable values. When observations are no longer necessary, the Event Manager daemon commands the resource monitor to stop supplying values. In this way, the Event Manager daemon performs no action for resource variables that are not of interest to clients.

Even if a resource monitor that has a connection type of **server** is running, it does not supply data to the Event Manager daemon except by command of the daemon.

The Event Manager daemon either observes a resource variable located in shared memory every X seconds, where X is the observation interval that is specified in the resource variable's resource class definition, or when the resource variable's value is sent to the Event Manager daemon by the resource monitor (transparently, via the RMAPI). The values of resource variables of value type Counter and Quantity are located in shared memory. The values of resource variables of value type State are not.

All resource variables that are located in shared memory with the same observation interval are observed on the same time boundary. This minimizes the observation overhead, no matter when the request for a resource variable is made.

Event Management Procedures

For the most part the Event Management subsystem runs itself without requiring administrator intervention. However, on occasion, you may need to check the status of the subsystem.

This section contains the procedure for displaying the status of the Event Manager daemon.

Displaying the Status of the Event Manager Daemon

You can display the operational status of the Event Manager daemon by issuing the lssrc command.

On a node, enter:

```
lssrc -l -s emsvcs
```

In response, the **lssrc** command writes the status information to standard output. The information includes:

- The information provided by the **lssrc -s emsvcs** command (short form)
- The names of any trace flags that are set

  For information on these flags, see the **haemtrcon** command in Appendix C.
- The EMCDB version
- The day and time the Event Manager daemon was started
- A report on the daemon's progress through initialization:

  ```
  Daemon connected to group services: TRUE/FALSE
  Daemon has joined peer group:       TRUE/FALSE
  Daemon communications enabled :     TRUE/FALSE
  ```
- A count of the peer daemons that are currently in the peer group. The count does not include this daemon
- A listing of the peer group state

- The number and type of EM client connections. Note that when a daemon relays a request to another daemon, the sending daemon is treated as a client by the receiving daemon.
- A list of each resource monitor that is defined in the EMCDB and the current status of each, as follows.

  A resource monitor may have multiple executing instances, the number of the resource monitor instance is specified in the Inst column. The connection type (C=client, S=server, I=internal) is found in the Type column. The connection status is indicated by the FD column; if the file descriptor is greater than or equal to 0, a connection is open. If a resource monitor has a shared memory segment used to transfer information to the Event Manager daemon, it has a shared memory ID greater than or equal to 0 in the SHMID column. The process ID of the resource monitor is listed in the PID column; it is interpreted as follows:

  ID greater than 0 Resource monitor has been successfully started by the Event Manager daemon

  ID equal to 0 A resource monitor started by the daemon has terminated (or the resource monitor forked, the parent process exited and the child process is the actual resource monitor)

  ID equal to -1 The resource monitor has never been started by the Event Manager daemon

  ID equal to -2 The resource monitor is not startable by the Event Manager daemon.

  The **Locked** column indicates whether or not a resource monitor is locked and the current count of start attempts and successful connections, in the form {mm}/{nn}, where {mm} is the count of start attempts and {nn} is the count of successful connections.

  If a resource monitor has more than one instance, information is present in the PID and Locked columns only for instance number 0. However, the count of successful connections is for all instances of the resource monitor.
- The highest file descriptor in use.
- The peer daemon status.

  This lists the status of peer daemons by node number, in node number order. Note that this list only includes peer daemons that have joined the peer group since the local daemon started.

  Following the node number are two characters. If both characters are S, the specified node is the number of the node on which this daemon is running. Otherwise, the characters can take on values as follows.

  The first character is I, O where:

  - I indicates that the peer on the specified node is a peer group member.
  - O indicates that the peer is no longer a peer group member (but was at one time).

  The second character is either A or R, where:

  - A indicates that this daemon is accepting join requests from the peer on the specified node.
  - R means this daemon is rejecting join requests.
- A list of internal daemon counters, for use by IBM service personnel.

Chapter 32 The Topology Services Subsystem

This chapter introduces you to the Topology Services subsystem. It includes information about the subsystem's components, its configuration, other components that depend on it, and its operation. It also discusses the relationship of the Topology Services subsystem to the other RSCT high availability subsystems. Finally, it describes a procedure you can use to check the status of the subsystem.

After reading this chapter, you will be able to manage the Topology Services subsystem and, if necessary, perform problem determination.

For additional information on diagnosing RSCT problems on the RS/6000 SP, see the *PSSP Diagnosis Guide,*GA22-7350-02.

Introducing Topology Services

Topology Services is a distributed subsystem of the IBM RS/6000 Clustering Technology Software (RSCT). It is one of the three RSCT subsystems that provide a set of high availability services. The other subsystems are Group Services and Event Management.

The Topology Services subsystem provides the other high availability subsystems with network adapter status, node connectivity information, and a reliable messaging service. The adapter status and node connectivity information is provided to the Group Services subsystem upon request, Group Services then makes it available to its client subsystems. The Reliable Messaging Services, which takes advantage of node connectivity information to reliably deliver a message to a destination node, is available to the other high availability subsystems.

The Topology Services on each cluster node shares this adapter and node connectivity information with its neighbor nodes, forming a heartbeat ring. Each node sends a heartbeat message to one of its neighbors and expects to receive a heartbeat from its other neighbor. Actually each subsystem instance may form multiple rings, one for each network it is monitoring. This depends on the number of networks in the HACMP configuration.

Each heartbeat ring has a leader called the Group Leader. This is a Topology Services daemon whose interface on this ring has the highest IP address. Heartbeat messages allow each member to monitor one of its neighbors and to report to the Group Leader if a loss of heartbeats is detected. The Group Leader responds by forming a new heartbeat ring based on such reports. It also periodically advertises its group’s presence to adapters not currently in the group in an attempt to maximize the size of this heartbeat ring. Every time a new group is formed, it indicates which adapters are present and which adapters are absent. This information is provided to Group Services as adapter status notifications.

In addition to the heartbeat messages, connectivity messages are sent around on all the heartbeat rings. Connectivity messages for each ring are forwarded to other rings if necessary, so that all nodes can construct a connectivity graph. This graph determines node connectivity and defines a set of routes that Reliable Messaging can use to send a message between this node and any other node in the cluster.

Topology Services Components

The Topology Services subsystem consists of the following components:

Topology Services Daemon The central component of the Topology Services subsystem.

Port numbers TCP/IP port numbers that the Topology Services subsystem uses for daemon-to-daemon communications. The Topology Services subsystem also uses UNIX domain sockets for server-to-client communication.

Control script A shell script that is used to add, start, stop, and delete the Topology Services subsystem, which operates under the System Resource Controller (SRC) control.

Start-up script A Perl script that is used to obtain the HACMP/ES configuration from the Global ODM and start up the Topology Services Daemon. This script is invoked by HACMP/ES startup via the SRC.

Files and directories Various files and directories that are used by the Topology Services subsystem to maintain run-time data.

The sections that follow contain more details about each of these components.

The Topology Services Daemon (hatsd)

The Topology Services daemon is contained in the executable file **/usr/sbin/rsct/bin/hatsd.** This daemon runs on each node of a HACMP/ES cluster. If the daemon is running on a node that is part of a SP system, then there will be two daemons running. One will be for HACMP/ES the other will be for the SP system. The SRC can help distinguish between the two instances. The daemons have different subsystem names. The SP version has a subsystem name of **hats** and the HACMP/ES version has a subsystem name of **topsvcs**.

When each daemon starts up, it first reads its configuration from a file setup by the Startup script **topsvcs**, the same as the subsystem name. This file is called the machines list file, because it has all the machines (nodes) listed that are part of the configuration and the IP addresses for each adapter in the configuration for each of those nodes. From this file, the daemon knows the IP address and node number of all potential heartbeat ring members. This file is built from the HACMP/ES Configuration stored in the Global ODM.

The Topology Services subsystem directive is to form as large a heartbeat ring as possible. To form this ring, daemons must alert the other daemons on the other nodes of their presence using a PROCLAIM message. According to a hierarchy defined by Topology Services, daemons may only proclaim to IP addresses that are lower than its own and a daemon may only accept a PROCLAIM message from an IP address higher than its own. Also, a daemon only proclaims if it is the leader of a ring. When a daemon first starts up, it builds a heartbeat ring for every local adapter, containing only that local adapter. This is called a singleton group and this daemon is the leader in each one of these singleton groups.

To manage the changes in these groups, Topology Services defines the following roles:

Group Leader The daemon whose local adapter in this group has the highest IP address of all its group members. The Group Leader Proclaims, handles requests for JOINS, handles DEATH notifications, coordinates group membership changes with group members, and send around connectivity information.

Crown Prince The daemon whose local adapter in this group has the second highest IP address of all its group members. It is also the daemon that could detect the death of the Group Leader and has the authority to ascend to the leader of the group, if that happens.

Mayor A daemon, with a local adapter present in this group that has been picked by the Group Leader to broadcast a message to all the adapters in the group that are also members of his local sub-net. The way that a daemon knows that it has been picked is when it receives a message that has to be broadcasted.

Generic This is any daemon with a local adapter in the heartbeat ring. The role of the Generic daemon is to monitor the heartbeat of the upstream neighbor and inform the Group Leader if the correct number of heartbeats did not arrive. Also to send heartbeats to its downstream neighbor.

Each one of these roles is dynamic, which means that every time a new heartbeat ring is formed, the roles of each member are evaluated and assigned.

In summary, Group Leaders both send and receive PROCLAIM messages. If the PROCLAIM is from a leader with a higher IP address, then the subordinate leader replies with a JOIN request. The superior leader forms a new group with all members from both groups. All members monitor their neighbors for heartbeats. If a sufficient number of heartbeats are missed, a message is sent to the Group Leader and the offending adapter will be dropped from the group. Anytime that here is a membership change, Group Services will be notified.

The Group Leader also accumulates node connectivity information, constructs connectivity graph and set of routes for connections from its node to every other node in the partition. The group connectivity information is sent out to all nodes so that they can use it to update their graphs and also compute routes from their node to any other node. It is this traversal of the graph on each node that determines the node membership notification that will be provided on each node. Whenever the graph changes, routes are re-calculated, and a list of nodes that have connectivity is generated and made available to Group Services.

Port Numbers and Sockets

The Topology Services subsystem uses several types of communications:

- UDP port numbers for intra-cluster communications between Topology Services daemons.
- UNIX domain sockets for communication between Topology Services Clients and Topology Services daemon.

Intra-cluster Port Numbers

For communication between Topology Services daemons within a cluster the Topology Services subsystem uses a single UDP port number. This port number is fixed and is listed in **/etc/services**. Likewise, Group Services and Event Manager have port numbers listed in **/etc/services**. Topology Services uses the entry in **/etc/services** at initialization to establish what port number is going to be used by all Topology Services daemons.

UNIX Domain Sockets

Unix domain sockets are used for communication with Topology Services clients, Group Services and Event Management. These are connection-orientated sockets. The following socket name is used (*clstrname* is the name of the HACMP/ES cluster) to connect to the Topology Services daemon:

```
/var/ha/soc/topsvcs/server_socket.<clstrname>
```

The Control Script (topsvcsctrl)

The Topology Services control script is contained is the executable file **/usr/sbin/rsct/bin/topsvcsctrl.** This script is normally invoked when the cluster manager is started. If necessary, you can invoke it directly from the command line.

The purpose of the topsvcsctrl script is to add the Topology Services subsystem on a node. You can also use the command to remove the subsystem from a node, start the subsystem, stop the subsystem.

Files and Directories

The Topology Services subsystem uses the following directories:

- **/var/ha/log,** for log files
- **/var/ha/run,** for Topology Services daemon current working directory
- **/var/ha/soc,** for socket files

The /var/ha/log Directory (Log Files)

The **/var/ha/log** directory contains trace output from the Topology Services daemon. The logs are named **topsvcs.DD.HHMMSS.ClstrName, where:**

- **DD** is the Day of the Month that this daemon was started.
- **HHMMSS** is the Hour, Minute, and Second of the day that the daemon was started.
- **ClstrName** is the name of the cluster as provided by the **cllsclstr** utility

The /var/ha/run Directory (Daemon Working Files)

In the **/var/ha/run** directory, a directory named topsvcs.ClstrName is created, where ClstrName is the name of the cluster as provided by the cllsclstr utility. This directory is the current working directory for the Topology Services daemon. If the Topology Services daemon abnormally terminates, the core dump file is placed in this directory. Whenever the Topology Services starts, it renames any core files to core.**DD.HHMMSS.ClstrName,** where:

- **DD** is the Day of the Month that this daemon was started.
- **HHMMSS** is the Hour, Minute, and Second of the day that the daemon was started.
- **ClstrName** is the name of the cluster as provided by the **cllsclstr** utility.

Components on Which Topology Services Depends

The Topology Services subsystem depends on the following components:

Subsystem Resource Controller An AIX subsystem that can be used to define and control subsystems. The Topology Services subsystem used by HACMP/ES is named topsvcs

Global ODM This is where HACMP/ES stores its entire configuration, including the network topology.

HACMP/ES Utilities Used by Topology Services to retrieve information from the Global ODM.

Configuring and Operating Topology Services

The following sections describe how the components of the Topology Services subsystem work together to provide topology services. Topics include:

- Configuring Topology Services
- Initializing Topology Services Daemon
- Operating Topology Services.

Configuring Topology Services

The Topology Services subsystem is added to the SRC as part of the HACMP/ES post-installation processing. There should be no other configuration necessary other than configuring the HACMP/ES topology, i.e. nodes, adapter, networks, addresses, etc. It is possible to modify the default interval between heart messages or then number of missed heartbeats that constitute a failure, but these can be considered as tuning to be done later. This tuning data is stored in an ODM class named HACMPtopsvcs. The following are the attributes in this class:

runFixedPri Run the daemon with a fixed priority. Since Topology Services is a real-time application, there is a need to avoid scheduling conflicts. A value of 1 indicates that the daemon is running with fixed priority. 0 indicates that it is not.

fixedPriLevel This is the actual fixed priority level that is used. The daemon will accept values greater to or equal to 31. The default is 38.

In addition to this there are some per-network type tunables stored in the HACMPnim class. The following is a list of attributes in that class:

name The name of the network type that these tunables are for, such as ether, atm, token.

desc A description of the network type, such as Ethernet Protocol

addrtype Not Used.

path Not Used.

para Not Used.

grace Grace period for IP address takeover in seconds.

hbrate Interval between heartbeats in microseconds.

cycle Number of missed heartbeats that signifies an adapter death,

The important ones here are the **hbrate** and the **cycle**; they provide the per-network values for **hbinterval** and **fibrillateCount**.

Starting and Stopping the Subsystem

The start and stop functions of **topsvcsctrl** command run the **startsrc** and **stopsrc** commands, respectively. This is the way that it is started during HACMP/ES cluster manager startup. The startup script calls **topsvcsctrl** to start the Topology Services subsystem. Likewise, the Topology Services subsystem is shutdown by using the **topsvcsctrl** command.

If it ever is necessary to stop the subsystem manually, use:

/usr/sbin/rsct/bin/topsvcsctrl –k. It shouldn’t be necessary to start the subsystem, since this is done by cluster manager startup, but if it is, use: **/usr/sbin/rsct/bin/topsvcsctrl –s**.

Deleting and Adding the Subsystem

Adding the subsystem should not be necessary, since the subsystem is added as part of post processing of HACMP/ES install. However, it could be necessary, if the subsystem, for some reason was deleted. To add the subsystem you would use: **/usr/sbin/rsct/bin/topsvcsctrl –a**. This command basically makes the subsystem known to the SRC and establishes everything that the SRC needs to start the subsystem. To delete the subsystem, which undoes what was done during the add, use: **/usr/sbin/rsct/bin/topsvcsctrl –d**.

Tracing the Subsystem

The tracing function of the **topsvcsctrl** command supplies additional problem determination information when it is requested by the IBM Support Center. Normally, you should not turn tracing on because it may slightly degrade Topology Services subsystem performance and can consume large amounts of disk space in the **/var** file system.

Initializing Topology Services Subsystem

When the Topology Services Subsystem is started it goes through a 2-stage initialization. The start-up script, **topsvcs** does the first stage, by executing the following steps:

1. The startup program obtains the number of the node on which it is running using the HACMP utility **clhandle**. This gives the daemon part of its identity and is the means by which it identifies which adapters are its local adapters.
2. The startup program also retrieves the cluster ID and the cluster name using the HACMP/ES utility, **cllsclstr**. This will enable the daemon to create files that have the cluster name or cluster name as part of the file name.
3. Each daemon builds a condensed version of the HACMP topology configuration using the output from the **cllsif** utility. This condensed version has basically node number, interface name, and IP address for every network in the cluster. For the case of non-IP connections it has path names to the devices. In addition to this there is subsystem specific information and tunable information that is passed to the daemon.
4. The startup program performs file maintenance in the Log directory and Run directory. This includes removing old log files, renaming new core files and removing any old core files.
5. The startup program executes the Topology Services daemon **hatsd.**

The daemon then continues the initialization with the following steps.

1. Read the current Machines List file and initialize internal data structures.
2. Initialize daemon to daemon communications, as well as client communications.
3. For each local adapter defined in the configuration, form a membership consisting of only the local adapter.

The daemon is now in its initialized state and ready to communicate with Topology Services daemons on other nodes in the cluster. The intent is to expand each singleton membership group formed during initialization to contain as many members as possible. Each adapter has an offset associated with it. Only other adapter membership groups with the same offset can join together to form a larger membership group. Eventually, as long as al the adapters configured for a particular network can communicate with each other, all adapters will belong to a single group.

Merging All Adapters into a Single Group

Initially the subsystem starts out as N singleton groups, one for each node. Each of those daemons is a Group Leader of those singleton groups and is aware of the existence of other adapters that could join his group, because of the configuration information. The next step is to begin proclaiming to subordinate nodes.

The proclaim logic tries to lure in members as efficiently as possible. For the first 3 proclaim cycles, daemons proclaim to only their own sub-net, and if the sub-net is broadcast capable, that message is broadcasted. The result of this is that given the previous assumption that all daemons started out as singletons, this would evolve into M groups, where M is the number of sub-nets that span this heartbeat ring. On the fourth proclaim cycle, those M group Leaders send proclaims to adapters that are outside of their local sub-net. This will cause a merging of groups into larger and larger groups until they have coalesced into a single group.

From the time the groups were formed, until they reach a stabilization point, the groups will be considered Unstable. The stabilization point is reached when a heartbeat ring has no group changes for the interval of 10 times the heartbeat send interval. Up to that point the proclaim will continue to operate on a 4 cycle system, where 3 cycles only proclaim to the local sub-nets, and on cycle proclaims to adapters not contained on the local sub-net. Then after the heartbeat ring has reached stability, proclaim messages will go out to all adapters not currently in the group regardless of the sub-net to which they belong.

Operating Topology Services Daemon

Normal operation of the Topology Services subsystem does not require administrative intervention. The subsystem is designed to recover from temporary failures, such as node failures or failures of individual Topology Services daemons. However, there may be some operational characteristics of interest to administrators.

The maximum node number allowed is 2048 and the maximum number of networks that it has capability of monitoring is 16. Note: A Daisy chain of non-IP or point to point networks will occupy multiple entries.

Topology Services is meant to be sensitive to network response and this sensitivity is tunable. However, other conditions may degrade Topology Services ability to accurately report on either adapter or node membership. One such condition is failure of AIX to schedule the daemon process in a timely manner. This can cause daemons to be late in sending their heartbeats by a significant amount. This can be because of too high of an interrupt rate, high rate of paging activity, or other problems. Anytime these conditions exist, analyze the problem carefully to fully understand it.

Also since Topology Services is a real-time process, intentionally subverting its use of the CPU, can only lead to false indication and should be avoided.

Topology Services Procedures

Normally, the Topology Services subsystem runs itself without requiring administrator intervention. However, on occasion, you may need to check the status of the subsystem.

Displaying the Status of the Topology Services Daemon

You can display the operational status of the Topology Services daemon by issuing the **lssrc** command. This output will only show status of the networks configured on the node where the command was run. Suppose network n1 connects nodes A, B, and C, and network n2 connects nodes B, C, and D. You can only monitor network n1 on node A and you can only monitor network n2 on node D, but you can monitor both networks on nodes B or C.

On a node, enter:

```
lssrc -ls topsvcs
```

In response, the **lssrc** command writes the status information to the standard output. The information includes:

- The information provided by the **lssrc –s** command (short form)
- One line of column headings for following network status lines

- Two lines for each Network for which this node has an adapter and includes the following:
  - The network name
  - The network index
  - The number of defined members/adapters that the configuration reported existing for this network
  - The number of members/adapters currently in the heartbeat ring
  - The state of the heartbeat ring, denoted by S, U, or D. S stands for Stable, U stands for unstable, and D stands for disabled.
  - Adapter ID, the address and instance number for the local adapter in this heartbeat ring
  - Group ID, the address and instance number of the heartbeat ring. The address of the heartbeat ring is also the address of the group leader.
- HB Interval, which is the interval in seconds between heartbeats. This exists both on a per network basis and a default value which could be different. The per network value overrides the default value for that network if it exists.
- HB Sensitivity, which is the number of missed heartbeats that constitute a failed adapter.
- The number of clients connected with process name and process id.
- Configuration Instance, the Instance number of the Machines List file.

The following is an example of the output from the **lssrc –ls topsvcs** command:

```
Subsystem         Group            PID     Status
 topsvcs          topsvcs          5734    active
 Network Name   Indx Defd Mbrs St Adapter ID      Group ID
 SP_ether_0     [ 0]    2    2  S 9.114.61.69     9.114.61.70
 SP_ether_0     [ 0]              0x35dc7e0a      0x35dc7e0e
 HB Interval = 1 secs. Sensitivity = 4 missed beats
 HPS_net_0      [ 2]    2    2  S 1.1.1.21        1.1.1.22
 HPS_net_0      [ 2]              0x35dc884d      0x35dc8888
 HB Interval = 1 secs. Sensitivity = 4 missed beats
   2 locally connected Clients with PIDs:
 haemd( 18412) hagsd( 23782)
   Configuration Instance = 6
   Default: HB Interval = 1 secs. Sensitivity = 4 missed beats
```

The two networks being monitored are named SP_ether and HPS_net. The 0 suffix is provided by topology services. It distinguishes between multiple heartbeat rings for a single network. One example of this is when there is a service ring and one or more standby rings. The SP_ether_0 ring has 2 defined adapters and both adapters are currently in the heartbeat ring. The same is true for the HPS_net_0 heartbeat ring. Also both heartbeat rings are in the Stable state. The heartbeat interval is 1 second and adapters are declared dead if 4 heartbeat messages are missed. That translates to no heartbeat messages in 8 seconds. There are 2 connected clients. They are Event Management (**haemd**) and Group Services (**hagsd**). The instance number for the Machines List file is 6.

Part 6 Volume 2 Appendixes

These appendixes contain reference material on script utilities and RSCT commands and messages, information on maintaining your cluster on a 7x24 basis, and information about the sample program VSM.

Appendix E Script Utilities

This appendix describes the utilities called by the event and startup scripts supplied with HACMP/ES. These utilities are general-purpose tools that can be called from any script, or from the AIX command line. The examples assume they are called in a script.

This appendix also includes the reference pages for Cluster Resource Group Information commands.

Utilities

The script utilities are stored in the **/usr/es/sbin/cluster/events/utils** directory. The utilities described in this chapter are grouped in the following categories:

- Disk utilities
- RS/6000 SP system utilities
- Filesystem and volume group utilities
- Logging utilities
- Network utilities
- DARE utilities
- Emulation utilities
- Security utilities.

Disk Utilities

cl_disk_available

Syntax

```
cl_disk_available diskname ...
```

Description

Checks to see if a disk named as an argument is currently available to the system, and, if not, makes the disk available.

Parameters

diskname	List of one of more disks to be made available; for example, *hdisk1*.

Return Values

0	Successfully made the specified disk available.
1	Failed to make the specified disk available.
2	Incorrect or bad arguments were used.

cl_fs2disk

Syntax

```
cl_fs2disk [-lvip] mount_point
```

or

```
cl_fs2disk -g volume_group
```

where **-l** identifies and returns the logical volume, **-v** returns the volume group, **-i** returns the physical volume ID, **-p** returns the physical volume, and **-g** is the mount point of a filesystem (given a volume group).

Description

Checks the ODM for the specified logical volume, volume group, physical volume ID, and physical volume information.

Parameters

mount point	Mount point of filesystem to check.
volume group	Volume group to check.

Return Values

0	Successfully retrieved filesystem information.
1	Failed to retrieve filesystem information.

cl_get_disk_vg_fs_pvids

Syntax

```
cl_get_disk_vg_fs_pvids [filesystem_list volumegroup_list]
```

Description

Given filesystems and/or volume groups, the function returns a list of the associated PVIDs.

Parameters

filesystem_list	The filesystems to check.
volumegroup_list	The volume groups to check.

Return Values

0	Success.
1	Failure.
2	Invalid arguments.

cl_is_array

Syntax

```
cl_is_array diskname
```

Description

Checks to see if a disk is a READI disk array.

Parameters

diskname	Single disk to test; for example, *hdisk1*.

Return Values

0	Disk is a READI disk array.
1	Disk is not a READI disk array.
2	An error occurred.

cl_is_scsidisk

Syntax

```
cl_is_scsidisk diskname
```

Description

Determines if a disk is a SCSI disk.

Parameters

diskname	Single disk to test; for example, *hdisk1*.

Return Values

0	Disk is a SCSI disk.
1	Disk is not a SCSI disk.
2	An error occurred.

cl_raid_vg

Syntax

```
cl_raid_vg volume_group
```

Description

Checks to see if the volume group is comprised of RAID disk arrays.

Parameters

volume_group	Single volume group to check.

Return Values

0	Successfully identified a RAID volume group.
1	Could not identify a RAID volume group; volume group must be 9333.
2	An error occurred. Mixed volume group identified.

cl_scdiskreset

Syntax

```
cl_scdiskreset /dev/diskname ...
```

Description

Issues a reset (SCSI ioctl) to each SCSI disk named as an argument.

Parameters

/dev/diskname	List of one or more SCSI disks.

Return Values

0	All specified disks have been reset.
-1	No disks have been reset.
n	Number of disks successfully reset.

cl_scdiskrsrv

Syntax

```
cl_scsidiskrsrv /dev/diskname ...
```

Description

Reserves the specified SCSI disk.

Parameters

/dev/diskname	List of one or more SCSI disks.

Return Values

0	All specified disks have been reserved.
-1	No disks have been reserved.
n	Number of disks successfully reserved.

cl_sync_vgs

Syntax

```
cl_sync_vgs -b|f volume_group ...
```

Description

Attempts to synchronize a volume group by calling **syncvg** for the specified volume group.

Parameters

volume_group	Volume group list.
-b	Background sync.
-f	Foreground sync.

Return Values

0	Successfully started **syncvg** for all specified volume groups.
1	The **syncvg** of at least one of the specified volume groups failed.
2	No arguments were passed.

scdiskutil

Syntax

```
scdiskutil -t /dev/diskname ...
```

Description

Tests and clears any pending SCSI disk status.

Parameters

-t	Tests to see if a unit is ready.
/dev/diskname	Single SCSI disk.

Return Values

-1	An error occurred.
0	The disk is not reserved.
>0	The disk is reserved.
2	No arguments were passed.

ssa_fence

Syntax

```
ssa_fence -e event pvid
```

Description

Fences a node in or out.

Additionally, this command also relies on environment variables; the first node up fences out all other nodes of the cluster regardless of their participation in the resource group.

If it is not the first node up, then the remote nodes fence in the node coming up. The node joining the cluster will not do anything.

If it is a **node_down** event, the remote nodes will fence out the node that is leaving. The node leaving the cluster will not do anything.

The last node going down clears the fence register.

Environment Variables

PRE_EVENT_MEMBERSHIP	Set by cluster manager.
POST_EVENT_MEMBERSHIP	Set by cluster manager.
EVENT_ON_NODE	Set by calling script.

Parameters

-e event	1=up; 2=down.
pvid	Physical volume ID on which fencing will occur.

Return Values

0	Success.
1	Failure. A problem occurred during execution. A message describing the problem is written to stderr and to the cluster log file.
2	Failure. Invalid number of arguments. A message describing the problem is written to stderr and to the cluster log file.

ssa_clear

Syntax

```
ssa_clear -x | -d pvid
```

Description

Clears or displays the contents of the fence register. If **-d** is used, a list of fenced out nodes will be displayed. If **-x** is used, the fence register will be cleared.

Note: This command exposes data integrity of a disk, by unconditionally clearing its fencing register. It requires adequate operator controls and warnings, and should *not* be included within any takeover script.

Return Values

0	Success.
1	Failure. A problem occurred during execution. A message describing the problem is written to stderr and to the cluster log file.
2	Failure. Invalid number of arguments. A message describing the problem is written to stderr and to the cluster log file.

ssa_clear_all

Syntax

```
ssa_clear_all pvid1, pvid2 ... pvidn
```

Description

Clears the fence register on multiple physical volumes.

Return Values

0	Success.
1	Failure. A problem occurred during execution. A message describing the problem is written to stderr and to the cluster log file.
2	Failure. Invalid number of arguments. A message describing the problem is written to stderr and to the cluster log file.

ssa_configure

Syntax

```
ssa_configure
```

Description

Assigns unique node IDs to all the nodes of the cluster. Then it configures and unconfigures all SSA pdisks and hdisks on *all* nodes thus activating SSA fencing. This command is called from the SMIT screen during the sync of a node environment. If this command fails for any reason, that node should be rebooted.

Return Values

0	Success.
1	Failure. A problem occurred during execution. A message describing the problem is written to stderr and to the cluster log file.

RS/6000 SP Utilities

cl_swap_HPS_IP_address

Syntax

```
cl_swap_HPS_IP_address interface address netmask [delete]
```

Description

This script is used to specify an alias address to an SP Switch interface, or remove an alias address, during IP address takeover. Note that adapter swapping does not make sense for the SP Switch since all addresses are alias addresses on the same adapter.

Parameters

interface	The name of the interface.
alias	The alias address.
netmask	Netmask.
delete	When this parameter is specified, the alias is removed.

Return Values

0	Success.
1	The adapter could not be configured (using the **ifconfig** command) at the specified address.
2	Invalid syntax.

Examples

The following example specifies the alias *1.1.1.1* for the *css0* interface.

```
cl_swap_HPS_IP_address css0 1.1.1.1 255.255.255.128
```

The following example removes the alias *1.1.1.1* from the *css0* interface.

```
cl_swap_HPS_IP_address css0 1.1.1.1 255.255.255.128 delete
```

Filesystem and Volume Group Utilities

cl_activate_fs

Syntax

```
cl_activate_fs /filesystem_mount_point
```

Description

Mounts the filesystems passed as arguments.

Parameters

/filesystem_mount_point	A list of one or more filesystems to mount.

Return Values

0	All filesystems named as arguments were either already mounted or were successfully mounted.
1	One or more filesystems failed to **mount**.
2	No arguments were passed.

cl_activate_nfs

Syntax

```
cl_activate_nfs retry host /filesystem_mount_point
```

Description

NFS-mounts the filesystems passed as arguments. The routine backgrounds the mounts rather than specifying the **bg** option to the **mount** command because the standard **mount** command fails if the server host is not ready (that is, has not exported the filesystem). This command will retry in the background. Note that this route assumes the filesystem is already mounted if any mounted filesystem has a matching name.

Parameters

retry	Number of attempts. Sleeps 15 seconds between attempts.
host	NFS server host.
/filesystem_mount_point	List of one or more filesystems to activate.

Return Values

0	All filesystems passed were either mounted or mounts were scheduled in the background.
1	One or more filesystems failed to mount.
2	No arguments were passed.

cl_activate_vgs

Syntax

```
cl_activate_vgs [-n] volume_group_to_activate ...
```

Description

Initiates a **varyonvg** of the volume groups passed as arguments.

Parameters

-n	Do not sync the volume group when **varyon** is called.
volume_group_to_activate	List of one of more volume groups to activate.

Return Values

0	All of the volume groups are successfully varied on.
1	The **varyonvg** of at least one volume group failed.
2	No arguments were passed.

cl_deactivate_fs

Syntax

```
cl_deactivate_fs /filesystem_mount_point
```

Description

Attempts to **unmount** any filesystem passed as an argument that is currently mounted.

Parameters

/filesystem_mount_point	List of one or more filesystems to unmount.

Return Values

0	All filesystems were successfully unmounted.
1	One or more filesystems failed to unmount.
2	No arguments were passed.

cl_deactivate_nfs

Syntax

```
cl_deactivate_nfs file_system_to_deactivate ...
```

Description

Attempts to **unmount -f** any filesystem passed as an argument that is currently mounted.

Parameters

file_system_to_deactivate	List of one or more NFS-mounted filesystems to unmount.

Return Values

0 Successfully unmounted a specified filesystem.

2 No arguments were passed.

cl_deactivate_vgs

Syntax

```
cl_deactivate_vgs volume_group_to_deactivate ...
```

Description

Initiates a **varyoffvg** of any volume group that is currently varied on and that was passed as an argument.

Parameters

volume_group_to_deactivate List of one or more volume groups to vary off.

Return Values

0 All of the volume groups are successfully varied off.

1 The **varyoffvg** of at least one volume group failed.

2 No arguments were passed.

cl_export_fs

Syntax

```
cl_export_fs hostname file_system_to_export ...
```

Description

NFS-exports the filesystems given as arguments so that NFS clients can continue to work.

Parameters

hostname Hostname of host given root access.

filesystem_to_export List of one or more filesystems to NFS-export.

Return Values

0 Successfully exported all filesystems specified.

1 A runtime error occurred: unable to export or unable to startsrc failures.

2 No arguments were passed.

cl_nfskill

Syntax

```
cl_nfskill [-k] [-t] [-u] directory ...
```

Description

Lists the process numbers of local processes using the specified NFS directory.

Find and kill processes that are executables fetched from the NFS-mounted filesystem. Only the root user can kill a process of another user.

If you specify the **-t** flag, all processes that have certain NFS module names within their stack will be killed.

Warning: When using the **-t** flag it is not possible to tell which NFS filesystem the process is related to. This could result in killing processes which belong to NFS-mounted filesystems other than those which are cross-mounted from another HACMP node and under HACMP control. This could also mean that the processes found could be related to filesystems under HACMP control but not part of the current resources being taken. This flag should therefore be used with caution and only if you know you have a specific problem with umounting the NFS filesystems.

To help to control this, the **cl_deactivate_nfs** script contains the normal calls to **cl_nfskill** with the **-k** and **-u** flags and commented calls using the **-t** flag as well. If you choose to use the **-t** flag, you should uncomment those calls and comment the original calls.

Parameters

-k Sends the SIGKILL signal to each local process,

-u Provides the login name for local processes in parentheses after the process number.

-t Finds and kills processes that are just opening on NFS filesystems.

directory Lists of one or more NFS directories to check.

Return Values

None.

Logging Utilities

cl_log

Syntax

```
cl_log message_id, default_message, variables
```

Description

Logs messages to **syslog** and standard error.

Parameters

message_id	Message ID for the messages to be logged.
default_message	Default message to be logged.
variables	List of one or more variables to be logged.

Return Values

0	Successfully logged messages to **syslog** and standard error.
2	No arguments were passed.

Network Utilities

cl_echo

Syntax

```
cl_echo message_id, default_message, variables
```

Description

Logs messages to standard error.

Parameters

message_id	Message ID for the messages to be displayed.
default_message	Default message to be displayed.
variables	List of one or more variables to be displayed.

Return Values

0	Successfully displayed messages to stdout.
2	No arguments were passed.

cl_nm_nis_off

Syntax

```
cl_nm_nis_off
```

Description

Turns off name serving and NIS.

Parameters

None

Return Values

0	Successfully turned off name serving and NIS client services.
1	Could not turn off name serving, or could not turn off NIS.

cl_nm_nis_on

Syntax

```
cl_nm_nis_on
```

Description

Turns on name serving and NIS.

Parameters

None.

Return Values

0	Successfully turned on name serving and NIS client services.
1	Could not turn on name serving, NIS, or both.

cl_swap_HW_address

Syntax

```
cl_swap_HW_address address interface
```

Description

Checks to see if an alternate hardware address is specified for the address passed as the first argument. If so, it assigns the hardware address specified to the new adapter.

Parameters

address	Interface address or IP label.
interface	Interface name (for example, *en0* or *tr0*).

Return Values

0	Successfully assigned the specified hardware address to an adapter.
1	Could not assign the specified hardware address to an adapter.
2	Wrong number of arguments were passed.

Note: This utility is used during adapter swap and IP address takeover.

cl_swap_IP_address

Syntax

```
cl_swap_IP_address interface address netmask
cl_swap_IP_address interface1 address1 interface2 address2 netmask
```

Description

This routine is used during adapter swap and IP address takeover.

In the first form, the routine sets the specified interface to the specified address:

```
cl_swap_IP_address en0 1.1.1.1 255.255.255.128
```

In the second form, the routine sets two interfaces in a single call. This is sometimes necessary due to AIX routing quirks where an existing route gets spuriously deleted. An example where this is required is the case of swapping two interfaces:

```
cl_swap_IP_address en0 1.1.1.1 en1 2.2.2.2 255.255.255.128
```

Parameters

interface	Interface name.
address	IP address.
netmask	Network mask. Must be in decimal format.

Return Values

0	Successfully swapped IP addresses.
1	**ifconfig** failed.
2	Wrong or incorrect number of arguments.

Note: This utility is used for swapping the IP address of either a standby adapter with a local service adapter (called adapter swapping), or a standby adapter with a remote service adapter (called masquerading). For masquerading, the **cl_swap_IP_address** routine should sometimes be called *before* processes are stopped, and sometimes *after* processes are stopped. This is application dependent. Some applications respond better if they shutdown *before* the network connection is broken, and some respond better if the network connection is closed first.

cl_unswap_HW_address

Syntax

```
cl_unswap_HW_address interface
```

Description

Script used during adapter swap and IP address takeover. It restores an adapter to its boot address.

Parameters

interface	Interface name (for example, *en0* or *tr0*).

Return Values

0	Success.
1	Failure.
2	Invalid parameters.

DARE Utilities

cldare

Syntax

```
cldare [-n] [-u] [-i] [-r] [-t] [-v] [-f] [-M resgroup:[location|
[stop|default]] [:sticky] ...
```

Description

Updates the HACMP/ES Cluster Manager and associated daemons with new configuration information, provides an emulation feature to emulate cluster resource and cluster topology dynamic reconfiguration, allows resource group migration to other cluster nodes while the cluster is running.

Note: When run in emulation mode (using the -f flag), **cldare** makes no changes to the configuration information.

Parameters

-n If this flag is specified, cluster resources removed from the configuration will keep their configuration as part of the DARE event. The default behavior is for the resources not to keep their configuration.

-u If this flag is specified, the DARE will remove the stage ODM lock on all nodes. This may be used in case of a failure to remove the lock.

-i If this flag is specified, the DARE event will be performed whether or not the cluster verification finds configuration errors.

-r
If this flag is specified, Cluster Resources will be synchronized across cluster node.

-t
If this flag is specified, Cluster Topology will be synchronized across cluster nodes.

-v
If this flag is specified, verification will not run.

-f
If this flag is specified, the **cldare** command will run as an emulation only and no changes will be made to the Cluster Manager. You must specify either the **-r** flag to indicate the emulation of Synchronizing Cluster Resources, or the **-t** flag to indicate the emulation of Synchronizing Cluster Topology. The **-u** flag and the **-v** flag cannot be used if you specify emulation mode.

-M ***resgroup:*****[location | [stop | default]] [:sticky]**
Causes one or more specified resource groups migrate to other nodes. The stop keyword brings the resource group offline, taking down any service label, unmounting filesystems, etc. The use of the default keyword instead of a location causes the resource group to be started/moved to a cluster node based on the node priority.

The keyword *sticky* defines the start, stop, or move operation as sticky, causing the resource group to remain on a node and attempt to return to that node after fallover and reintegration. The sticky designation supersedes the normal resource policy and node priority list.

Emulation Utilities

cl_emulate

Syntax

```
cl_emulate -e node_up -n nodename
cl_emulate -e node_down -n nodename {f|g|t}
cl_emulate -e network_up -w networkname -n nodename
cl_emulate -e network_down -w networkname -n nodename
cl_emulate -e join_standby -n nodename -a ip_label
cl_emulate -e fail_standby -n nodename -a ip_label
cl_emulate -e swap_adapter -n nodename -w network -a ip_label
-d ip_label
```

Description

Emulates a specific cluster event and outputs the result of the emulation. The output is shown on the screen as the emulation runs, and is saved to an output file on the node from which the emulation was executed.

The Event Emulation utility does not run customized scripts such as pre- and post- event scripts. In the output file the script is echoed and the syntax is checked, so you can predict possible errors in the script. However, if customized scripts exist, the outcome of running the actual event may differ from the outcome of the emulation

When emulating an event which contains a customized script, the Event Emulator uses the **ksh** flags **-n** and **-v**. The **-n** flag reads commands and checks them for syntax errors, but does not execute them. The **-v** flag indicates verbose mode. When writing customized scripts that may be accessed during an emulation, be aware that the other **ksh** flags may not be compatible with the **-n** flag and may cause unpredictable results during the emulation. See the **ksh** man page for flag descriptions.

You can run only one instance of an event emulation at a time. If you attempt to start an emulation while an emulation is already running on a cluster, the integrity of the output cannot be guaranteed.

Parameters

-e ***eventname*** The name of the event to emulate: **node_up**, **node_down**, **network_up**, **network_down**, **join standby**, **fail standby**, **swap adapter**.

-n ***nodename*** The node name used in the emulation.

-f Forced shut down emulation. Cluster daemons terminate without running any local procedures.

-g Graceful shutdown with no takeover emulation.

-t Graceful shutdown emulation with the resources being released by this node and taken over by another node.

-w ***networkname*** The network name used in the emulation.

-a ***ip_label*** The standby adapter address with which to switch.

-d ***ip_label*** The service adapter to fail.

Note: The **cldare** command also provides an emulation feature for dynamic reconfiguration events. See DARE Utilities on page E-16 for more information on **cldare**.

Security Utilities

To simplify and automate the process of creating a secure system, two scripts for setting up Kerberos service principals are provided with HACMP/ES version 4.4:

- **cl_setup_kerberos**—Extracts the HACMP adapter labels from an already configured node and creates a file, **cl_krb_service**, that contains all of the HACMP adapter labels and additional format information required by the **add_principal** Kerberos setup utility. Also creates the **cl_adapters** file that contains a list of the adapters required to extract the service principals from the authentication database.
- **cl_ext_krb**—Prompts the user to enter the Kerberos password to be used for the new principals, and uses this password to update the **cl_krb_service** file. Checks for a valid **.k** file and alerts the user if one doesn't exist. Once a valid **.k** file is found, the **cl_ext_krb** script runs the **add_principal** utility to add all the adapter labels from the **cl_krb_service** file into the authentication database; extracts the service principals and places them in a

new Kerberos services file, **cl_krb-srvtab**; creates the **cl_klogin** file that contains additional entries required by the **.klogin** file; updates the **.klogin** file on the control workstation and all nodes in the cluster; and concatenates the **cl_krb-srvtab** file to each node's **/etc/krb-srvtab** file.

Cluster Resource Group Information Commands

The following commands are available for use by scripts or for execution from the command line:

- **clRMupdate**
- **clRGinfo**

HACMP/ES event scripts use the **clRMupdate** command to notify the Cluster Manager that it should process an event. Users or scripts can execute the **clRGinfo** command to get information about resource group status and location.

Warning: The resource group information is used by the event scripts to determine resource group movement during cluster events. Manually updating the resource group information may lead to incorrect fallover operations.

clRMupdate Command

Synopsis

```
clRMupdate {acquiring | releasing | rg_up | rg_down | rg_error } name
```

Description

The **clRMupdate** command is used by the HACMP event scripts to notify the Cluster Manager that it should process an event. These events are used internally by the Cluster Manager so it can report the current state and location of resource groups.

Note: When using this command to change the state of a resource group, this method will change the state of the resource group relative to the local node. The state of the resource group relative to a remote node cannot be changed without running this command on the remote node.

For example, running the following command on node n1 will change the state of the resource group "rg1" to ONLINE on node n1:

```
$ clRMupdate rg_up rg1
```

The exit values for this command are:

0	Success
1	Operation could not be performed

Parameters

acquiring rg_name	Resource group *rg_name* is being acquired
releasing rg_name	Resource group *rg_name* is being released
rg_up rg_name	Resource group *rg_name* is now up
rg_down rg_name	Resource group *rg_name* is now down
rg_error rg_name	Script failed – set rg to error state
node_error nodename	Script failed – set node to error state

See Also

`clRGinfo`

clRGinfo Command

Synopsis

```
clRGinfo [ -s ] [ groupname1 ] [ groupname2 ]
clRGinfo [ -h ]
```

Description

Use the **clRGinfo** command to display the location and state of the specified resource groups.

The exit values for this command are:

0	Success
1	Operation could not be performed

Parameters

s	Display in colon (shortened) format.
h	Print usage message.

A sample report is shown below in normal format.

```
---------------------------------------------------
GroupName      State           Location
---------------------------------------------------
group1         ONLINE          node1
               ONLINE          node2
               ACQUIRING       node3
               OFFLINE         node4
group2         ONLINE          node3
group3         RELEASING       node2
```

A condensed colon separated format follows.

```
group1:ONLINE:node1
group1:ONLINE:node2
group1:ACQUIRING:node3
group1:OFFLINE:node4
group2:ONLINE:node3
group3:RELEASING:node2
```

See Also

`clRMupdate`

Appendix F RSCT Commands and Utilities

This appendix describes the RSCT common commands and utilities.

emsvcsctrl Script

A control script that starts the Event Management subsystem.

Syntax

```
emsvcsctrl {-a | -s | -k | -d | -c | -t | -o | -r | -h}
```

-a	Adds the subsystem.
-s	Starts the subsystem.
-k	Stops the subsystem.
-d	Deletes the subsystem.
-c	Cleans the subsystems.
-t	Turns tracing on for the subsystem.
-o	Turns tracing off for the subsystem.
-r	Refreshes the subsystem.
-h	Displays usage information.

Operands

None.

Description

Event Management is a distributed subsystem of RSCT that provides a set of high availability services for the IBM RS/6000. By matching information about the state of system resources with information about resource conditions that are of interest to client programs, it creates events. Client programs can use events to detect and recover from system failures, thus enhancing the availability of the system.

The **emsvcsctrl** control script controls the operation of the Event Management subsystem. The subsystem is under the control of the System Resource Controller (SRC) and belongs to a subsystem group called **emsvcs**. Associated with each subsystem is a daemon.

The **emsvcsctrl** script also controls the operation of the AIX Resource Monitor subsystem. The subsystem is under SRC control and also belongs to the **emsvcs** subsystem group. Associated with each subsystem is a daemon.

Instances of the Event Management and AIX Resource Monitor subsystems execute on each node in the HACMP/ES cluster.

From an operational point of view, the Event Management subsystem group is organized as follows:

Subsystem Event Management

Subsystem Group **emsvcs**

SRC Subsystem **emsvcs**.The **emsvcs** subsystem is associated with the **haemd** daemon.

emaixos The **emaixos** is associated with the **harmad** daemon

Daemons The **haemd** daemon provides the Event Management services. The **harmad** daemon is the resource monitor for AIX operating system resources.

The **emsvcsctrl** script is not normally executed from the command line. It is normally called by the HACMP/ES startup script command during installation of the system.

The **emsvcsctrl** script provides a variety of controls for operating the Event Management subsystem:

- Adding, starting, stopping, and deleting the subsystem
- Cleaning up the subsystems
- Turning tracing on and off.

Adding the Subsystem

When the **-a** flag is specified, the control script uses the **mkssys** command to add the Event Management and AIX Resource Monitor subsystems to the SRC. The control script operates as follows:

1. It makes sure that the **emsvcs** and **emaixos** subsystems are stopped.
2. It removes the **emsvcs** and **emaixos** subsystems from the SRC (just in case they are still there).
3. It adds the **emsvcs** subsystem to the SRC.
4. It adds the **emaixos** subsystem to the SRC.
5. It adds **haemrm** group using the **mkgroup** command, if it does not already exist. Any errors that occur are written to a log file named **/var/ha/log/em.mkgroup**.
6. It creates the **/var/ha/lck/haem** and **/var/ha/soc/haem** directories, if they don't already exist. Any errors that occur are written to a log file named **/var/ha/log/em.mkdir**.
7. It copies the Event Management Configuration Database, (EMCDB) from its install location, **/usr/sbin/rsct/install/config/em.HACMP.cdb** to its run-time location, **/etc/ha/cfg/em.HACMP.cdb**. Any errors resulting from the copy are written to a log file named **/var/ha/log/em.cp**.

Starting the Subsystem

When the **-s** flag is specified, the control script uses the **startsrc** command to start the Event Management subsystem, **emsvcs**, and the AIX Resource Monitor subsystem, **emaixos**.

Stopping the Subsystem

When the **-k** flag is specified, the control script uses the **stopsrc** command to stop the Event Management subsystem, **emsvcs**, and the AIX Resource Monitor subsystem, **emaixos**.

Deleting the Subsystem

When the **-d** flag is specified, the control script uses the **rmssys** command to remove the Event Management and AIX Resource Monitor subsystems from the SRC. The control script operates as follows:

1. It makes sure that the **emsvcs** and **emaixos** subsystems are stopped.
2. It removes the **emsvcs** and **emaixos** subsystems from the SRC using the **rmssys** command.

Cleaning Up the Subsystems

When the **-c** flag is specified, the control script stops and removes the Event Management subsystems for all system partitions from the SRC. The control script operates as follows:

1. It stops all instances of subsystems in the subsystem group by using the **stopsrc -g emsvcs** command.
2. It removes all instances of subsystems in the subsystem group from the SRC using the **rmssys** command.
3. It removes the Event Management Configuration Database (EMCDB) from its run-time location, **/etc/ha/cfg/em.HACMP.cdb**.

Turning Tracing On

When the **-t** flag is specified, the control script turns tracing on for the **haemd** daemon, using the **haemtrcon** command. Tracing for the **harmad** daemon is also enabled, using the **traceson** command.

Turning Tracing Off

When the **-o** flag is specified, the control script turns tracing off for the **haemd** daemon, using the **haemtrcoff** command. Tracing for the **harmad** daemon is also disabled, using the **tracesoff** command.

Refreshing the Subsystem

The **-r** flag has no effect for this subsystem.

Logging

While it is running, the Event Management daemon normally provides information about its operation and errors by writing entries to the AIX error log. If it cannot, errors are written to a log file called **/var/ha/log/em.default.cluster_name**.

Files

/var/ha/log/em.default.*cluster_name* Contains the default log of the **haemd** daemon on the cluster named *cluster_name*.

/var/ha/log/em.cp Contains a log of any errors that occurred while copying the Event Management Configuration Database.

/var/ha/log/em.trace.*cluster_name* Contains the trace log of the **haemd** daemon on the cluster named *cluster_name*.

/var/ha/log/em.mkgroup Contains a log of any errors that occurred while creating the **haemrm** group.

/var/ha/log/em.mkdir Contains a log of any errors that occurred while creating the /var/ha/lck/haem and **/var/ha/soc/haem** directories.

Standard Error

This command writes error messages (as necessary) to standard error.

Exit Values

0 Indicates the successful completion of the command.

1 Indicates that an error occurred.

Security

You must be running with an effective user ID of root.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 31, The Event Management Subsystem
- *IBM RS/6000 Cluster Technology: Event Management Programming Guide and Reference*
- *AIX Version 4 Commands Reference.*
- Information about the System Resource Controller (SRC) in *AIX Version 4 General Programming Concepts: Writing and Debugging Programs*

Location

/usr/sbin/rsct/bin/emsvcsctrl

Related Information

haemd, haemtrcoff, haemtrcon, lssrc, startsrc, stopsrc

Examples

1. To add the Event Management subsystem to the SRC, enter:

   ```
   emsvcsctrl -a
   ```

2. To start the Event Management subsystem, enter:

   ```
   emsvcsctrl -s
   ```

3. To stop the Event Management subsystem, enter:

   ```
   emsvcsctrl -k
   ```

4. To delete the Event Management subsystem from the SRC, enter:

   ```
   emsvcsctrl -d
   ```

5. To clean up the Event Management subsystem, enter:

   ```
   emsvcsctrl -c
   ```

6. To turn tracing on for the Event Management daemon, enter:

   ```
   emsvcsctrl -t
   ```

7. To turn tracing off for the Event Management daemon, enter:

   ```
   emsvcsctrl -o
   ```

8. To display the status of all of the subsystems in the Event Management SRC group, enter:

   ```
   lssrc -g emsvcs
   ```

9. To display the status of an individual Event Management subsystem on a node, enter:

   ```
   lssrc -s emsvcs
   ```

10. To display detailed status about an individual Event Management subsystem on a node, enter:

    ```
    lssrc -l -s emsvcs
    ```

In response, the system returns information that includes the running status of the subsystem, the settings of trace flags, the version number of the Event Management Configuration Database, the time the subsystem was started, the connection status to Group Services and peer Event Management subsystem, and the connection status to Event Management clients, if any.

11. To display the status of all of the daemons under SRC control, enter:

    ```
    lssrc -a
    ```

grpsvcsctrl Script

A control script that starts the Group Services subsystems.

Syntax

```
grpsvcsctrl {-a | -s | -k | -d | -c | -u | -t | -o | -r | -h}
```

-a Adds the subsystems.

-s Starts the subsystems.

-k Stops the subsystems.

-d Deletes the subsystems.

-c Cleans the subsystems, that is, delete them from all system partitions.

-t Turns tracing on for the subsystems.

-o Turns tracing off for the subsystems.

-r Refreshes the subsystem.

-h Displays usage information.

Operands

None.

Description

Group Services provides distributed coordination and synchronization services for other distributed subsystems running on a set of nodes in the HACMP/ES cluster. The **grpsvcsctrl** control script controls the operation of the subsystems that are required for Group Services. These subsystems are under the control of the System Resource Controller (SRC) and belong to a subsystem group called **grpsvcs**. A daemon is associated with each subsystem.

From an operational point of view, the Group Services subsystem group is organized as follows:

Subsystem Group Services

Subsystem Group **grpsvcs**

SRC Subsystems **grpsvcs** and **grpsvcsglsm** The **grpsvcs** subsystem is associated with the **hagsd** daemon. The **grpsvcsglsm** subsystem is associated with the **hagsglsmd** daemon.

The subsystem names on the nodes are **grpsvcs** and **grpsvcsglsm**. There is one of each subsystem per node and it is associated with the cluster to which the node belongs.

Daemons	**hagsd** and **hagsglsmd** The **hagsd** daemon provides the majority of the Group Services functions.The **hagsglsmd** daemon provides global synchronization services for the switch adapter membership group.

The **grpsvcsctrl** script is not normally executed from the command line. It is normally called by the startup command during installation of the cluster.

The **grpsvcsctrl** script provides a variety of controls for operating the Group Services subsystems:

- Adding, starting, stopping, and deleting the subsystems
- Turning tracing on and off

Before performing any of these functions, the script obtains the current cluster name.

Adding the Subsystem

When the **-a** flag is specified, the control script uses the **mkssys** command to add the Group Services subsystems to the SRC. The control script operates as follows:

1. It makes sure that both the **grpsvcs** and **grpsvcsglsm** subsystems are stopped.
2. It gets the port number for the **grpsvcs** subsystem for this cluster from the global ODM and ensures that the port number is set in the **/etc/services** file. The range of valid port numbers is 10000 to 10100, inclusive.

   The service name that is entered in the **/etc/services** file is **grpsvcs.*cluster_name***.
3. It removes the **grpsvcs** and **grpsvcsglsm** subsystems from the SRC (just in case they are still there).
4. It adds the **grpsvcs** and **grpsvcsglsm** subsystems to the SRC. The cluster name is configured as a daemon parameter on the **mkssys** command.

Starting the Subsystem

When the **-s** flag is specified, the control script uses the **startsrc** command to start the Group Services subsystems, **grpsvcs** and **grpsvcsglsm**.

Stopping the Subsystem

When the **-k** flag is specified, the control script uses the **stopsrc** command to stop the Group Services subsystems, **grpsvcs** and **grpsvcsglsm**.

Deleting the Subsystem

When the **-d** flag is specified, the control script uses the **rmssys** command to remove the Group Services subsystems from the SRC. The control script operates as follows:

1. It makes sure that both the **grpsvcs** and **grpsvcsglsm** subsystems are stopped.
2. It removes the **grpsvcs** and **grpsvcsglsm** subsystems from the SRC using the **rmssys** command.
3. It removes the port number from the **/etc/services** file.

Cleaning Up the Subsystems

When the **-c** flag is specified, the control script stops and removes the Group Services subsystems for all system partitions from the SRC. The control script operates as follows:

1. It stops all instances of subsystems in the subsystem group in all partitions, using the **stopsrc -g grpsvcs** command.
2. It removes all instances of subsystems in the subsystem group in all partitions from the SRC using the **rmssys** command.

Turning Tracing On

When the **-t** flag is specified, the control script turns tracing on for the **hagsd** daemon, using the traceson command. Tracing is not available for the hagsglsmd daemon.

Turning Tracing Off

When the -o flag is specified, the control script turns tracing off (returns it to its default level) for the **hagsd** daemon, using the **tracesoff** command. Tracing is not available for the **hagsglsmd** daemon.

Refreshing the Subsystem

The **-r** flag has no effect for this subsystem.

Logging

While they are running, the Group Services daemons provide information about their operation and errors by writing entries in a log file in the **/var/ha/log** directory.

Each daemon limits the log size to a pre-established number of lines (by default, 5,000 lines). When the limit is reached, the daemon appends the string .**bak** to the name of the current log file and begins a new log. If a .**bak** version already exists, it is removed before the current log is renamed.

Files

/var/ha/log/grpsvcs_nodenum_instnum.cluster_name

Contains the log of the **hagsd** daemons on the nodes.

/var/ha/log/grpsvcsglsm_nodenum_instnum.cluster_name

Contains the log of the **hagsglsmd** daemons on the nodes.

The file names include the following variables:

- nodenum is the node number on which the daemon is running
- instnum is the instance number of the daemon
- *cluster_name* is the name of the cluster in which the daemon is running.

Error Notification

This command writes error messages (as necessary) to standard error.

0	Indicates the successful completion of the command.
1	Indicates that an error occurred.

You must be running with an effective user ID of root.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 30, The Group Services Subsystem
- *AIX Version 4 Commands Reference.*
- Information about the System Resource Controller (SRC) in *AIX Version 4 General Programming Concepts: Writing and Debugging Programs*

Location

/usr/sbin/rsct/bin/grpsvcsctrl

Related Information

`hagsd, hagsglsmd, lssrc, startsrc, stopsrc`

Examples

1. To add the Group Services subsystems to the SRC enter:
   ```
   grpsvcsctrl -a
   ```
2. To start the Group Services subsystems, enter:
   ```
   grpsvcsctrl -s
   ```
3. To stop the Group Services subsystems, enter:
   ```
   grpsvcsctrl -k
   ```
4. To delete the Group Services subsystems from the SRC, enter:
   ```
   grpsvcsctrl -d
   ```
5. To clean up the Group Services subsystems, enter:
   ```
   grpsvcsctrl -c
   ```
6. To turn tracing on for the Group Services daemon, enter:
   ```
   grpsvcsctrl -t
   ```
7. To turn tracing off for the Group Services daemon, enter:
   ```
   grpsvcsctrl -o
   ```
8. To display the status of all of the subsystems in the Group Services SRC group, enter:
   ```
   lssrc -g grpsvcs
   ```

9. To display the status of an individual Group Services subsystem, enter:

   ```
   lssrc -s subsystem_name
   ```

10. To display detailed status about an individual Group Services subsystem, enter:

    ```
    lssrc -l -s subsystem_name
    ```

    In response, the system returns information that includes the running status of the subsystem, the number and identity of connected GS clients, information about the Group Services domain, and the number of providers and subscribers in established groups.

11. To display the status of all of the daemons under SRC control, enter:

    ```
    lssrc -a
    ```

haemd Daemon

The Event Manager daemon, which observes resource variable instances that are updated by Resource Monitors and generates and reports events to client programs.

Syntax

```
haemd
```

The daemon has no specifiable flags or operands.

Description

The **haemd** daemon is the Event Manager daemon. The daemon observes resource variable instances that are updated by Resource Monitors and generates and reports events to client programs.

One instance of the **haemd** daemon executes on every node of a cluster. The **haemd** daemon is under System Resource Controller (SRC) control.

Because the daemon is under SRC control, it cannot be started directly from the command line. It is normally started by the **emsvcsctrl** command. If you must start or stop the daemon directly, use the **emsvcsctrl** command.

When SRC creates the **haemd** daemon, the actual program started is **haemd_HACMP**. The **haemd_HACMP** program, after collecting information needed by the daemon, then executes the **haemd** program. In other words, the **haemd_HACMP** program is replaced by the **haemd** program in the process created by SRC.

For more information about the Event Manager daemon, see the **emsvcsctrl** man page.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 31, The Event Management Subsystem
- *IBM RS/6000 Cluster Technology: Event Management Programming Guide and Reference*
- *AIX Version 4 Commands Reference.*

Location

/usr/sbin/rsct/bin/haemd

Related Information

emsvcsctrl and **haemd_HACMP**

Examples

See the **emsvcsctrl** command.

haemd_HACMP

Start-up program for the Event Manager daemon.

Syntax

```
haemd_HACMP [-d trace_arg ]°
```

-d **trace_arg**	Enables tracing for the daemon activity specified by trace_arg. This flag may be specified multiple times.

This command has no operands

Description

The **haemd_HACMP** program is the start-up program for the **haemd** daemon. When the Event Management subsystem is configured in the System Resource Controller (SRC) by the **emsvcsctrl** command, **haemd_HACMP** is specified as the program to be started.

This program can only be invoked by the SRC. To start the Event Management subsystem use the **emsvcsctrl** command.

The **-d** flag should only be used under the direction of the IBM Support Center. The possible trace arguments are the same as for the **haemtrcon** command, except for **regs** and **dinsts**. To use this flag the **emsvcs** subsystem definition in the SRC must be changed using the **chssys** command with the **-a** argument. Then the daemon must be stopped and then restarted.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 31, The Event Management Subsystem
- *IBM RS/6000 Cluster Technology: Event Management Programming Guide and Reference*
- *AIX Version 4 Commands Reference.*

Location

/usr/sbin/rsct/bin/haemd_HACMP

Related Information

emsvcsctrl, haemd and **haemtrcon**

Examples

See the **emsvcsctrl** command.

haemqvar

Queries resource variables.

Syntax

```
haemqvar [-H domain | -S domain] [-c | -d | -i] [-f file] [-h]
                            [class  var  rsrcID [º]]
```

-H *domain* Queries resource variables in the HACMP domain specified by *domain*.

-S *domain* Queries resource variables in the SP domain specified by *domain*.

-c Queries current resource variable values.

-d Queries resource variable definitions but produces short form output.

-i Queries instances of resource variables.

-f *file* Queries resource variables specified in *file*.

-h Displays a usage statement only.

Operands

class Specifies the name of the resource variable class or a null string.

var Specifies the name of the resource variable or a null string.

rsrcID Specifies a resource ID or an asterisk.

Description

The **haemqvar** command queries the Event Management subsystem for information about resource variables. By default, the command writes to standard output the definitions for all resource variables in the current SP domain, that is, the current SP system partition as defined by the **SP_NAME** environment variable. If **SP_NAME** is not set the default system partition is used. The **-S** flag can be used to specify another SP domain (system partition). To query variables in an HACMP domain, use the **-H** flag. For an SP domain, the domain flag argument is a system partition name. For an HACMP domain, the domain flag argument is an HACMP cluster name. When the **-H** flag is specified, the command must be executed on one of the nodes in the HACMP/ES cluster.

The following information is reported for each resource variable definition:

- Variable Name
- Value Type
- Data Type
- SBS Format (if data type is Structured Byte String)
- Initial Value
- Class
- Locator
- Variable Description
- Resource ID and its description
- Default Expression (if defined) and its description

Since the default behavior of this command can produce a large amount of output, standard output should be redirected to a file.

If the **-d** flag is specified only the resource variable name and a short description are written to standard output, one name and description per line.

If the **-c** flag is specified the current values of all resource variables instances are written to standard output, one per line. The line of output contains the location of the resource variable instance (node number), the resource variable name, the resource ID of the instance and the resource variable instance value. If the resource variable is a Structured Byte String (SBS) data type, then the value of each SBS field is reported.

The **-i** flag reports the same information as the **-c** flag except that the value of the variable instance is the last known value rather than the current value. The **-i** flag is useful for determining what resource variable instances exist.

For both the **-c** and the **-i** flags, if an error is encountered in obtaining information about a resource variable instance, the output line contains an error message, symbolic error codes, the location of where the error originated (if it can be determined), the resource variable name and the resource ID.

To return information about specific resource variables, specify the *class, var* and *rsrcID* operands. These operands can be repeated to specify additional resource variables. In addition, the *var* and *rsrcID* operands can be wildcarded to match a number of resource variables. Note that null string operands or an asterisk must be quoted in the shells.

If *class* is not a null string, then all variables in the specified class, as further limited by the *var* and *rsrcID* arguments, are targets of the query. If class is a null string, then variables of all classes, as further limited by the *var* and *rsrcID* arguments, are targets of the query.

The *var* argument can be wildcarded in one of two ways:

- Specify the variable name as a null string
- Truncate the name after any component

When the resource variable name is wildcarded in the first manner, then all resource variables, as further limited by the *class* and *rsrcID* arguments, are targets of the query. When the resource variable name is wildcarded in the second manner, all resource variables whose high-order (leftmost) components match the *var* argument, as further limited by the *class* and *rsrcID* arguments, are targets of the query.

All resource variable instances, or definitions if neither the **-c** nor the **-i f**lags are specified, of the variables specified by the *class* and *var* arguments that match the *rsrcID* argument are the targets of the query.

If neither the **-c** nor the **-i** flags are specified, the *rsrcID* argument is a semicolon-separated list of resource ID element names. If either the **-c** or the **-i** flags is specified, the rsrcID argument is a semicolon-separated list of name/value pairs. A name/value pair consists of a resource ID element name followed by an equal sign followed by a value of the resource ID element. An element value may consist of a single value, a range of values, a comma-separated list of single values or a comma-separated list of ranges. A range takes the form a-b and is valid only for resource ID elements of type integer (the type information can be obtained from the variable definition). There can be no blanks in the resource ID.

A resource ID element is wildcarded by specifying its value as the asterisk character. Only variables that are defined to contain the elements, and only the elements, specified in the rsrcID argument are targets of the query. If any element of the resource ID consists of the asterisk character, rather than a name/value pair (or just a name if querying for definitions), all variables that are defined to contain at least the remaining specified elements are targets of the query. The entire resource ID is wildcarded if it consists of only the asterisk character; all instances of all resource variables, as further limited by the class and var arguments, are targets of the query.

Note that the *rsrcID* argument must be quoted in the shells if it contains semicolons or asterisks.

The *class, var* and *rsrcID* operands can be placed in a file, one set of operands per line, instead of being specified as command arguments. Use the **-f** flag to specify the name of the file to the command. If the **-f** flag is used, any operands to the command are ignored. Within the file, null strings are specified as two adjacent double quote characters and a completely wildcarded resource ID can either be a single asterisk or a double quoted asterisk (“*”). On each line the arguments must be separated by white space (blanks or tabs).

```
Following are some examples of using wildcards in the rsrcID argument:
NodeNum=5;VG=rootvg;LV=hd4
NodeNum=*;VG=rootvg;LV=hd4
NodeNum=*;VG=*;LV=*
NodeNum=9
NodeNum=*
NodeNum=9;VG=*;*
NodeNum=*;*
```

For these examples, assume the *class* and *var* arguments are null strings. If either the *class* or *var* arguments or both are not null strings, targets for the query are restricted accordingly.

In the first three examples, all variables whose resource IDs are defined to contain the elements *NodeNum, VG* and *LV*, and only those elements, are matched. In the first example, only one instance is matched. In the second example, one instance from each node is matched. In the third example, all instances of the matching resource variables are matched.

In the fourth example, all variables whose resource IDs are defined to contain only the element *NodeNum* are matched. The instances matched are associated with node 9. In the fifth example, the same set of variables are matched, but all instances of each variable are matched.

In the sixth example, all variables whose resource IDs are defined to contain elements *NodeNum* and *VG,* as well as zero or more additional elements, are matched. The instances matched are associated with node 9. In the last example, all variables whose resource IDs are defined to contain the element *NodeNum*, as well as zero or more additional elements, are matched. All instances of the variables are matched.

Given the flexibility in specifying resource variables for query, it is possible that no resource variable instance or resource variable definition will match. If there is no match appropriate error information is reported, either in the form described above or as follows.

If the specification of the *class*, *var* or *rsrcID* arguments are in error, the output line contains an error message, symbolic error codes and the specified class name, resource variable name and resource ID.

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM Highly Available Cluster Multi-Processing: Enhanced Scalability (HACMP)/ES) Licensed Program Product (LPP).

Location

/usr/sbin/rsct/bin/haemqvar

Examples

1. To obtain the definitions of all resource variables in the current cluster and place the output in a file, enter:

   ```
   haemqvar -H HAcluster > vardefs.out
   ```

2. To obtain a short form list of all resource variables whose resource IDs contain the element VG, in the HACMP cluster named HAcluster, enter:

   ```
   haemqvar -H HAcluster -d "" "" "VG;*"
   ```

3. To obtain resource variables whose resource IDs contain only the elements VG and NodeNum, enter:

   ```
   haemqvar -H HAcluster -d "" "" "VG;NodeNum"
   ```

haemtrcoff

Turns tracing off for the Event Manager daemon.

Syntax

```
haemtrcoff -s subsys_name -a trace_list
```

FLAGS

-s **subsys_name** Specifies the name of the Event Management subsystem. On a node this is **emsvcs**. This argument must be specified.

-a **trace_list** Specifies a list of trace arguments. Each argument specifies the type of activity for which tracing is to be turned off. At least one argument must be specified. If more than one argument is specified, the arguments must be separated by commas. The list may not include blanks.

OPERANDS The following trace arguments may be specified:

init Stops tracing the initialization of the Event Manager daemon.

config Stops dumping information from the configuration file.

insts Stops tracing resource variable instances that are handled by the daemon.

rmctrl Stops tracing Resource Monitor control.

cci Stops tracing the client communication (internal) interface.

emp Stops tracing the event manager protocol.

obsv Stops tracing resource variable observations.

evgn Stops tracing event generation and notification.

reg Stops tracing event registration and unregistration.

pci Stops tracing the peer communication (internal) interface.

msgs Stops tracing all messages that come to and are issued from the daemon.

query Stops tracing queries that are handled by the daemon.

gsi Stops tracing the Group Services (internal) interface.

eval Stops tracing expression evaluation.

rdi Stops tracing the reliable daemon (internal) interface.

sched Stops tracing the internal scheduler.

shm Stops tracing shared memory management activity.

all Stops tracing all activities.

all_but_msgs Stops tracing all activities except for messages. Message activity is defined by the msgs argument.

Description

The **haemtrcoff** command is used to turn tracing off for specified activities of the Event Manager daemon. Trace output is placed in an Event Management trace log for the system partition.

Use this command only under the direction of the IBM Support Center. It provides information for debugging purposes and may degrade the performance of the Event Management subsystem or anything else that is running in the system partition. Do **not** use this command during normal operation.

Files

/var/ha/log/em.trace.*cluster_name*

Contains the trace log of the **haemd** daemon on the cluster named *cluster_name*.

/var/ha/log/em.msgtrace.*cluster_name*

Contains message trace output from the Event Manager daemon on the cluster named *cluster_name*.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 31, The Event Management Subsystem

Location

/usr/sbin/rsct/bin/haemtrcoff

Related Information

emsvcsctrl, haemd, haemtrcon

Examples

1. To turn off all tracing for the Event Management subsystem on one of the cluster nodes, login to the node and enter:

   ```
   haemtrcoff -s emsvcs -a all
   ```

2. To turn off all tracing of initialization and configuration for the Event Management subsystem on a cluster node, login to the node and enter:

   ```
   haemtrcoff -s emsvcs -a init,config
   ```

haemtrcon

Turns tracing on for the Event Manager daemon.

Syntax

```
haemtrcon -s subsys_name -a trace_list
```

Flags

-s **subsys_name** Specifies the name of the Event Management subsystem. On a node, this is **emsvcs**. This argument must be specified.

-a **trace_list** Specifies a list of trace arguments. Each argument specifies the type of activity for which tracing is to be turned on. At least one argument must be specified. If more than one argument is specified, the arguments must be separated by commas. The list may not include blanks.

OPERANDS. The following trace arguments may be specified:

init Traces the initialization of the Event Manager daemon.

config Dumps information from the configuration file.

insts Traces resource variable instances that are handled by the daemon.

rmctrl Traces Resource Monitor control.

cci Traces the client communication (internal) interface.

emp Traces the event manager protocol.

obsv Traces resource variable observations.

evgn Traces event generation and notification.

reg Traces event registration and unr

pci Traces the peer communication (internal) interface.

msgs Traces all messages that come to and are issued from the daemon.

query Traces queries that are handled by the daemon.

gsi Traces the Group Services (internal) interface.

eval Traces expression evaluation.

rdi Traces the reliable daemon (internal) interface.

sched Traces the internal scheduler.

shm Traces shared memory management activity.

all Traces all activities.

all_but_msgs Traces all activities except for messages. Message activity is defined by the msgs argument.

regs Traces currently registered events.

dinsts Traces all resource variable instances known to the daemon.

iolists Traces immediate observation lists

olists Traces observation lists

Description

The **haemtrcon** command is used to turn tracing on for specified activities of the Event Manager daemon. Trace output is placed in an Event Management trace log for the system partition. When used, the regs, dinsts, iolists, and olists arguments perform a one-time trace. The specified information is placed in the trace log, but no further tracing is done.

Use this command only under the direction of the IBM Support Center. It provides information for debugging purposes and may degrade the performance of the Event Management subsystem or anything else that is running in the system partition. Do **not** use this command to turn tracing on during normal operation.

Files

/var/ha/log/em.trace.*cluster_name*

Contains the trace log of the **haemd** daemon on the cluster named *cluster_name.*

/var/ha/log/em.msgtrace.c*luster_name*

Contains message trace output from the Event Manager daemon on the cluster named *cluster_name*.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 31, The Event Management Subsystem

Location

/usr/sbin/rsct/bin/haemtrcon

Related Information

emsvcsctrl, haemd, haemtrcoff

Examples

1. To turn on all tracing for the Event Management subsystem on one of the cluster nodes, login to the node and enter:

   ```
   haemtrcon -s emsvcs -a all
   ```

2. To turn on all tracing of initialization and configuration for the Event Management subsystem on a cluster node, login to the node and enter:

   ```
   haemtrcon -s emsvcs-a init,config
   ```

haemunlkrm

Unlocks and starts a Resource Monitor.

Syntax

```
haemunlkrm -s subsys_name -a resmon_name
```

-s **subsys_name** Specifies the name of the Event Management subsystem. On a node this is **emsvcs**. This argument must be specified.

-a **resmon_name** Specifies the name of the Resource Monitor to unlock and start.

Description

If the Event Management daemon cannot successfully start a resource monitor after three attempts within a two hour interval, or if the daemon has successfully connected to the instances of a resource monitor N times within a two hour interval, the resource monitor is “locked” and no further attempts are made to start it or to connect to any of its instances. N is three in an HACMP/ES cluster. Once the cause of the failure is determined and the problem corrected, the haemunlkrm command can be used to unlock the Resource Monitor and attempt to start it or connect to the resource monitor instances.

The status of the Event Manager daemon, as displayed by the **lssrc** command, indicates if a Resource Monitor is locked.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Location

/usr/sbin/rsct/bin/haemunlkrm

Examples

If the output of the **lssrc** command indicates that the program Resource Monitor IBM.PSSP.harmpd is locked, then after correcting the condition that prevented the Resource Monitor from being started, enter:

```
haemunlkrm -s emsvcs -a IBM.PSSP.harmpd
```

Note: This example applies to unlocking a Resource Monitor on a node.

hagsd Daemon

A Group Services daemon that provides a general purpose facility for coordinating and monitoring changes to the state of an application that is running on a set of nodes.

Syntax

```
hagsd daemon_name
```

daemon_name specifies the name used by the daemon to name log files and identify its messages in the error log.

Description

The **hagsd** daemon is part of the Group Services subsystem, which provides a general purpose facility for coordinating and monitoring changes to the state of an application that is running on the cluster nodes. This daemon provides most of the services of the subsystem.

One instance of the **hagsd** daemon executes on each cluster node. The **hagsd** daemon is under System Resource Controller (SRC) control.

Because the daemon is under SRC control, it is better not to start it directly from the command line. It is normally called by the **grpsvcsctrl** command, which is in turn called by the HACMP/ES cluster start up process. If you must start or stop the daemon directly, use the **startsrc** or **stopsrc** command.

For more information about the Group Services daemons, see the **grpsvcsctrl** man page.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 30, The Group Services Subsystem
- *AIX Version 4 Commands Reference.*

Location

/usr/sbin/rsct/bin/hagsd

Related Information

grpsvcsctrl, hagsglsmd

Examples

See the **grpsvcsctrl** command.

hagsglsmd Daemon

A Group Services daemon that provides global synchronization services for the switch adapter membership group.

Syntax

```
hagsglsmd daemon_name
```

daemon_name	Specifies the name used by the daemon to name log files and identify its messages in the error log.

Description

The **hagsglsmd** daemon is part of the Group Services subsystem, which provides a general purpose facility for coordinating and monitoring changes to the state of an application that is running on cluster nodes.

One instance of the **hagsglsmd** daemon executes on every cluster node. The **hagsglsmd** daemon is under System Resource Controller (SRC) control.

Because the daemon is under SRC control, it is better not to start it directly from the command line. It is normally called by the **grpsvcsctrl** command, which is in turn called by the HACMP/ES cluster start up process. If you must start or stop the daemon directly, use the **startsrc** or **stopsrc** command.

For more information about the Group Services daemons, see the **grpsvcsctrl** man page.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 30, The Group Services Subsystem
- *AIX Version 4 Commands Reference.*

Location

/usr/sbin/rsct/bin/hagsglsmd

Related Information

grpsvcsctrl, hagsd

Examples

See the **grpsvcsctrl** command.

topsvcs Script

Starts or restarts Topology Services on a cluster node.

Syntax

```
topsvcs
```

Description

Use this command to start the operation of Topology Services for a cluster.

The **topsvcs** script is not normally executed from the command line. It is normally called by the **topsvcsctrl** command, which is in turn called by the HACMP/ES startup process.

Note that the **topsvcs** script issues the following commands:

```
no -o nonlocsrcroute=1
no -o ipsrcroutesend=1
no -o ipsrcrouterecv =1
no -o ipsrcrouteforward=1
```

These commands enable IP source routing. Do **not** change this setting, because the Topology Services subsystem requires this setting to work properly. If you change the setting, the Topology Services subsystem and a number of other subsystems that depend on it will no longer operate properly.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 32, The Topology Services Subsystem
- *AIX Version 4 Commands Reference.*
- Information about the System Resource Controller (SRC) in *AIX Version 4 General Programming Concepts: Writing and Debugging Programs*

Location

/usr/sbin/rsct/topsvcs

Related Information

topsvcsctrl, lssrc, startsrc, stopsrc

Examples

See the **topsvcsctrl** command.

topsvcsctrl Script

A control script that starts the Topology Services subsystem.

Syntax

```
topsvcsctrl {-a | -s | -k | -d | -c | -u | -t | -o | -r | -h}
```

-a	Adds the subsystem.
-s	Starts the subsystem.
-k	Stops the subsystem.
-d	Deletes the subsystem.
-c	Cleans the subsystems.
-t	Turns tracing on for the subsystem.
-o	Turns tracing off for the subsystem.
-r	Refreshes the subsystem.
-h	Displays usage information.

This command has no operands.

Description

Topology Services is a distributed subsystem of RSCT that provides information to other subsystems about the state of the nodes and adapters in the HACMP/ES cluster.

The **topsvcsctrl** control script controls the operation of the Topology Services subsystem. The subsystem is under the control of the System Resource Controller (SRC) and belongs to a subsystem group called **topsvcs**. This script is normally started by the HACMP/ES startup process.

An instance of the Topology Services subsystem executes every node of a cluster.

From an operational point of view, the Topology Services subsystem group is organized as follows:

Subsystem Topology Services

Subsystem Group **topsvcs**

SRC Subsystem **topsvcs** The topsvcs subsystem is associated with the **hatsd** daemon and the **topsvcs s**cript. The **topsvcs** script configures and starts the **hatsd** daemon. The subsystem name on the nodes is **topsvcs**.

There is one of each subsystem per node and it is associated with the cluster to which the node belongs.

Daemons **hatsd** The **hatsd** daemon provides the Topology Services. The **topsvcs** script configures and starts the **hatsd** daemon.

Description

The **topsvcsctrl** script is not normally executed from the command line. It is normally called by the HACMP/ES startup command.

The **topsvcsctrl** script provides a variety of controls for operating the Topology Services subsystem:

- Adding, starting, stopping, and deleting the subsystem
- Cleaning up the subsystems, that is, deleting them from all system partitions
- Turning tracing on and off
- Refreshing the subsystem

Before performing any of these functions, the script obtains the current cluster name (using the **cllsclstr** command) and the node number (using the **clhandle** command). If the node number is zero, the control script is running on the control workstation.

Except for the clean and unconfigure functions, all functions are performed within the scope of the current system partition.

Adding the Subsystem

When the **-a** flag is specified, the control script uses the **mkssys** command to add the Topology Services subsystem to the SRC. The control script operates as follows:

1. It makes sure that the **topsvcs** subsystem is stopped.
2. It gets the port number for the topsvcs subsystem for this cluster system from the Global ODM and ensures that the port number is set in the **/etc/services** file. The range of valid port numbers is 10000 to 10100, inclusive.
3. The service name that is entered in the **/etc/services** file is **topsvcs**.*cluster_name*.

4. It checks to see if the subsystem is already configured. If not, it creates an instance of the TS_Config class for this subsystem with default values. The default values are:
   - Heartbeats are sent out a rate of 1 per second (Frequency attribute = 1)
   - The number of heartbeats from the neighboring node that can be missed before the neighbor is declared inoperative is 4 (Sensitivity attribute = 4)
   - The execution priority is fixed (Run_FixPri attribute = 1)
   - The value of the execution priority used on the **set_priority** system call is 38 (FixPri_Value attribute = 38).
5. It removes the **topsvcs** subsystem from the SRC (just in case it is still there).
6. It adds the **topsvcs** subsystem to the SRC.

Starting the Subsystem

When the **-s** flag is specified, the control script uses the **startsrc** command to start the Topology Services subsystem, **topsvcs**.

Stopping the Subsystem

When the **-k** flag is specified, the control script uses the **stopsrc** command to stop the Topology Services subsystem, **topsvcs**.

Deleting the Subsystem

1. When the **-d** flag is specified, the control script uses the **rmssys** command to remove the Topology Services subsystem from the SRC. The control script operates as follows:
2. It makes sure that the **topsvcs** subsystem is stopped.
3. It removes the **topsvcs** subsystem from the SRC using the **rmssys** command.
4. It removes the port number from the **/etc/services** file.

Cleaning Up the Subsystems

When the **-c** flag is specified, the control script stops and removes the Topology Services subsystems for all clusters from the SRC. The control script operates as follows:

1. It stops all instances of subsystems in the clusters, using the **stopsrc -g topsvcs** command.
2. It removes all instances of subsystems in the subsystem group in all clusters from the SRC using the **rmssys** command.
3. It removes all entries for the **topsvcs** subsystems from the **/etc/services** file.

Turning Tracing On

When the **-t** flag is specified, the control script turns tracing on for the **hatsd** daemon, using the **traceson** command.

Turning Tracing Off

When the **-o** flag is specified, the control script turns tracing off (returns it to its default level) for the **hatsd** daemon, using the **tracesoff** command.

Refreshing the Subsystem

When the **-r** flag is specified, the control script refreshes the subsystem, using the **topsvcs** refresh command and the **refresh** command.

It rebuilds the information about the node and adapter configuration in the Global ODM and signals the daemon to read the rebuilt information.

Logging

While it is running, the Topology Services daemon provides information about its operation and errors by writing entries in a log file. The **hatsd** daemon uses a log file called **/var/ha/log/topsvcs**.*cluster_name*.

Files

/var/ha/log/topsvcs.*cluster_name*.	Contains the log of the **hatsd** daemon on the cluster named *cluster_name*.

Errors

This command writes error messages (as necessary) to standard error.

0	Indicates the successful completion of the command.
1	Indicates that an error occurred.

You must be running with an effective user ID of root.

Implementation Specifics

This command is part of RS/6000 Cluster Technology (RSCT), which is included with the IBM High Availability Cluster Multi-Processing for AIX: Enhanced Scalability (HACMP/ES) Licensed Program Product (LPP).

Prerequisite Information

- Chapter 32, The Topology Services Subsystem
- *AIX Version 4 Commands Reference.*
- Information about the System Resource Controller (SRC) in *AIX Version 4 General Programming Concepts: Writing and Debugging Programs*

Location

/usr/sbin/rsct/bin/topsvcsctrl

Related Information

topsvcs, lssrc, startsrc, stopsrc

Examples

1. To add the Topology Services subsystem to the SRC, enter:

   ```
   topsvcsctrl -a
   ```

2. To start the Topology Services subsystem, enter:

   ```
   topsvcsctrl -s
   ```

3. To stop the Topology Services subsystem, enter:

   ```
   topsvcsctrl -k
   ```

4. To delete the Topology Services subsystem from the SRC, enter:

   ```
   topsvcsctrl -d
   ```

5. To clean up the Topology Services subsystem, enter:

   ```
   topsvcsctrl -c
   ```

6. To turn tracing on for the Topology Services daemon, enter:

   ```
   topsvcsctrl -t
   ```

7. To turn tracing off for the Topology Services daemon, enter:

   ```
   topsvcsctrl -o
   ```

8. To display the status of all of the subsystems in the Topology Services SRC group, enter:

   ```
   lssrc -g topsvcs
   ```

9. To display the status of an individual Topology Services subsystem, enter:

   ```
   lssrc -s subsystem_name
   ```

10. To display detailed status about an individual Topology Services subsystem, enter:

    ```
    lssrc -l -s subsystem_name
    ```

    In response, the system returns information that includes the running status of the subsystem, the number of defined and active nodes, the required number of active nodes for a quorum, the status of the group of nodes, and the IP addresses of the source node, the group leader, and the control workstation.

11. To display the status of all of the daemons under SRC control, enter:

    ```
    lssrc -a
    ```

Appendix G RSCT Messages

This appendix describes the RSCT error messages.

2520 – Group Services Messages

2520-201 **Cannot remove service entry service_name from services_file.**

EXPLANATION

Either the service name service_name was not found in the services-file, or the service name service_name could not be removed from the services file. The **/tmp** directory or the file system containing the services file service_file could be full.

USER RESPONSE

Add the service name before deleting it, or ensure that the file systems are not full.

2520-206 **Cannot register service service_name, protocol protocol, port port_number.**

EXPLANATION

The subsystem being installed by the named command was not able to install the specified service name service_name protocol protocol port number port_number in the /etc/services directory, which is required for the subsystem to operate.

USER RESPONSE

If there is a port number specified in the error message, verify that this port number is not being used by another subsystem. If no port number is specified in the error message, all of the port numbers that this command selected were already in use.

2520-207 **daemon_file_name is not executable.**

EXPLANATION

The daemon file name daemon_file_name must be an executable file to allow the subsystem being installed by the named command to operate, and the daemon file name daemon_file_name may also not exist.

USER RESPONSE

As root, run chmod +x daemon_file_name and try the command again, or run installp again to install the subsystem.

2520-208 **The subsystem_name subsystem must be stopped.**

EXPLANATION

The subsystem name subsystem_name must be stopped to allow the named command to operate on it.

USER RESPONSE

As root, run stopsrc -s subsystem_name and try the command again.

2520-209 **Could not add subsystem_name to resource_controller.**

EXPLANATION

The named command failed in defining the subsystem name subsystem_name to the resource controller resource_controller. If the resource controller is inittab, mkitab subsystem name failed. If resource controller is SRC, **mkssys -s *subsystem name*** failed.

USER RESPONSE

Fix what caused the resource controller command listed above to fail and try the command again.

2520-211 **Internal error.**

EXPLANATION

The named command encountered an internal error.

USER RESPONSE

Contact the IBM Support Center.

2521 – Event Management Messages

2521-000 **The trace argument trace_argument is not valid.**

EXPLANATION

A trace argument that is not valid was specified to the named program.

USER RESPONSE

Specify a valid trace argument. See the documentation for the named program.

2521-001 **The option_flag flag requires an argument.**

EXPLANATION

The named option flag option_flag did not have a required option argument.

USER RESPONSE

Specify a valid option argument. See the documentation for the named program.

2521-002 **Undefined option flags specified.**

EXPLANATION

Undefined option flags were specified to the named program.

USER RESPONSE

See the documentation for the named program for valid option flags.

2521-003 **Cannot obtain node number.**

EXPLANATION

The named program cannot obtain the number of the node upon which it is processing.

USER RESPONSE

Validate that the node number has been set and that **/usr/sbin/cluster/utilities/clhandle** exists and is executable.

2521-005 **Reserved error number.**

EXPLANATION

The named program encountered an internal error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-006 **System call system_call failed with error error_number - error_message.**

EXPLANATION

The named system call system_call failed with the specified error number error_number and message error_message when invoked by the named program.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-007 **Internal error (additional_error_information).**

EXPLANATION

The named program encountered an internal error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-008 **Internal error (error_info1; error_info2).**

EXPLANATION

The named program encountered an internal error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-009 **Cannot make directory directory_name, mkdir() error is error_number - error_message.**

EXPLANATION

The named program attempted to create the specified directory directory_name using the **mkdir()** system call. The system call failed with the named error error_number.

USER RESPONSE

If possible, correct the condition that resulted in the error, such as missing directories in the path of the directory to be created, and start the Event Management daemon again. Refer to the Event Management Subsystem chapter for information. If the problem cannot be corrected contact the IBM Support Center.

2521-010 **Cannot open or create file_name, open() error is error_number - error_message.**

EXPLANATION

The named program attempted to open or create the specified file file_name using the **open()** system call. The system call failed with the named error error_number.

USER RESPONSE

If possible, correct the condition that resulted in the error and start the Event Management daemon again (if it has ended). Refer to the Event Management Subsystem chapter for information regarding the specified file. If the problem cannot be corrected, record the above information and contact the IBM Support Center.

2521-015 **Event Manager configuration database version mismatch (expected_version_string; data_base_version_string).**

EXPLANATION

The version of the Event Manager Configuration Database (EMCDB) read by the Event Manager daemon is not the expected version.

USER RESPONSE

Refer to the Event Management Subsystem chapter.

2521-016 **Event Manager configuration database checksum error.**

EXPLANATION

The version of the Event Manager Configuration Database (EMCDB) read by the Event Manager daemon produced a checksum error.

USER RESPONSE

The EMCDB has become corrupted. Remove the file **/etc/ha/cfg/em.*domain_name.cdb*** on the node where this error occurred, where *domain_name* is the name of the domain of the node. After the file is removed restart the Event Management daemon. Refer to the Event Management subsystem chapter for more information about the EMCDB.

2521-017 **Cannot connect to resource monitor resource monitor name (resource monitor instance number) error error number - error message.**

EXPLANATION

The Event Manager daemon could not connect to the specified instance of the named resource monitor.

USER RESPONSE

Record the above information. Contact the IBM Support Center if the named resource monitor is supplied by IBM. Otherwise, contact the vendor that supplied the resource monitor. Also, examine the error log for any error messages from the named resource monitor.

2521-018 **Resource monitor resource_monitor_name terminated due to signal signal_number.**

EXPLANATION

The named resource monitor resource_monitor_name terminated as a result of receiving the specified signal signal_number.

USER RESPONSE

Record the above information. Contact the IBM Support Center if the named resource monitor is supplied by IBM. Otherwise, contact the vendor that supplied the resource monitor. Also, examine the error log for any error messages from the named resource monitor.

2521-019 **Resource monitor resource monitor name (resource monitor ID) terminated with exit code exit code.**

EXPLANATION

The named resource monitor terminated with the specified exit code.

USER RESPONSE

If the exit code is 127, the resource monitor could not be executed. Examine the file **/var/ha/log/em.default.*domain_name*.RMid** for further information, where domain_name is the name of the Event Management domain (identical to the domain name in the name of the file containing this error) and RMid is the resource monitor ID specified in this error message. If the exit code is not 127, then contact the IBM Support Center if the named resource monitor is supplied by IBM. Otherwise, contact the vendor that supplied the resource monitor. Also, examine the error log for any error messages from the named resource monitor.

2521-020 **Command based resource monitor terminated early.**

EXPLANATION

A command based resource monitor connected to the Event Manager daemon, but terminated without sending data to the daemon.

USER RESPONSE

Record the above information. Contact the IBM Support Center if the named resource monitor is supplied by IBM. Otherwise, contact the vendor that supplied the resource monitor.

2521-021 **Corrupted registration cache file (additional_error_information).**

EXPLANATION

A file used to cache event registration requests has become corrupted.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-022 **Cannot get port number for service_name.**

EXPLANATION

A port number for the specified service name service_name cannot be obtained from the /etc/services file.

USER RESPONSE

Refer to the Event Management Subsystem chapter.

2521-023 **Cannot initialize Group Services (error_code).**

EXPLANATION

Event Management could not initialize Group Services, due to the specified error code error_code returned from Group Services.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-024 **Received count_of_unrecognized_message unrecognized messages in last time minutes.**

EXPLANATION

The Event Management daemon has received the specified number of unrecognized messages count_of_unrecognized_message within the specified time interval. These messages were received on the UDP port used for communication among Event Management daemons. The most likely cause of this error is that this port number is being used by another application.

USER RESPONSE

Validate that the port number configured for use by the Event Management daemon is only being used by the Event Management daemon. Refer to the Event Management Subsystem chapter.

2521-026 **Cannot read file_name, read() error is error_number - error_message.**

EXPLANATION

The named program attempted to read the specified file file_name using the read() system call. The system call failed with the named error error_number. An error number of 0 indicates that the read() call ended with a short count. For example, the named file file_name did not contain as much data as was expected.

USER RESPONSE

If possible, correct the condition that resulted in the error and start the Event Management daemon again (if it has ended). Refer to the Event Management Subsystem chapter for information regarding the specified file. If the problem cannot be corrected, record the above information and contact the IBM Support Center.

2521-027 **Socket connection type socket_type rejected, too many open descriptors.**

EXPLANATION

Event Management rejected a socket connection of the indicated type due to insufficient available file descriptors:

- Type 'C': connection from client
- Type 'R': connection from/to resource monitor
- Type 'H': connection to Host Respond daemon.

USER RESPONSE

End unnecessary Event Management client applications. Refer to the Event Management Subsystem chapter.

2521-028 **Reserved error number.**

EXPLANATION

The named program encountered an internal error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-029 **Reserved error number.**

EXPLANATION

The named program encountered an internal error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-030 **Resource monitor resource monitor name (resource monitor instance number) added unsolicited instance of variable resource variable name.**

EXPLANATION

The specified instance of the named resource monitor attempted to add an instance of the named resource variable that had not been requested by the Event Management session. This indicates a programming error in the resource monitor.

USER RESPONSE

Event Management clients should refrain from referencing this resource variable until the resource monitor has been corrected. Record the above information. Contact the IBM Support Center if the named resource monitor is supplied by IBM. Otherwise, contact the vendor that supplied the resource monitor.

2521-031 **Resource monitor resource monitor name (resource monitor instance number) deleted unsolicited instance of variable resource variable name.**

EXPLANATION

The specified instance of the named resource monitor attempted to delete an instance of the named resource variable that had not been requested by the Event Management session. This indicates a programming error in the resource monitor.

USER RESPONSE

Event Management clients should refrain from referencing this resource variable until the resource monitor has been corrected. Record the above information. Contact the IBM Support Center if the named resource monitor is supplied by IBM. Otherwise, contact the vendor that supplied the resource monitor.

2521-032 **Cannot dispatch group services (group_services_return_code).**

EXPLANATION

The Event Management daemon could not dispatch the Group Services function: the **ha_gs_dispatch()** function returned a nonzero value.

USER RESPONSE

The Event Management daemon should have ended and then restarted. Check the Group Services Subsystem for errors.

2521-033 **Peer daemon on node node number is not responding to Group Services.**

EXPLANATION

The Event Management daemon on the specified node has not responded to a Group Services “ping” within 120 seconds.

USER RESPONSE

The Event Management daemon on the specified node should eventually terminate and then be restarted. If it does not then the administrator must terminate the unresponsive daemon (using the **kill** command). Until the daemon on the specified node is restarted, no other Event Management daemons can join the Event Management peer group and provide Event Management services.

2521-034 **Not responding to Group Services - terminating.**

EXPLANATION

The Event Management daemon has been informed that it has not responded to a Group Services “ping” within 120 seconds. The daemon is terminating so that it can be restarted automatically.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-035 **The daemon is executing with user ID incorrect user ID instead of root.**

EXPLANATION

The Event Management daemon is executing with the specified user ID instead of with the root ID.

USER RESPONSE

The error could be caused by a user other than root executing the Event Management daemon **haemd** or by an incorrect configuration of the Event Management subsystem in the System Resource Controller (SRC). Validate that the file **/usr/sbin/rsct/bin/haemd** is not readable or executable by 'others'.

To correct this problem, the daemon can be reconfigured in the SRC by executing the **emsvcsctrl -d** command, then executing the **emsvcsctrl -a** command and, finally, executing the **emsvcsctrl -s** command. Refer to the Event Management Subsystem chapter and the man page for the **emsvcsctrl** command for more information.

2521-036 **The Event Management group 'haemrm' cannot be found.**

EXPLANATION

The Event Management daemon could not find the group name **haemrm**. This group must exist in order for Resource Monitors other than root to properly connect to the Event Management daemon.

USER RESPONSE

Correct this problem by executing the **emsvcsctrl -c** command, then executing the **emsvcsctrl -a** command and, finally, executing the **emsvcsctrl -s** command.

Refer to the Event Management Subsystem chapter and the man page for the emsvcsctrl command for more information.

2521-037 **Internal error (XXXXXXXX ; XXXXXXXX ; XXXXXXXX; XXXXXXXX ; XXXXXXXX).**

EXPLANATION

The named program encountered an internal error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-038 **Resource monitor resource monitor name (resource monitor instance number) added duplicate instance of variable resource variable name.**

EXPLANATION

The specified instance of the named resource monitor attempted to register an instance of the named resource variable that has already been registered by another instance of the named resource monitor. This indicates a programming error in the resource monitor.

USER RESPONSE

Event Management clients should refrain from referencing this resource variable until the resource monitor has been corrected. Contact the IBM Support Center if the named resource monitor is supplied by IBM. Otherwise, contact the vendor that supplied the resource monitor.

2521-039 **Shared memory (shared memory ID) has been corrupted, dump file is dump file name.**

EXPLANATION

The Event Management daemon has detected a corrupted shared memory segment with the specified shared memory ID. A copy of the first page of the segment has been placed in the named dump file in the**/var/ha/run/haem.*domain_name*** directory, where *domain_name* is the name of the Event Management domain. The dump file name includes the name and instance number of the resource monitor instance that was using the shared memory segment, and also includes a time stamp value.

USER RESPONSE

The most probable cause of this error is a defect in the named resource monitor. Record the above information. Contact the IBM Support Center if the named resource monitor is supplied by IBM. Otherwise, contact the vendor that supplied the resource monitor. A less probable cause is that some other application has attached to the specified shared memory segment and corrupted it.

2521-040 Cannot execute cmd name to obtain node number or domain name (error code).

EXPLANATION

The named program cannot execute the specified command. The **clhandle** command obtains the number of the node upon which the named program is executing. The **cldomain** command obtains the name of the domain containing the node upon which the named program is executing.

USER RESPONSE

Validate that the specified command exists and is executable.

2521-041 **command name exited with error code error code.**

EXPLANATION

The named program executed the specified command, which then returned the specified error code.

USER RESPONSE

Refer to the HACMP product documentation for the specified command to determine the appropriate action for the returned error code.

2521-042 **cmd name had no output.**

EXPLANATION

The named program executed the specified command but the command did not provide any output. The **clnodenum** command returns the number of the node upon which the named program is executing. The cldomain command returns the name of the domain containing the node upon which the named program is executing.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-043 **cmd name terminated due to signal number signal number.**

EXPLANATION

The named program executed the specified command, which then terminated as a result of the specified signal number.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-044 **The system call name system call failed, see error 2521-045 in error file name.**

EXPLANATION

The indicated system call failed. Additional information can be found in the specified error file for Event Management.

USER RESPONSE

Record the above information and contact the IBM Support Center Provide this message and message 2521-045 from the default log file.

2521-045 **The system call name system call failed with error error number - error message when loading load module name. Additional error information: additional error information.**

EXPLANATION

The indicated system call failed when attempting to load the specified load module. Additional error information may follow.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-048 **Internal error: 0xXXXXXXXX 0xXXXXXXXX 0xXXXXXXXX s1=XXXXXXXX s2=XXXXXXXX s3=XXXXXXXX**

EXPLANATION

The named program encountered an internal error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-300 **The -option flag flag requires an argument.**

EXPLANATION

The named option flag did not have a required option argument.

USER RESPONSE

Specify a valid option argument. See the documentation for the named program.

2521-301 **Internal error (additional error information, additional error information).**

EXPLANATION

The named program encountered an internal error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-302 **The EMAPI function function name failed (error number) with the following error message: error message**

EXPLANATION

The named program invoked the specified function, which returned the specified error number and error message.

USER RESPONSE

If possible, perform the corrective action for the specified error message. If not, retry the command. If the error is repeated, record the above information and contact the IBM Support Center.

2521-303 **Cannot obtain information for domain domain name (domain type).**

EXPLANATION

The named program could not obtain information about the specified domain.

USER RESPONSE

The most likely cause of this error is that the specified domain name does not match any existing domains of the specified type. Provide a valid domain name and retry the command.

2521-304 **Cannot connect to the domain name Event Management subsystem.**

EXPLANATION

The named program could not establish a connection with the Event Management subsystem in the specified domain.

USER RESPONSE

The most likely cause of this error is that the Event Management subsystem in the specified domain is not running. Start the Event Management subsystem, if necessary, and retry the command.

2521-305 **Lost connection to the domain name Event Management subsystem.**

EXPLANATION

The named program lost its connection with the Event Management subsystem in the specified domain.

USER RESPONSE

The most likely cause of this error is that the Event Management subsystem in the specified domain has stopped running. Start the Event Management subsystem, if necessary, and retry the command.

2521-306 **Incorrect number of command line arguments (number of arguments).**

EXPLANATION

The named program requires that the number of command line arguments be a multiple of three.

USER RESPONSE

Specify either no command line arguments or three arguments for each resource variable to be queried: the class name, resource variable name and the resource ID. Refer to the manual page for the named program for more information.

2521-307 **Cannot open input file file name. fopen() error is error number - error message**

EXPLANATION

The named program could not open the specified input file.

USER RESPONSE

Retry the command after correcting the problem indicated by the fopen() error message.

2521-308 **Error in input file at line line number, position line position.**

EXPLANATION

The named program detected an error in the input file on the indicated line number at the indicated position.

USER RESPONSE

Refer to the manual page for the named program for more information.

2521-551 **daemon-file-name is not executable.**

EXPLANATION

daemon-file-name must be an executable file to allow the subsystem being installed by the above command to operate. The problem could also be that daemon-file-name does not exist.

USER RESPONSE

As root, run c**hmod +x** ***daemon-file-name***, and retry the command. If daemon-file-name does not exist, rerun **installp** of the RS/6000 Cluster Technology package to reinstall the subsystem.

2521-552 **Could not add subsystem-name to resource-controller.**

EXPLANATION

The above command failed to define the subsystem to the resource controller. If the resource controller is inittab, the command mkitab subsystem-name failed. If the resource controller is SRC, **mkssys -s** ***subsystem-name*** failed.

USER RESPONSE

Fix whatever caused the resource controller command listed above to fail, and retry the command.

2521-553 **Cannot create the Event Management group EM-group-name.**

EXPLANATION

The above command failed to create the specified Event Management group.

USER RESPONSE

The file /var/ha/log/em.mkgroup contains the error message from the **mkgroup** command. Correct the problem indicated by this message and rerun this command. Refer to the Event Management Subsystem chapter and the man page for this command for more information.

2521-554 **Cannot create the Event Management directory EM-directory-name.**

EXPLANATION

The above command failed to create the specified Event Management directory.

USER RESPONSE

The file **/var/ha/log/em.mkdir** contains the error message from the **mkdir** command. Correct the problem indicated by this message and rerun this command. Refer to the Event Management Subsystem chapter and the man page for this command for more information.

2521-555 **Cannot copy the Configuration Database.**

EXPLANATION

The above command failed to copy the Event Management Configuration Database.

USER RESPONSE

The file **/var/ha/log/em.cp** contains a **cp** command error message. Correct the problem indicated by this error message. Refer to the Event Management Subsystem chapter and the man page for this command for more information.

2521-601 **An error occurred in function name while getting domain information from command: standard error message.**

EXPLANATION

While the Event Manager API was preparing to run, attempted to run, or ran the specified command to obtain information about a domain, a call to the named function failed. Standard error message indicates the error returned by the function.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-602 **Unexpected end-of-file detected while getting domain information from command.**

EXPLANATION

The Event Manager API ran the command identified above to obtain information about a domain. The Event Manager API did not receive all the expected information from the command.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-603 **A data format error was detected while getting domain information from command.**

EXPLANATION

The Event Manager API ran the command identified above to obtain information about a domain. The Event Manager API received information from the command that was not of the expected format.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-614 **Attempt to allocate number_of_bytes bytes with a call to function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to allocate number_of_bytes bytes with the function_name function. The allocation was not successful. Note that standard_error_message indicates the error returned by function name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-615 **Conversion of domain address domain address to network form failed.**

EXPLANATION

The Event Manager API could not covert the above domain address to network form. The address must not be valid.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-616 **Conversion of domain port domain port to network form failed.**

EXPLANATION

The Event Manager API could not covert the above domain port to network form. The address must not be valid.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-617 **An attempt to create a UNIX domain socket with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to create a UNIX domain socket with the function_name function. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-618 **An attempt to create an Internet domain socket with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to create a Internet domain socket with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-619 **An attempt to get file descriptor flags with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to get file descriptor flags with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-620 **An attempt to set file descriptor flags with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to set file descriptor flags with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-621 **An attempt to initialize a lock with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to initialize a lock with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-622 **An attempt to lock a lock with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to lock a lock with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-623 **An attempt to unlock a lock with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to unlock a lock with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-624 **An attempt to initialize a mutex with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to initialize a mutex with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-625 **An attempt to lock a mutex with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to lock a mutex with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-626 **An attempt to unlock a mutex with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to unlock a mutex with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-627 **A session file descriptor was apparently closed without calling ha_em_end_session().**

EXPLANATION

A file descriptor was created for a new Event Manager API session, but the file descriptor was already associated with an existing Event Manager API session. The program may have closed the session file descriptor without calling ha_em_end_session().

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-628 **Attempt to close the session file descriptor with a call to function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to close a file descriptor with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-629 **The UNIX domain socket file name generated from file_name is too long.**

EXPLANATION

The name of the UNIX domain socket file used to communicate with the Event Manager daemon is too long to fit in an address structure.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-630 **Attempt to connect to an Event Manager daemon with a UNIX domain socket failed: function_name: standard_error_message**

EXPLANATION

The Event Manager API attempted to connect to an Event Manager daemon through a UNIX domain socket, but failed. The function_name function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-631 **Attempt to connect to the Event Manager daemon with an Internet domain socket failed: function_name: standard_error_message.**

EXPLANATION

The Event Manager API attempted to connect to an Event Manager daemon through an Internet domain socket, but failed. The function_name function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-632 **Attempt to get current time with function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API attempted to obtain the current time with the function_name function, but failed. The function returned standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-633 **Session with file descriptor file_descriptor not found.**

EXPLANATION

The Event Manager API could not find a session with file descriptor file_descriptor An Event Manager API routine must have been called specifying a session file descriptor that was not valid.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-635 **Undefined command specified: command.**

EXPLANATION

An undefined command value was specified on a call to **ha_em_send_command()**.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-636 **Undefined subcommand specified for event registration request: subcommand.**

EXPLANATION

An undefined subcommand value was specified on a call to **ha_em_send_command().** The undefined value was subcommand.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-637 **Unexpected number of elements specified for event registration request: number_of_elements.**

EXPLANATION

An unexpected number of elements was specified in a call to **ha_em_send_command().** The unexpected number of elements was number_of_elements.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-638 **Undefined subcommand specified for event unregistration request: subcommand.**

EXPLANATION

An undefined subcommand value was specified on a call to **ha_em_send_command()**. The undefined value was subcommand.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-639 **Unexpected number of elements specified for event unregistration request: number_of_elements**

EXPLANATION

An unexpected number of elements was specified in a call to **ha_em_send_command().** The unexpected number of elements was number_of_elements.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-640 **Undefined subcommand specified for query request: subcommand.**

EXPLANATION

An undefined subcommand value was specified on a call to **ha_em_send_command().** The undefined value was subcommand.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-641 **Unexpected number of elements specified for query request: number_of_elements.**

EXPLANATION

An unexpected number of elements was specified in a call to **ha_em_send_command().** The unexpected number of elements was number_of_elements.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-642 **The session's connection to the Event Manager daemon has been previously lost.**

EXPLANATION

An Event Manager session's connection has been lost, and an improper attempt was made to perform some action within the session. Once a session's connection has been lost, the only proper action within the session is to restart or end the session.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-643 **This process does not own the specified session.**

EXPLANATION

A process made an improper attempt to perform some action within an Event Manager session that was not started by the process. Once a process inherits an Event Manager session from its parent, the only proper action on that session within the child process is to end the session.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-644 **An event registration array element could not be allocated.**

EXPLANATION

Internally, the Event Manager API keeps track of event registration requests. The Event Manager API could not find a free array element to use to keep track of a new event registration request.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-645 **A query control array element could not be allocated.**

EXPLANATION

Internally, the Event Manager API keeps track of query requests. The Event Manager API could not find a free array element to use to keep track of a new query request.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-646 **Attempt to send command message unsuccessful; function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API failed in an attempt to send a command message to an Event Manager daemon. The failing function was function_name. The error message returned by the function was standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-647 **Attempt to send command message unsuccessful; function_name sent no bytes.**

EXPLANATION

The Event Manager API failed in an attempt to send a command message to an Event Manager daemon. The failing function was function_name. No bytes were sent.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-648 **Attempt to receive command response unsuccessful; function_name failed: standard_error_message.**

EXPLANATION

The Event Manager API failed in an attempt to receive a command response from an Event Manager daemon. The failing function was function_name. The error message returned by the function was standard_error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-649 **Attempt to receive command response unsuccessful; function_name detected end-of-file; connection with Event Manager lost.**

EXPLANATION

The Event Manager API failed in an attempt to receive a command response from an Event Manager daemon. The failing function was function_name, which reported that it detected end-of-file. The Event Manager API connection with the Event Manager daemon has been lost.

USER RESPONSE

Use the **ha_em_restart_session()** function to reconnect the session with the Event Manager daemon.

2521-650 **No resource variable name was specified for an event that was to be registered.**

EXPLANATION

The Event Manager API was called to register one or more events. One of the events did not specify a resource variable name.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-651 **No rearm expression was specified for an event that was to be registered.**

EXPLANATION

The Event Manager API was called to register one or more events with rearm expressions. One of the events did not specify a rearm expression.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-652 **Could not find environment_variable in the environment.**

EXPLANATION

The Event Manager was looking for a specific environment variable, environment_variable, but failed to find it. This message should only occur when using debugging versions of the Event Manager API.

USER RESPONSE

If you are intentionally using a debugging version of the Event Manager API, set the specified environment variable to a value. Otherwise, record the above information and contact the IBM Support Center.

2521-653 **Format of value found in environment_variable environment variable not valid; value was: value.**

EXPLANATION

The Event Manager obtained the value, value, of an environment variable, environment_variable, and its format is not valid. This message should only occur when using debugging versions of the Event Manager API.

USER RESPONSE

If you are intentionally using a debugging version of the Event Manager API, set the specified environment variable to a valid value. Otherwise, record the above information and contact the IBM Support Center.

2521-654 **Value found in environment_variable environment variable too long; value was: value.**

EXPLANATION

The Event Manager obtained the value, value of an environment variable, environment_variable, and its value is too long. This message should only occur when using debugging versions of the Event Manager API.

USER RESPONSE

If you are intentionally using a debugging version of the Event Manager API, set the specified environment variable to a valid value. Otherwise, record the above information and contact the IBM Support Center.

2521-656 **Unexpected event ID specified in event unregistration request: event_ID.**

EXPLANATION

The event ID event_ID was specified in an event unregistration request, but the event ID does not identify a registered event.

USER RESPONSE

Determine if the event ID is still valid.

2521-657 **Duplicate event ID specified in event unregistration request: event_ID.**

EXPLANATION

The event ID event_ID was specified more than once in an event unregistration request, or it was specified in a previous event unregistration request.

USER RESPONSE

Determine if the event ID has already been specified for unregistration.

2521-658 **Lost track of an event ID specified in an event unregistration request: event_ID.**

EXPLANATION

The event ID event_ID was specified in an event unregistration request, and the EMAPI has lost track of it.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-659 **Could not duplicate a file descriptor while getting domain information from command.**

EXPLANATION

The Event Manager API was preparing to run the command identified above, to obtain information about a domain. The Event Manager API could not associate the write end of a pipe with the standard output file descriptor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-660 **After getting domain information from command, a call to routine () returned an unexpected value: value.**

EXPLANATION

The Event Manager API had run the specified command to obtain information about a domain. The Event Manager API called the above routine to wait for the command to finish. The routine returned an unexpected value: value.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-661 **The command command returned unexpected status: 0xXXXXXXXX.**

EXPLANATION

The Event Manager API had run the specified command to obtain information about a domain. The command unexpectedly terminated with the above status.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-662 **The session has become unusable.**

EXPLANATION

An Event Manager session has become unusable. The session may have become unusable as a result of a previously reported error. If the thread safe version of the EMAPI is being used, the session may have become unusable because a thread using the session was canceled.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-663 **The session cannot be restarted, because the session's connection has not been lost.**

EXPLANATION

The ha_em_restart_session() routine was called for a session that has not lost its connection to the Event Manager daemon.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-664 **Attempt to send command message unsuccessful; function_name failed: Connection with Event Manager lost.**

EXPLANATION

The Event Manager API failed in an attempt to send a command message to an Event Manager daemon. The failing function was function_name. The send failed because the Event Manager API connection with the Event Manager daemon may have been lost.

USER RESPONSE

The **ha_em_restart_session()** function can be used to reconnect the session with the Event Manager daemon.

2521-665 **Undefined domain type specified: 0x domain type.**

EXPLANATION

An undefined domain type value was specified on a call to ha_em_start_session(). The undefined value was domain type.

USER RESPONSE

Specify a defined domain type. See the HA_EM_DOMAIN_* definitions in **ha_emapi.h**.

2521-666 **The command command terminated with the exit value exit value.**

EXPLANATION

The Event Manager API had run the specified command to obtain information about the HACMP domain. The exit value of the command indicates an error occurred.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-667 **The command command failed to return a domain name.**

EXPLANATION

The Event Manager API had run the specified command to obtain information about the HACMP domain. It did not return a domain name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-668 **The specified domain, specified domain, does not match the HACMP domain, actual domain, returned by command command.**

EXPLANATION

The specified domain above does not match the HACMP domain above.

USER RESPONSE

Specify the correct HACMP domain on the call to **ha_em_start_session()**, or specify a null string.

2521-801 **The RMAPI has already been initialized.**

EXPLANATION

The ha_rr_init() routine has already been called successfully.

USER RESPONSE

The ha_rr_init() should only be called once to initialize the RMAPI.

2521-802 **The system call system call failed while attempting to allocate number of bytes bytes: standard error message**

EXPLANATION

The system call failed when trying to allocate the number of bytes.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-803 **The system call system call failed while attempting to open file file name : errno -standard error message.**

EXPLANATION

An error occurred when trying to open the specified file.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-804 **Unable to obtain a lock in the file name. Attempted to lock a monitor instance id in the range lower -upper.**

EXPLANATION

An error occurred when the RMAPI tried to lock the specified file. The lock file is used to ensure that only the number of monitor copies that the monitor is configured to allow, may execute at the same time. The lower and upper bound of resource monitor ids that were attempted is provided.

USER RESPONSE

Make sure that only the number of copies the resource monitor is configured to allow are executing on the node.

2521-805 **Checksum failed on EMCDB file: name.**

EXPLANATION

The version of the EMCDB file read by the RMAPI produced a checksum error.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-806 **Resource monitor rname not found in EMCDB file: cname.**

EXPLANATION

The resource monitor rname was not found in the EMCDB file for this domain.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-807 **Spmi function function_name failed with error error_number.msg.**

EXPLANATION

The specified Spmi routine failed.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-808 **The RMAPI has not been initialized.**

EXPLANATION

The ha_rr_init() routine has not been called to initialize the RMAPI.

USER RESPONSE

The RMAPI must be initialized by calling the **ha_rr_init()** routine before calling any other RMAPI routines.

2521-809 **Bad variable vector argument.**

EXPLANATION

A ha_rr_variable argument to an RMAPI routine was NULL or the number of variables specified to the routine was 0.

USER RESPONSE

Check that both the vector and vector size arguments are being passed correctly.

2521-810 **The resource class class name was not found in the EMCDB for resource monitor monitor name.**

EXPLANATION

The requested class name was not found in the EMCDB for the monitor.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-812 **Bad variable value vector argument.**

EXPLANATION

An ha_rr_val argument to an RMAPI routine was NULL or the number of variables specified to the routine was 0.

USER RESPONSE

Check that the vector and vector size arguments are being passed correctly.

2521-813 **Resource monitor server socket does not exist.**

EXPLANATION

A request was made to start a session before the resource monitor server socket has been created. This message applies only to server type resource monitors.

USER RESPONSE

Check that monitor is defined properly and that ha_rr_makserv() is being called prior to calling **ha_rr_start_session()**.

2521-814 **Notification protocol argument does not match.**

EXPLANATION

The ha_rr_notify_proto argument has been passed differently between functions calls.

USER RESPONSE

Check that the ha_rr_notify_proto arguments to the **ha_rr_makserv()** routine and the **ha_rr_start_session()** routine are the same.

2521-815 **No connection pending to accept.**

EXPLANATION

A request was made to accept a connection on the resource monitor's server socket, but no connection request was found.

USER RESPONSE

This message is returned only to server type resource monitors and does not represent an error condition. Refer to the ha_rr_start_session man page details for action to be taken for the HA_RR_EAGAIN error.

2521-816 **Additional Event Manager client session request rejected.**

EXPLANATION

An additional request was made to create a client session with the Event Manger daemon.

USER RESPONSE

ha_rr_start_session() should only be called once to create a session with the Event Manager daemon for resource monitors defined as type client.

2521-817 **System call system call (args) failed with error: error number -error message.**

EXPLANATION

The specified system call failed with errno value.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-818 **No SIGIO signal handler found.**

EXPLANATION

No signal handler was defined for SIGIO.

USER RESPONSE

If the resource monitor's notification protocol is HA_RR_SIGIO, a SIGIO handler must be defined before calling the function.

2521-819 **A server resource monitor can have a maximum of maximum_number_of_sessions active sessions.**

EXPLANATION

The RMAPI can serve up to the maximum number of sessions.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-820 **No session found matching the session file descriptor: file_descriptor.**

EXPLANATION

No sessions were found matching the session file descriptor argument passed to the RMAPI routine.

USER RESPONSE

Check that the descriptor argument passed to the function was the one returned by a previous call to ha_rr_start_session().

2521-821 **Connection closed by the resource monitor manager.**

EXPLANATION

A session socket was found to be closed by the resource monitor manager.

USER RESPONSE

Review the actions to be taken for the type of resource monitor. If this is a server monitor that is started by the Event Manager daemon, and this was the last manager connection active, the monitor should exit. If the monitor will continue processing, all registered variables should be deleted for this session by calling **ha_rr_del_var()**. The session should then be ended by calling **ha_rr_end_session()**.

2521-822 **Resource monitor name is not defined to be a server in the EMCDB file.**

EXPLANATION

ha_rr_makserv() failed because the specified resource monitor is not defined as a server.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-823 **Resource monitor server socket already exists.**

EXPLANATION

The **ha_rr_makserv()** routine failed because it has already been successfully called.

USER RESPONSE

Check for multiple calls to the ha_rr_makserv() routine.

2521-824 **Unknown notification protocol argument.**

EXPLANATION

The value of a notification protocol argument passed to a RMAPI routine was not a valid value.

USER RESPONSE

Ensure that only the HA_RR_NOTIFY_SELECT or HA_RR_NOTIFY_SIGIO values are used as the notification protocol argument to the ha_rr_makserv and ha_rr_start_session routines.

2521-825 **PTX context(s) not found in the EMCDB for variable: variable_name.**

EXPLANATION

A PTX context was not found in the EMCDB file for the variable.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-829 **EMCDB version does not match the version in use by the Event Manager daemon.**

EXPLANATION

The timestamp or sequence number in the EMCDB file does not match the EMCDB used by the Event Manager daemon.

USER RESPONSE

Call ha_rr_terminate() then start the RMAPI session again to access the correct EMCDB.

2521-830 **The system call system call failed while attempting to send a response message to the manager session with socket descriptor socket descriptor : errno -standard error message.**

EXPLANATION

The system call failed to send a response message to the resource monitor manager.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-831 **The RMAPI received an unknown manager id from session with socket descriptor socket descriptor.**

EXPLANATION

The RMAPI received an unknown resource monitor manager id message on the specified socket.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-832 **RMAPI called with effective user id of (effective user id of caller). Effective id of 0 (root) is required.**

EXPLANATION

The resource monitor was not running with an effective id of 0 (root) when the RMAPI was called and the monitor has an monitor instance id of 0, and is configured to supply Counter/Quantity variables to PTX shared memory. Root authority is required by the Spmi library. Only monitors with an instance id of 0, and which are configured to supply PTX variables are required to have root authority due to the Spmi restriction.

USER RESPONSE

Check that the resource monitor executes with the required effective user id to use the RMAPI for supplying variables to PTX shared memory.

2521-837 **The load() system call failed while attempting to load module module name used for determining the domain name. The load() error is errno - err string. additional error strings (if any).**

EXPLANATION

The load system call failed to load an RMAPI module used to determine the domain name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-838 **The value value of environment variable name_envvar, does not match the environment domain name value name_value determined by the RMAPI.**

EXPLANATION

If the domain name environment variable is set, the RMAPI loads a module specific to the environment which is used to determine the domain name. The value returned by the load module is compared to the value set in the environment variable as a validation check. This messages indicates the values did not match.

USER RESPONSE

Make sure the environment variables are set correctly. If the resource monitor is started by the Event Management daemon, record the above information and contact the IBM Support Center.

2521-839 **The domain name environment variable, name_envvar, was set to value, but the domain type environment variable, type_envvar, was not set.**

EXPLANATION

The domain name environment variable, name_envvar, was set to specify the domain name to be used by the RMAPI. However, the domain type variable, type_envvar, was not set. If the name variable is set, the type variable must also be set so that the RMAPI can perform validation checking.

USER RESPONSE

Make sure that the process starting the resource monitor sets the domain environment variables correctly. If the monitor is started by the Event Management daemon, record the above information and contact the IBM Support Center.

2521-840 **Unknown value value found in domain type environment variable envvar.**

EXPLANATION

The environment variable, envvar, used to specify the RMAPI environment, was set to an unknown value. Valid values are SP, and HACMP.

USER RESPONSE

Make sure that the process starting the resource monitor sets the domain type environment variable correctly. If the monitor is started by the Event Management daemon, record the above information and contact the IBM Support Center.

2521-841 **The command command, used to obtain the domain name, was not found.**

EXPLANATION

The above command, which is used by the RMAPI to determine the domain name, was not found on the local system.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-842 **Unable to execute or determine the exit status of command command, used to obtain the domain name.**

EXPLANATION

The RMAPI was unable to determine the exit status of the above command, used to determine the domain name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-843 **The command, command, used to obtain the domain name, terminated with an exit value of value.**

EXPLANATION

The RMAPI executed the above command to determine the domain name. The command terminated with the exit value above.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-844 **The command, command, did not return a domain name.**

EXPLANATION

The RMAPI executed the above command to determine the domain name. The command terminate normally (with a 0 exit code), but did not return a domain name to the RMAPI.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-845 **Cannot add variables to an unidentified resource monitor manager session.**

EXPLANATION

The resource monitor attempted to add variables to a session that has not yet been identified as being with haemd or PTPE.

USER RESPONSE

If the monitor is a server, only call ha_rr_add_var() in response to receiving a HA_RR_CMD_ADDV or HA_RR_CMD_ADDALL command.

2521-846 **Cannot add State variable resource variable name, with resource ID resource ID, to the specified resource monitor manager session.**

EXPLANATION

The resource monitor attempted to add variables with a value type of State, to a Performance Monitor client session. This condition could occur if the resource monitor attempts to add State variables in response to receiving a HA_RR_CMD_ADDALL.

USER RESPONSE

The resource monitor should only add variables of value type Counter or Quantity in response to the command HA_RR_CMD_ADDALL.

2521-847 **Internal RMAPI error: internal error message**

EXPLANATION

An error internal to the RMAPI has occurred. The error message contains details of the error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-848 **The RMAPI subroutines have been locked following a previously returned error: lock condition**

EXPLANATION

Following a severe error detected by the RMAPI, api routines were locked from executing in order to keep the RMAPI resources in a consistent state from which it can successfully process ha_rr_terminate() to free its resources.

USER RESPONSE

Review the error returned by the RMAPI before receiving this error. Call ha_rr_terminate() to terminate the RMAPI session before continuing. Once the RMAPI has been terminated, ha_rr_init() may again be called.

2521-849 **No data was available to be read from the resource monitor manager session socket socket file descriptor.**

EXPLANATION

The RMAPI attempted to read data from the specified socket descriptor. The read() system call returned with an errno of EAGAIN indicating that there was no data available to be read.

USER RESPONSE

If the notify protocol of the monitor is HA_NOTIFY_SIGIO then this does not represent an error as an attempt to read each socket must be made whenever the SIGIO signal is received. If the protocol is HA_NOTIFY_SELECT the message also does not represent an error, but it is recommended for efficiency that the monitor only call ha_rr_get_ctrlmsg() for sockets that are ready to be read.

2521-850 **An error was found in the value of the resource variable variable name with resource ID variable resource ID. RMAPI errno - details.**

EXPLANATION

The RMAPI detected an error in the value of the specified variable. Additional details regarding the nature of the error follow.

USER RESPONSE

Correct the problem with the value before supplying it to RMAPI routines.

2521-851 **A pointer parameter passed to the RMAPI was NULL.**

EXPLANATION

The RMAPI was expecting a valid pointer as one of it's parameters, but was passed a NULL pointer value.

USER RESPONSE

Make sure that valid parameters are being passed to the RMAPI routine.

2521-852 **The command parameter passed to the RMAPI was not a valid value.**

EXPLANATION

The command parameter passed to the RMAPI routine ha_rr_rm_ctl() was not a valid value.

USER RESPONSE

Make sure that command parameter is a value as specified in the ha_rmapi.h header file and *RS/6000 CT: Event Management Programming Guide and Reference*.

2521-853 **The command to set values cannot be called after the RMAPI has been initialized.**

EXPLANATION

The command parameter passed to the RMAPI routine ha_rr_rm_ctl() specified that the routine should set one or more RMAPI values. Value may only be set by this routine prior to calling ha_rr_init().

USER RESPONSE

Call ha_rr_rm_ctl() to set values before calling ha_rr_init() to initialize the RMAPI.

2521-854 **The resource monitor instance id specified, instance-id is not a valid value. Resource monitor ids must be in the range legal range - legal range.**

EXPLANATION

The ha_rr_rm_ctl() routine was called to set the instance id of the resource monitor. The value passed to the RMAPI was outside the range of legal values.

USER RESPONSE

Call ha_rr_rm_ctl() with an instance id in the range noted. Refer to the ha_rmapi.h header file and *RS/6000 CT: Event Management Programming Guide and Reference* for details.

2521-855 **The resource monitor requested a nonzero resource monitor instance id, but is configured to allow only a single instance of the monitor to execute at a time.**

EXPLANATION

The ha_rr_rm_ctl() routine was called prior to ha_rr_init() to set the resource monitor id. The id that was specified was either a nonzero value, or the value of the HA_RR_RM_INSTID_NOPERF constant. When ha_rr_init() was called the RMAPI detected that the resource monitor was configured to allow only one instance to execute at a time. Such monitors must execute with an instance id of 0.

USER RESPONSE

Record the above information and contact the IBM Support Center

2521-856 **The resource monitor requested the instance id requested id, but is configured to only allow monitor instances in the range legal range - legal range.**

EXPLANATION

The ha_rr_rm_ctl() routine was called prior to ha_rr_init() to set the resource monitor id. Resource monitors that are configured to allow multiple copies to execute, must specify instance ids in the range 0 to (number of instances the RM is configured for minus one). The value specified by the resource monitor was not within this range.

USER RESPONSE

Specify a value for the resource monitor instance id which is valid for the configuration of the monitor.

2521-857 **The UNIX domain socket file name, path name , is too long.**

EXPLANATION

The name of the UNIX domain socket file to use to communicate with a resource monitor manager is too long to fit in an address structure.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-858 **The calling process must have an effective user id of 0 (root), or have the name of haem rm group group in its group list.**

EXPLANATION

Resource monitors must either have an effective id of 0 (root) or have the haemrm group in their group list in order to use the RMAPI.

USER RESPONSE

Check that the resource monitor executes with the required effective user id or has the haem rm group in its group list.

2521-859 **A session with the Event Management daemon already exists.**

EXPLANATION

The resource monitor accepted a connection with a resource monitor manager using the ha_rr_start_session() routine. The RMAPI determined that the session was with the Event Management daemon (haemd), but that there was an existing session with the haemd. Since only 1 session with haemd is allowed, the error probably occurred because the ha_rr_end_session() was not called to end the existing haemd session, prior to calling ha_rr_start_session().

USER RESPONSE

When a RMAPI routine returns the HA_RR_EDISCONNECT error, be sure to call the ha_rr_del_var() and ha_rr_end_session() routines to end the closed manager session.

2521-860 **The system call system call failed when attempting to invoke a function on file: file name. errno - standard message.**

EXPLANATION

The specified system call failed when attempting to perform a function on the above file name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-861 **The haem shared memory segment segment id was found to be corrupted at page page number cell offset cell offset.**

EXPLANATION

The RMAPI detected that the shared memory page had been corrupted. The shared memory may have been overwritten by the RM process or another process.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-862 **The load() system call failed while attempting to load the Spmi library (libSpmi.a). The load() error is errno - err string. additional error strings (if any).**

EXPLANATION

The load system call failed to load the Spmi library used to supply resource variable values to PTX shared memory for performance monitoring.

USER RESPONSE

Make sure that the perfagent.tools fileset is correctly installed. Record the above information and contact the IBM Support Center.

2521-863 **The size, file size, of CDB file file name, is not a valid CDB file size.**

EXPLANATION

Either the actual size of the CDB file, or the file size as reported within the CDB file, is not a valid size for a CDB file. This may indicate that the file was not copied correctly or has been corrupted.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-864 **The internal reported size, reported file size, of CDB file file name is larger than the actual file size of actual file size.**

EXPLANATION

The CDB file size reported within the CDB file was found to be larger than the actual file. This may indicate a corrupted or truncated CDB file.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2521-865 **The Event Management daemon detected the following error in the resource monitor: error number - details.**

EXPLANATION

The Event Management deamon sent the HA_RR_CMD_ERROR command to the RMAPI indicating it detected an error in the operations of the resource monitor or the RMAPI. The error number is a value defined in ha_rmapi.h following the form, HA_RR_CMD_ERR_*. The error number indicates the suspected source and cause of the error.

USER RESPONSE

Consult *RS/6000 CT: Event Management Programming Guide and Reference* for details regarding the specific error number. If the cause indicates a problem in the implementation of the resource monitor, correct the problem as described in the guide and retry the monitor. If the error indicates the problem is caused by the RMAPI, record the above information and contact the IBM Support Center.

2521-866 **The RMAPI was passed a resource variable handle value which was not valid.**

EXPLANATION

On a call to the RMAPI routine, the resource monitor passed a variable handle value used to identify a resource variable instance which has been added to one or more sessions. The value passed to the RMAPI was not a valid handle. The value may have been NULL, or did not reference variable instance. This error can occur if the monitor passes a handle which was no longer valid due to a previous call to ha_rr_del_var() or ha_rr_unreg_var().

USER RESPONSE

Correct the resource monitor so that it passes only valid handles returned by the ha_rr_add_var() routine. Refer to the ha_rr_del_var() and ha_rr_unreg_var() man pages regarding conditions which cause a handle to no longer be usable.

2521-867 **Attempt to connect to an Event Manager daemon with a UNIX domain socket failed: function name : errno - standard error message.**

EXPLANATION

The RMAPI attempted to connect to an Event Manager daemon through a UNIX domain socket, but failed. The specified function name returned the above standard error message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522 – Resource Monitor Messages

2522-000 **Could not start signal pipe for process poll signal.**

EXPLANATION

A Program Resource Monitor routine could not perform the initialization needed to handle a signal used to determine when the state of system processes should be polled.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-001 **Could not start signal pipe for process reap signal.**

EXPLANATION

A Program Resource Monitor routine could not perform the initialization needed to handle a signal used to determine when a monitored process has ended.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-002 **Could not start signal pipe for terminate signal.**

EXPLANATION

A Program Resource Monitor routine could not perform the initialization needed to handle a signal used to end the program.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-003 **Error from kernel extension routine KE_routine: error_message.**

EXPLANATION

A Program Resource Monitor routine called the Program Resource Monitor Kernel Extension routine KE_routine, which returned an error. The error message returned was error_ message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-004 **RMAPI routine RMAPI_routine returned error value error_number: error_message.**

EXPLANATION

A Program Resource Monitor routine called the RMAPI routine RMAPI_routine, which returned an error. The error number returned was error_number; the error message returned was error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-005 **flag argument flag_argument not two comma separated numbers.**

EXPLANATION

A Program Resource Monitor routine found a problem with the specified argument, flag_argument, for the flag flag. The argument is expected to be two comma separated numbers.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-006 **flag polling frequency argument specified_polling_frequency not positive.**

EXPLANATION

A Program Resource Monitor routine found a problem with the flag flag's argument. The first number of the argument, the polling frequency, is expected to be a positive number; specified_polling_frequency was specified instead.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-007 **flag re-polling frequency argument specified_re-polling_frequency not positive.**

EXPLANATION

A Program Resource Monitor routine found a problem with the flag flag's argument. The second number of the argument, the re-polling frequency, is expected to be a positive number; specified_re-polling_frequency was specified instead.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-008 **flag re-polling frequency argument specified_re-polling_frequency greater than polling frequency argument specified_polling_frequency.**

EXPLANATION

A Program Resource Monitor routine found a problem with the flag flag's argument. The second number of the argument, the re-polling frequency, is expected to be less than or equal to the first number of the argument, the polling frequency, however, it was not. The specified numbers were specified_polling_frequency and specified_re-polling_frequency.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-009 **flag flag requires an argument.**

EXPLANATION

A Program Resource Monitor routine found a problem with the flag flag. The flag is expected to have an argument, however, none was specified.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-010 **flag flag not recognized.**

EXPLANATION

A Program Resource Monitor routine found a problem with the program arguments. The flag flag was specified, but is not valid.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-011 **system_routine failed to change signal signal disposition: error_message.**

EXPLANATION

A Program Resource Monitor routine called system_routine to change the disposition of signal signal. However, system_routine returned an error described by error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-012 **system routine system_routine returned error: error_message**

EXPLANATION

A Program Resource Monitor routine called the system routine system_routine. However, the system routine returned an error described by error_message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-013 **Could not add a new session to the list of active sessions.**

EXPLANATION

A Program Resource Monitor routine could not get a new session structure added to the list of active sessions. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center. .

2522-014 **RMAPI_routine failed to returned a control message.**

EXPLANATION

A Program Resource Monitor routine called the RMAPI routine RMAPI_routine to get a control message. The RMAPI routine failed to return a control message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-015 **Lengths returned by RMAPI_routine do not agree.**

EXPLANATION

A Program Resource Monitor routine called the RMAPI routine RMAPI_routine to get a control message. The RMAPI routine returns the length of the control message returned, and the control message itself includes its length. These two lengths do not agree.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-016 **A command (command) solely meant for Counter or Quantity resource variables has been received from the RMAPI.**

EXPLANATION

A Program Resource Monitor routine received a command, command, from a Resource Monitor Manager solely meant for Counter or Quantity resource variables. Since the Program Resource Monitor provides only State resource variables, this command was not expected.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-017 **An unrecognized command (command) was received from the RMAPI.**

EXPLANATION

A Program Resource Monitor routine received an unrecognized command, command, from a Resource Monitor Manager.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-018 **Number of variables (number_of_variables) in control message from RMAPI not positive.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager that did not contain a positive number of variables. The number of variables specified was number_of_variables.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-019 **Error encountered converting user name to user identifier for resource variable (resource ID).**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable resource variable with resource ID resource ID be instantiated. The user name in the resource ID could not be converted to a user ID. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-020 **Could not start signal pipe for debug signal.**

EXPLANATION

A Program Resource Monitor routine could not perform the initialization needed to handle a signal used for debugging.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-021 **Could not create new variable structure for resource variable (resource ID).**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be instantiated. The resource variable was resource variable with resource ID resource ID. A variable structure could not be created to describe the new resource variable. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-022 **Could not add new variable structure for resource variable (resource ID) to instantiated resource variable list.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be instantiated. The resource variable was resource variable with resource ID resource ID. The variable structure describing the resource variable could not be added to the list of instantiated resource variables. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-023 **RMAPI routine could not register resource variable (resource ID).**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be instantiated. The resource variable was resource variable with resource ID resource ID . The above RMAPI routine was called to register the instantiated resource variable, but the RMAPI routine failed.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-024 **RMAPI routine RMAPI_routine reported registration error registration_error_number.**

EXPLANATION

A Program Resource Monitor routine called the RMAPI routine RMAPI_routine to register an instantiated resource variable. The RMAPI routine returned the registration error registration_error_number.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-025 **Requested to add resource variable (instance_identifier), which is not instantiated.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager, which requested that a resource variable be added (monitored). However, the resource variable is not instantiated. The instance identifier specified was instance_identifier.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-026 **Could not create new program structure for resource variable (instance_identifier).**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be added (monitored). However, a program structure could not be created to describe the program to be monitored on behalf of the resource variable. A previous error message may describe the cause of the problem. The instance identifier of the resource variable that was to be added was instance_identifier.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-027 **Could not add program (program_name;user_ID) to monitored programs list.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be added (monitored). However, the program structure describing the program to be monitored on behalf of the resource variable could not be added to the list of monitored programs. A previous error message may describe the cause of the problem. The program name specified was program_name, and the user identifier specified was user_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-028 **Could not add variable (instance_identifier) to program (program_name,user_ID) variable list.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager, which requested that a resource variable be added (monitored). However, the variable structure describing the resource variable could not be added to a list anchored by the program structure describing the program to be monitored on behalf of the resource variable. A previous error message may describe the cause of the problem. The instance identifier of the resource variable was instance_identifier. The program name was program_name, and the user ID was user_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-029 **RMAPI routine could not add resource variable (resource ID).**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be added (monitored). The resource variable was resource variable with resource ID resource ID. The RMAPI routine was called to add the monitored resource variable, but the RMAPI routine failed.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-030 **RMAPI routine RMAPI_routine reported add error add_error_number.**

EXPLANATION

A Program Resource Monitor routine called the RMAPI routine RMAPI_routine to add a monitored resource variable. The RMAPI routine returned the add error add_error_number.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-031 **RMAPI routine RMAPI routine did not return non-NULL handle for resource variable (resource ID).**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be added (monitored). The RMAPI routine was called to add the monitored resource variable. However, the RMAPI returned a NULL handle. The resource variable was resource variable with resource ID resource ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-032 **RMAPI routine RMAPI routine changed handle for resource variable (resource ID).**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be added (monitored). The RMAPI routine was called to add the monitored resource variable. However, the RMAPI returned a handle that differed from the resource variable's previous handle. The resource variable was resource variable with resource ID resource ID .

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-033 **Could not create new deferred add structure.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager, which requested that a resource variable be added (monitored). However, a deferred add structure could not be created to describe the add request. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-034 **Could not add new deferred add structure to list of deferred add requests.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager, which requested that a resource variable be added (monitored). However, the deferred add structure describing the add request could not be added to the list of deferred add requests. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-035 **Could not add instance identifier to deferred add structure.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager, which requested that a resource variable be added (monitored). However, the instance identifier of the resource variable to be monitored could not be added to a list anchored by a deferred add request structure. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-036 **Requested to delete resource variable (instance_identifier), which is not instantiated.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be deleted (not monitored). However, the resource variable is not instantiated. The instance identifier specified was instance_identifier.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-037 **Requested to delete resource variable (instance_identifier), which is not being monitored.**

EXPLANATION

There are two possibilities:

- A Program Resource Monitor received a control message from a Resource Monitor Manager, which requested that a resource variable be deleted (not monitored). However, the routine determined that the resource variable was not being monitored. The instance identifier specified was instance_identifier.
- A Program Resource Monitor was ending a RMAPI session because a disconnect had been detected. While ending the session, the routine was told to delete a resource variable that was not being monitored. The instance identifier specified was instance_identifier.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-038 **Unexpected return value from RMAPI_routine: return_value.**

EXPLANATION

A Program Resource Monitor routine called the RMAPI routine RMAPI_routine, which returned an unexpected return value, return_value.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-039 **RMAPI routine RMAPI_routine did not return NULL handle, as expected.**

EXPLANATION

A Program Resource Monitor routine called the RMAPI routine RMAPI_routine to delete a resource variable. The RMAPI routine returned a value indicating that the resource variable should no longer be monitored. However, the RMAPI routine also returned a handle that was not NULL.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-040 **Could not remove variable (instance_identifier) from program's variable list.**

EXPLANATION

The variable structure describing a resource variable could not be removed from a list anchored by the program structure describing the program being monitored on behalf of the resource variable. A previous error message may describe the cause of the problem. The instance identifier of the resource variable was instance_identifier.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-041 **Could not remove program (program_name;user_ID) from monitored programs list.**

EXPLANATION

The program structure describing the program that was monitored on behalf of a resource variable could not be removed from the list of monitored programs. A previous error message may describe the cause of the problem. The program name was program_name, and the user identifier was user_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-042 **Could not remove program (program_name;user_ID) from changed programs list.**

EXPLANATION

The program structure describing a program that is being monitored on behalf of a resource variable could not be removed from the list of changed programs. A previous error message may describe the cause of the problem. The program name was program_name, and the user identifier was user_ID

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-043 **Could not delete unneeded program structure (program_name;user_ID).**

EXPLANATION

The program structure describing a program that was being monitored on behalf of a resource variable could not be destroyed. A previous error message may describe the cause of the problem. The program name was program_name, and the user identifier was user_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-044 **Could not remove process (process_ID) from program (program_name;user_ID) process list.**

EXPLANATION

The process structure describing a process being monitored on behalf of a program could not be removed from a list anchored by the program structure describing the program. A previous error message may describe the cause of the problem. The process identifier was process_ID. The program name was program_name, and the user identifier was user_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-045 **Could not remove process (process_ID) from monitored process list.**

EXPLANATION

The process structure describing a process being monitored on behalf of a program could not be removed from the list of monitored processes. A previous error message may describe the cause of the problem. The process identifier was process_ID

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-046 **Could not delete unneeded process structure (process_ID).**

EXPLANATION

The process structure that had described a process being monitored on behalf of a program could not be destroyed. A previous error message may describe the cause of the problem. The process identifier was process_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-047 **Found instance identifier (instance_identifier) with no resource variable.**

EXPLANATION

A Program Resource Monitor routine encountered what was believed to be a valid instance identifier, instance_identifier, which had no resource variable associated with it.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-048 **Could not remove session (session_file_descriptor) from active session list.**

EXPLANATION

The session structure describing a RMAPI session could not be removed from the list of active sessions. A previous error message may describe the cause of the problem. The session file descriptor was session_file_descriptor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-049 **Kernel extension reported the death of a process (process_ID) that is not being monitored.**

EXPLANATION

A Program Resource Monitor Kernel Extension reported to a Program Resource Monitor routine the end of a process, whose PID was process_ID, which the Program Resource Monitor was not monitoring.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-050 **Monitored process (process_ID) does not have an associated monitored program.**

EXPLANATION

A Program Resource Monitor routine encountered a monitored process that did not have an associated monitored program. The PID of the process was process_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-051 **Could not put program (program_name;user_ID) on changed program list.**

EXPLANATION

A Program Resource Monitor routine could not place a program structure on the changed program list. The program name was program_name, and the user identifier was user_ID. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-052 **Could not access system process table entry.**

EXPLANATION

A Program Resource Monitor routine could not access the system's process table. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-053 **Could not create new process structure for process_ID.**

EXPLANATION

A Program Resource Monitor routine could not create a process structure to describe the process with PID process_ID A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-054 **Could not add process (process_ID) to program (program_name;user_ID) process list.**

EXPLANATION

A Program Resource Monitor routine could not add the process structure describing the process with PID process_ID to a list anchored by the program structure describing the program for which the process is to be monitored. A previous error message may describe the cause of the problem. The program name was program_name, and the user ID was user_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-055 **Process (process_ID) was to be reused, but it is in an unexpected state.**

EXPLANATION

A Program Resource Monitor routine discovered a process with PID process_ID already associated with a program, but with an unexpected state.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-056 **Could not add process (process_ID) to monitored process list.**

EXPLANATION

A Program Resource Monitor routine could not add the process structure describing the process with PID process_ID to the list of monitored processes. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-057 **Could not remove new deferred add structure from list of deferred add requests.**

EXPLANATION

A deferred add structure could not be removed from the list of deferred add requests. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-058 **Could not delete unneeded deferred add structure.**

EXPLANATION

A deferred add structure could not be destroyed. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-059 **Cannot do deferred add of resource variable (instance_identifier); it is not instantiated.**

EXPLANATION

A Program Resource Monitor routine encountered what was believed to be a valid instance identifier, instance_identifier, that had no resource variable associated with it. The resource variable was supposed to be added (monitored).

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-060 **Cannot do deferred add of resource variable (instance_identifier); no associated program found.**

EXPLANATION

A Program Resource Monitor routine encountered a resource variable, with instance identifier instance_identifier, for which deferred add processing was needed, however, it had no program associated with it.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-061 **Could not initialize process lists for program (program_name;user_ID).**

EXPLANATION

A Program Resource Monitor routine could not initialize some process lists. The program name was program_name, and the user identifier was user_ID. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-062 **For report phase report_phase, unexpected process state (process_state) for process (process_ID) associated with program (program_name;user_ID).**

EXPLANATION

A Program Resource Monitor routine was running in report phase report_phase, and found a process with PID process_ID that had an unexpected state: process_state. The process was associated with a program with the name program_name and the user identifier user_ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-063 **Could not create new structured byte string for program (program_name;user_ID).**

EXPLANATION

A Program Resource Monitor routine could not create a new structured byte string for a program. The program name was program_name, and the user identifier was user_ID. A previous error message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-064 **User name in resource variable (resource ID) is unknown.**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be instantiated. The resource variable was resource variable with resource ID . The user name in the resource ID could not be converted to a user ID, because the user name is unknown.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-065 **Program Resource Monitor cannot monitor itself: resource variable (resource ID).**

EXPLANATION

A Program Resource Monitor routine received a control message from a Resource Monitor Manager which requested that a resource variable be instantiated. The resource variable was resource variable with resource ID . The resource ID describes the Program Resource Monitor, which cannot monitor itself.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-066 **Could not remove variable (instance_identifier) from instantiated resource variable list.**

EXPLANATION

A Program Resource Monitor routine could not remove the resource variable with instance identifier instance_identifier . from the instantiated resource variable list. A previous message may describe the cause of the problem.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-067 **Could not delete unneeded variable structure (instance_identifier).**

EXPLANATION

A Program Resource Monitor routine could not delete the variable structure representing the resource variable with instance identifier instance_identifier.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-068 **Cannot delete object_type object because object_field_name field is not free.**

EXPLANATION

A Program Resource Monitor routine could not delete an object of type object_type because the memory associated with the object's object_field_name field has not been freed.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-069 **Number of elements used is greater than number allocated.**

EXPLANATION

A Program Resource Monitor routine found an array where the number of elements used is greater than the number of elements allocated.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-070 **Null resource variable name received.**

EXPLANATION

A Program Resource Monitor routine received an empty resource variable name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-071 **Unknown resource variable name received: resource_variable_name.**

EXPLANATION

A Program Resource Monitor routine received an unknown resource variable name, resource_variable_name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-072 **Null resource ID received.**

EXPLANATION

A Program Resource Monitor routine received an empty resource ID.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-073 **Resource ID resource ID contains one of the following illegal characters illegal characters.**

EXPLANATION

A Program Resource Monitor routine received a resource ID resource ID, which includes one or more of the above illegal characters.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-074 **Resource ID resource ID does not contain number name/value pairs.**

EXPLANATION

A Program Resource Monitor routine received a resource ID, resource ID which does not include number name/value pairs.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-075 **Resource ID resource ID contains more than number name/value pairs.**

EXPLANATION

A Program Resource Monitor routine received a resource ID, resource ID, which includes more than number name/value pairs.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-076 **Resource ID resource ID does not contain the name name/value pair.**

EXPLANATION

A Program Resource Monitor routine received a resource ID, resource ID, which does not include a name/value pair for name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-077 **Resource ID resource ID : value for name exceeds size limit characters.**

EXPLANATION

A Program Resource Monitor routine received a resource ID, which includes a name/value pair for name. The value for this name exceeds the size limit above.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-078 **Name/value pair name/value pair is missing name.**

EXPLANATION

A Program Resource Monitor routine received a resource ID name/value pair, name/value pair, without a name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-079 **Name/value pair name/value pair is missing value.**

EXPLANATION

A Program Resource Monitor routine received a resource ID name/value pair, name/value pair, without a value.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-080 **Name/value pair name/value pair has extra character.**

EXPLANATION

A Program Resource Monitor routine received a resource ID name/value pair with an extra character.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-081 **Cannot support new session. Can only support limit sessions.**

EXPLANATION

A Program Resource Monitor routine detected a new RMAPI session established with the resource monitor, and the number of sessions established exceeds the maximum number of RMAPI sessions supported by the resource monitor, limit.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-082 **Cannot remove unknown session session_descriptor.**

EXPLANATION

A Program Resource Monitor routine was requested to remove an unknown RMAPI session, session_descriptor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-083 **The process table has not been obtained.**

EXPLANATION

A Program Resource Monitor routine was called to provide a process table entry, but it was called at an improper time.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-084 **Specification as to where to start scanning the process table is not valid.**

EXPLANATION

A Program Resource Monitor routine was called to provide a process table entry, but the specification as to where to start scanning the process table was not valid.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-085 **Instance identifier (instance_identifier) is out of range.**

EXPLANATION

A Program Resource Monitor routine received an instance identifier, instance_identifier, which is outside the range of currently valid instance identifiers.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-086 **Instance identifier (instance_identifier) not in use.**

EXPLANATION

A Program Resource Monitor routine received an instance identifier, instance_identifier, which is not currently in use.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-087 **Instance identifier mismatch (instance_ID_1/instance_ID_2).**

EXPLANATION

A Program Resource Monitor routine found a resource variable structure that should have represented an instantiated resource variable whose instance identifier was instance_ID_1. However, the resource variable structure indicated that the resource variable's instance identifier was instance_ID_2.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-088 **No instance identifiers in use.**

EXPLANATION

A Program Resource Monitor routine unexpectedly found that no instance identifiers are currently in use.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-089 **The object is already linked to some list.**

EXPLANATION

A Program Resource Monitor routine was requested to link an object into a list using a certain field of the object. However, the object appears to be already linked into a list with the specified field.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-090 **Duplicate object found on the list.**

EXPLANATION

A Program Resource Monitor routine was requested to link an object into a list. However, an object was found to already be on the list with the same key as the object to be placed into the list.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-091 **The object is not on the list.**

EXPLANATION

A Program Resource Monitor routine was requested to remove an object from a list. However, the field specified through which the object should be linked into the list indicates that the object is not linked into the list.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-092 **Could not trim structured byte string to number_of_bytes bytes.**

EXPLANATION

A Program Resource Monitor routine could not trim the length of a structured byte string to number_of_bytes bytes.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-601 **daemon support routine : error message.**

EXPLANATION

An AIXOS Resource Monitor routine called the daemon support routine above. However, the support routine returned an error described by error message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-602 **Could not start signal pipe for terminate signal.**

EXPLANATION

An AIXOS Resource Monitor routine could not perform the initialization needed to handle a signal used to terminate the program.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-603 **Could not start signal pipe for alarm signal.**

EXPLANATION

An AIXOS Resource Monitor routine could not perform the initialization needed to handle a signal used to receive alarms.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-604 **RMAPI routine RMAPI routine error error number : error message.**

EXPLANATION

An AIXOS Resource Monitor routine called the RMAPI routine , which returned an error. The error number returned and associated message are stated above.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-605 **flag argument flag argument not a number.**

EXPLANATION

An AIXOS Resource Monitor routine found a problem with the specified argument for the above flag. The argument is expected to be a number.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-606 **flag instantiation interval argument specified instantiation interval is negative.**

EXPLANATION

An AIXOS Resource Monitor routine found a problem with the argument of the above flag. The argument is the instantiation interval and it is expected to be 0 or a positive number. The value specified is not.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-607 **flag domain type argument specified domain type not valid.**

EXPLANATION

An AIXOS Resource Monitor routine found a problem with the argument of the above flag. The argument, the domain type, is expected to be SP or HACMP.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-608 **flag domain name argument specified domain name not valid.**

EXPLANATION

An AIXOS Resource Monitor routine found a problem with the argument of the above flag. The argument is the domain name and it is not expected to be a null string.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-609 **-flag flag requires an argument.**

EXPLANATION

An AIXOS Resource Monitor routine found a problem with the above flag. The flag is expected to have an argument, but none was specified.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-610 **-flag flag not recognized.**

EXPLANATION

An AIXOS Resource Monitor routine found a problem with the program arguments. The above flag was specified, but is not valid.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-611 **flag flag not specified.**

EXPLANATION

An AIXOS Resource Monitor routine found a problem with the program arguments. The above flag was not specified, but it is required.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-612 **system routine system routine returned error: error message.**

EXPLANATION

An AIXOS Resource Monitor routine called the system routine above. However, the system routine returned an error described by error message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-613 **assertion failure: (failed assertion).**

EXPLANATION

An AIXOS Resource Monitor routine found that the above assertion was not true. The assertion should always be true. This error indicates a logic error in the resource monitor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-614 **assertion failure: (value 1 string operator value 2 string) value1: value 1 value2: value 2 .**

EXPLANATION

An AIXOS Resource Monitor routine found that an assertion was not true. The assertion should always be true. This error indicates a logic error in the resource monitor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-615 **assertion failure: (value 1 string operator value 2 string) value1: value 1 value2: value 2 .**

EXPLANATION

An AIXOS Resource Monitor routine found that an assertion was not true. The assertion should always be true. This error indicates a logic error in the resource monitor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-616 **assertion failure: (value 1 string operator value 2 string) value1: value 1 value2: value 2 .**

EXPLANATION

An AIXOS Resource Monitor routine found that an assertion was not true. The assertion should always be true. This error indicates a logic error in the resource monitor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-617 **An unrecognized command (command) was received from the RMAPI.**

EXPLANATION

An AIXOS Resource Monitor routine received an unrecognized command, from a Resource Monitor Manager.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-618 **Number of variables (number of variables) in control message from RMAPI not positive.**

EXPLANATION

An AIXOS Resource Monitor routine received a control message from a Resource Monitor Manager that did not contain a positive number of variables.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-619 **EMCDB has statistic statistic name unexpectedly at top of SPMI context tree.**

EXPLANATION

An AIXOS Resource Monitor routine found a statistic named in the Event Management configuration database at the top of the SPMI context tree. This is unexpected.

USER RESPONSE

If this was defined locally, fix the configuration. Otherwise, record the above information and contact the IBM Support Center.

2522-620 **EMCDB has context statistic name not in SPMI context tree.**

EXPLANATION

An AIXOS Resource Monitor routine found a statistic named in the Event Management configuration database that is not in the SPMI context tree.

USER RESPONSE

If this was defined locally, fix the configuration. If not, record the above information and contact the IBM Support Center.

2522-621 **SPMI routine SPMI routine returned error (error code): error message .**

EXPLANATION

An AIXOS Resource Monitor routine called the above SPMI routine . However, the SPMI routine returned an error described by error code and error message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-622 **SPMI-like routine SPMI-like routine failed.**

EXPLANATION

An AIXOS Resource Monitor routine called the SPMI-like routine above, which failed. The details of the failure are given in a prior message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-623 **ha_rr_reg_var() registered actually registered of attempted to register variables.**

EXPLANATION

An AIXOS Resource Monitor routine attempted to register resource variable instances with the RMAPI. Only actually registered of attempted to register resource variable instances were registered.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-624 **RMAPI routine ha_rr_reg_var() reported registration error error number for name, resource ID, instance ID.**

EXPLANATION

An AIXOS Resource Monitor routine called ha_rr_reg_var() to register a resource variable instance with the above name, resource ID, and instance ID. The RMAPI routine returned the registration error number above.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-625 **ha_rr_add_var() added actually added of attempted to add variables.**

EXPLANATION

An AIXOS Resource Monitor routine attempted to add resource variable instances with the RMAPI. Only actually added of attempted to add resource variable instances were added.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-626 **RMAPI routine ha_rr_add_var() reported add error error number for name, resource ID.**

EXPLANATION

An AIXOS Resource Monitor routine called ha_rr_add_var() to add a resource variable instance with the above name and resource ID. The RMAPI routine returned the add error error number.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-627 **ha_rr_unreg_var() unregistered actually unregistered of attempted to unregister variables.**

EXPLANATION

An AIXOS Resource Monitor routine attempted to unregister resource variable instances with the RMAPI. Only actually unregistered of attempted to unregister resource variable instances were unregistered.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-628 **RMAPI routine ha_rr_unreg_var() reported unregistration error error number for name, resource ID.**

EXPLANATION

An AIXOS Resource Monitor routine called ha_rr_unreg_var() to unregister a resource variable instance with the above name and resource ID . The RMAPI routine returned the unregistration error error number.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-629 **SPMI routine SPMI routine returned error error number : error message.**

EXPLANATION

An AIXOS Resource Monitor routine called the SPMI routine above . However, the SPMI routine returned an error described by error number and error message.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-630 **SPMI statistic returned error code error number.**

EXPLANATION

An AIXOS Resource Monitor routine encountered a statistic with error code error number.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-631 **Cannot support new session. Can only support limit sessions.**

EXPLANATION

An AIXOS Resource Monitor routine detected that a new RMAPI session has been established with the resource monitor, and that the number of sessions established exceeds the maximum number of RMAPI sessions supported by the resource monitor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-632 **Cannot remove unknown session session descriptor.**

EXPLANATION

An AIXOS Resource Monitor routine was requested to remove an unknown RMAPI session, with the above session descriptor.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-633 **ODM routine ODM routine returned an error: ODM error message.**

EXPLANATION

An AIXOS Resource Monitor routine called the above ODM routine, which returned an error. ODM error message describes the error.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-634 **LVM routine LVM routine returned an error: LVM error number.**

EXPLANATION

An AIXOS Resource Monitor routine called the LVM routine, which returned the LVM error number above.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-635 **The command command terminated with the exit value exit value.**

EXPLANATION

An AIXOS Resource Monitor routine had run the command identified above to obtain information about the HACMP domain. The exit value of the command, exit value, indicates an error occurred.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-636 **The command command returned unexpected status: 0x status.**

EXPLANATION

An AIXOS Resource Monitor routine had run the command identified above to obtain information about the HACMP domain. The termination status of the command, was unexpected.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2522-637 **The command command failed to return a domain name.**

EXPLANATION

An AIXOS Resource Monitor routine had run the command identified above to obtain information about the HACMP domain. It did not return a domain name.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523 – Topology Services Messages

2523-001 **My address missing from the adapter information.**

EXPLANATION

My address, which is found by examining the output of netstat -i, could not be found in the generated machines list file.

USER RESPONSE

Verify that the adapter's address matches that listed in the global ODM (adpater class).

2523-002 **GROUP_PROCLAIM message from outside the configuration, ID=adapter_id, ignored.**

EXPLANATION

Daemon received a Group Proclaim message from a node outside the current configuration. One or more daemons may have outdated configurations.

USER RESPONSE

If the message persists, contact the IBM Support Center.

2523-003 **Got a heartbeat from someone that is not my neighbor, ID=adapter_id GID=group_id.**

EXPLANATION

A node did not respond to committing of new groups.

USER RESPONSE

If the problem persists, the daemon on the offending node should be reset.

2523-005 **Failed to send heartbeat to adapter_id.**

EXPLANATION

Failure to send a heartbeat message, possibly due to network problems.

USER RESPONSE

Fix any network problems.

2523-006 **Unable to fork child process.**

EXPLANATION

Attempt to fork failed, possibly due to too many running processes.

USER RESPONSE

Determine and correct the resource problem.

2523-007 **argument_value: Not a valid argument_flag flag; positive integer expected.**

EXPLANATION

The daemon was invoked with a bad value for the indicated argument.

USER RESPONSE

Correct the argument value to a positive integer value. This may be a problem with the topsvcs script. Contact IBM Support.

2523-008 **argument_value: Not a valid argument_flag flag; non-negative integer <upper_bound expected.**

EXPLANATION

The daemon was invoked with a bad value for the indicated argument.

USER RESPONSE

Correct the argument value to a nonnegative less than the upper bound. This may be a problem with the topsvcs script. Contact IBM Support.

2523-009 **argument_value: Not a valid argument_flag flag.**

EXPLANATION

Daemon was invoked with an incorrect flag.

USER RESPONSE

Specify a flag with the correct value. This may be a problem with the topsvcs script. Contact IBM Support.

2523-010 **argument_value: Not a valid argument_flag flag; integer expected.**

EXPLANATION

Daemon was invoked with an incorrect flag.

USER RESPONSE

Specify a flag with an integer value. This may be a problem with the topsvcs script. Contact IBM Support.

2523-011 **Unrecognized parameter parameter**

EXPLANATION

Daemon was invoked with an unrecognized parameter.

USER RESPONSE

Specify the parameter correctly. This may be a problem with the topsvcs script. Contact IBM Support.

2523-012 **Unable to obtain socket port number from /etc/services.**

EXPLANATION

topsvcs.<partition name> entry missing in /etc/services

USER RESPONSE

The cluster was not correctly configured.

2523-013 **adapter_offset: Not a valid HB_SOCKET port number: expecting lower_bound - upper_bound.**

EXPLANATION

The specified port number is not in the range lower_bound - upper_bound.

USER RESPONSE

Specify HB_SOCKET port number in the indicated range. This may be a problem with the topsvcs script. Contact IBM Support.

2523-014 **My node number is not defined.**

EXPLANATION

This node was not specified in the configuration.

USER RESPONSE

Correct the configuration error.

2523-015 **Don't know how to broadcast message_type messages.**

EXPLANATION

The daemon is not capable of broadcasting a message of the indicated type.

USER RESPONSE

Contact the IBM Support Center.

2523-016 **Unable to send message, remote socket is not ready for writing.**

EXPLANATION

The attempt to write to the Remote Socket failed.

USER RESPONSE

Verify the network is working correctly.

2523-017 **Received a message_type message for an unknown address address.**

EXPLANATION

Received a message that is not destined for any node in the configuration.

USER RESPONSE

Contact the IBM Support Center.

2523-018 **Could not obtain HB_SHARED_MEM_KEY shared_memory_id.**

EXPLANATION

Unable to obtain a shared memory key.

USER RESPONSE

Use ipcrm to clean up shared memory.

2523-019 **Could not obtain HB_SHARED_SEM_KEY.**

EXPLANATION

Unable to obtain a shared memory key.

USER RESPONSE

Use ipcrm to clean up shared memory.

2523-020 **Could not create shared memory segment shared_memory_segment_number.**

EXPLANATION

Unable to create a shared memory segment.

USER RESPONSE

Use ipcrm to clean up shared memory.

2523-021 **Could not attach shared memory segment shared_memory_segment_number.**

EXPLANATION

Unable to attach to a shared memory segment.

USER RESPONSE

Use ipcrm to clean up shared memory.

2523-022 **Could not create semaphore.**

EXPLANATION

Unable to create a semaphore.

USER RESPONSE

Use ipcrm to clean up semaphores.

2523-023 **Could not initialize semaphore.**

EXPLANATION

Unable to initialize a semaphore.

USER RESPONSE

Use ipcrm to clean up semaphores.

2523-024 **key_number: Not a valid HB_SHARED_MEM_KEY number: expecting lower_bound - upper_bound.**

EXPLANATION

Shared memory key number is not in the range lower-bound — upper-bound.

USER RESPONSE

Specify the shared memory key number correctly.

2523-025 **key_number: Not a valid HB_SHARED_SEM_KEY number: expecting lower_bound - upper_bound.**

EXPLANATION

Shared memory key number is not in the range lower-bound — upper-bound.

USER RESPONSE

Specify the shared memory key number correctly.

2523-026 **Could not open Local Socket.**

EXPLANATION

An attempt to open the connection socket failed, possibly because of permissions.

USER RESPONSE

Remove the local socket file.

2523-027 **option: setsockopt on Local Socket failed.**

EXPLANATION

Unable to set the socket option on the local socket.

USER RESPONSE

Try to set the socket option on the local socket again.

2523-028 **Could not bind Local Socket = adapter_offset to file_descriptor.**

EXPLANATION

Unable to bind the local socket with the indicated file descriptor.

USER RESPONSE

Remove the socket file and try again.

2523-029 **Unable to send message on Remote Socket, message too long.**

EXPLANATION

The sendto command returned an errno of EMSGSIZE due to the size of the message.

USER RESPONSE

None.

2523-030 **Client message from pid=(PID) on an unregistered socket, message ignored.**

EXPLANATION

Received a client message from the indicated pid that is no longer a registered client, or possibly registered with a different pid.

USER RESPONSE

This process must register to have access.

2523-031 **Incomplete client message sent, bytes_sent out of msg_size bytes sent.**

EXPLANATION

Unable to send the entire message.

USER RESPONSE

None.

2523-033 **Unable to send Client message, the message pointer is not valid.**

EXPLANATION

The message pointer is not valid.

USER RESPONSE

Start the hats daemon again in this partition.

2523-035 **Listening Socket Error ID = error_id, not recognized.**

EXPLANATION

An error was generated that was not recognized.

USER RESPONSE

None.

2523-036 **Error Accepting Client Connection, errno=errno.**

EXPLANATION

The accept of a connection request failed.

USER RESPONSE

None.

2523-037 **Error sending data to pid=pid, rc = return_code, errno = errno.**

EXPLANATION

Error sending data to indicated process id.

USER RESPONSE

None.

2523-038 **Error receiving data from pid=pid, rc = return_code, errno = errno.**

EXPLANATION

Error receiving data from indicated process id.

USER RESPONSE

None.

2523-039 **Could not open Remote Socket.**

EXPLANATION

Unable to open remote socket, may indicate lack of resources.

USER RESPONSE

Identify and correct the problem and try again.

2523-040 **option: setsockopt on Remote Socket failed.**

EXPLANATION

Unable to set the socket option on the remote socket.

USER RESPONSE

Try again.

2523-041 **Could not bind Remote Socket = file_descriptor.**

EXPLANATION

Unable to bind the remote socket with the indicated file descriptor.

USER RESPONSE

Check to see if port number is already in use.

2523-042 **Getsockname on Remote Socket failed.**

EXPLANATION

An internal daemon error may be the cause of the failure.

USER RESPONSE

None.

2523-043 **Unmatching From ID.**

EXPLANATION

The address of the sender and the from address in the message do not match.

USER RESPONSE

None.

2523-044 **Received message_size bytes from sender_address with unmatching from ID.**

EXPLANATION

The address of the sender and the from address in the message do not match.

USER RESPONSE

Verify the source of the message.

2523-045 **Bad Socket Descriptor.**

EXPLANATION

The socket descriptor has been corrupted.

USER RESPONSE

Start the topology services daemon again.

2523-046 **Internal Data Error.**

EXPLANATION

An internal data error was detected.

USER RESPONSE

None.

2523-047 **Caught Signal.**

EXPLANATION

The select call ended due to a signal.

USER RESPONSE

None.

2523-048 **Select data value not valid.**

EXPLANATION

A parameter to a select call was not valid.

USER RESPONSE

Start the daemon again.

2523-049 **Select data pointer not valid.**

EXPLANATION

The data pointer to the select call is not valid.

USER RESPONSE

Start the daemon again.

2523-050 **Unexpected error from select.**

EXPLANATION

Unexpected errno set by the select call.

USER RESPONSE

Start the daemon again.

2523-051 **Could not open machine file file_name.**

EXPLANATION

Unable to open the machines list file indicated.

USER RESPONSE

Check to see if the file was created and has appropriate permissions.

2523-052 **Could not open a pipe to netstat.**

EXPLANATION

Unable to start the netstat command or create the pipe to it to gather output from the netstat command. Uses /usr/bin/netstat.

USER RESPONSE

Verify the existence of the netstat command.

2523-055 **Node number duplicated: string.**

EXPLANATION

The string is the line in the machines list that duplicates the node number.

USER RESPONSE

Fix the inconsistency in the configuration.

2523-056 **Gethostbyname host_name Failed, errno = errno.**

EXPLANATION

Unable to convert the host name to an IP address.

USER RESPONSE

Verify the host name is known to the network/name server.

2523-057 **IP address duplicated: node node_number address IP_address.**

EXPLANATION

The indicated IP address was already encountered in the configuration.

USER RESPONSE

Fix the inconsistency in the configuration.

2523-058 **Ignoring Adapter adapter_name at offset adapter_offset on node node_number.**

EXPLANATION

Unable to convert the adapter name adapter_name of the adapter at the adapter offset adapter_offset on node indicated to an IP address.

USER RESPONSE

Verify that adapter is known to the network/name server.

2523-059 **IP address changed for adapter at offset adapter_offset on node node_number.**

EXPLANATION

The IP address for the adapter on the indicated node changed at adapter offset adapter_offset.

USER RESPONSE

If this is an intended change, start all daemons again. Otherwise, correct the inconsistency.

2523-060 **Configuration error - no nodes defined.**

EXPLANATION

No nodes or adapters defined in the machines list file.

USER RESPONSE

Check to see if the configuration has nodes defined.

2523-061 **Refresh: deleted myself!**

EXPLANATION

An interface on this node has been omitted from the configuration.

USER RESPONSE

Verify if this is intended.

2523-062 **Operation: ioctl failed.**

EXPLANATION

The ioctl operation failed.

USER RESPONSE

Determine the reason for the failure, if possible.

2523-063 **Node is not in my group, ignoring Node Connectivity message!**

EXPLANATION

Received a node connectivity message from a node that is not in my group.

USER RESPONSE

Start the daemon the message is coming from, if problem persists.

2523-064 **Node number for adapter_offset is not valid, ignoring message!**

EXPLANATION

The indicated adapter ID does not correspond to a valid node number.

USER RESPONSE

Verify the source of the message.

2523-065 **NCT unable to compute node reachability!**

EXPLANATION

Failure to compute node reachability.

USER RESPONSE

Start the daemon again.

2523-066 **Node number of adapter_offset in my group not found.**

EXPLANATION

Unable to map an adapter in my group to a node number.

USER RESPONSE

If the problem persists, contact IBM Support.

2523-067 **Adapter(adapter_id) at offset adapter_offset not found.**

EXPLANATION

Unable to find this group member in NCT.

USER RESPONSE

If the problem persists, boot the daemon again.

2523-069 **No members in current group.**

EXPLANATION

No members were found in the current group.

USER RESPONSE

None.

2523-070 **My adapter not in current group.**

EXPLANATION

The adapter is not a member of my own group.

USER RESPONSE

None.

2523-071 **Adapter adapter_offset is already in the current group.**

EXPLANATION

Attempted to merge groups that had a member in common.

USER RESPONSE

None.

2523-072 **Missing adapter not found in my group.**

EXPLANATION

Members to be removed from the group were already missing.

USER RESPONSE

None.

2523-073 **My node number is not defined.**

EXPLANATION

The node number is not defined.

USER RESPONSE

Contact IBM Support.

2523-074 **Adapter offset adapter_offset is not valid.**

EXPLANATION

Adapter offset indicated is out of range.

USER RESPONSE

Specify a valid adapter offset.

2523-075 **Old subscription entry old_subscription_entry still exists.**

EXPLANATION

An old subscription entry already exists.

USER RESPONSE

The old subscription entry must be removed before a new one can be created.

2523-076 **Unknown subscription type.**

EXPLANATION

Not a valid subscription type.

USER RESPONSE

Specify a valid subscription type.

2523-077 **Cannot send Hb_No_Event to clients.**

EXPLANATION

Not a valid event.

USER RESPONSE

Specify a valid subscription event.

2523-078 **Cannot send Hb_All_Events to clients.**

EXPLANATION

Not a valid event.

USER RESPONSE

Specify a valid subscription event.

2523-079 **Cannot send Unknown event to clients.**

EXPLANATION

Not a valid event.

USER RESPONSE

Specify a valid subscription event.

2523-080 **Cannot delete a NULL entry.**

EXPLANATION

Attempted to delete a null interest entry.

USER RESPONSE

None.

2523-081 **Cannot delete entry from an empty list.**

EXPLANATION

Attempted to delete a subscription entry from an empty list.

USER RESPONSE

None.

2523-082 **ACK retry count exhausted, no ACK sent.**

EXPLANATION

Already sent maximum number of retry ACKs.

USER RESPONSE

None.

2523-083 **Leader died and originator is not the crown prince.**

EXPLANATION

Group leader died and the PTC message was not from the crown prince.

USER RESPONSE

None.

2523-084 **From Id = sender_ID From Group Id = sender_GID.**

EXPLANATION

Display sender's ID and GID.

USER RESPONSE

None.

2523-085 **Crown Prince Id = ID.**

EXPLANATION

Display the crown prince's ID.

USER RESPONSE

None.

2523-086 **PTC NAK ignored, I am not the group leader.**

EXPLANATION

Received a PTC negative acknowledgment, but am not the group leader.

USER RESPONSE

None.

2523-087 **PTC NAK ignored, I am not committing.**

EXPLANATION

Received a PTC negative acknowledgment, but not processing a group change.

USER RESPONSE

None.

2523-088 **PTC NAK ignored, no longer in PTC mode.**

EXPLANATION

Receive a PTC negative acknowledgment, but no longer accepting them.

USER RESPONSE

None.

2523-089 **PTC ACK ignored, I am not the group leader.**

EXPLANATION

Received a PTC acknowledgment, but am not the group leader.

USER RESPONSE

None.

2523-090 **PTC ACK ignored, I am not committing.**

EXPLANATION

Received a PTC acknowledgment, but not processing a group change.

USER RESPONSE

None.

2523-091 **PTC ACK ignored, no longer in PTC mode.**

EXPLANATION

Received a PTC acknowledgment, but no longer accepting them.

USER RESPONSE

None.

2523-092 **JOIN request rejected, I am not the group leader.**

EXPLANATION

Received a JOIN request, but am not the group leader.

USER RESPONSE

None.

2523-093 **JOIN request rejected, I am currently busy.**

EXPLANATION

Received a JOIN request, but am currently busy handling group changes.

USER RESPONSE

None.

2523-094 **DEATH_IN_FAMILY ignored, not from a group member.**

EXPLANATION

Received a DEATH_IN_FAMILY message, but it wasn't from a group member.

USER RESPONSE

None.

2523-095 **DEATH_IN_FAMILY ignored, I am not the group leader.**

EXPLANATION

Received a DEATH_IN_FAMILY, but I am not the group leader.

USER RESPONSE

None.

2523-096 **DISSOLVE GROUP ignored, not from a group member.**

EXPLANATION

Received a DISSOLVE GROUP message from outside our group.

USER RESPONSE

None.

2523-097 **JOIN time has expired.**

EXPLANATION

Time is up for waiting for a reply to our JOIN request. Wait for another proclaim.

USER RESPONSE

None.

2523-098 **PTC ACK time has expired, committing group.**

EXPLANATION

Time is up for waiting for PTC acknowledgments, commit the group with the members that did acknowledge.

USER RESPONSE

None.

2523-099 **COMMIT BROADCAST ACK time has expired, ACKS missing.**

EXPLANATION

Not everyone acknowledged the commit broadcast message.

USER RESPONSE

None.

2523-100 **COMMIT ACK time has expired, not everyone ACKed.**

EXPLANATION

Everyone in the new group did not respond.

USER RESPONSE

None.

2523-101 **PTC ACK retry count exhausted.**

EXPLANATION

Commit never received.

USER RESPONSE

None.

2523-102 **COMMIT time has expired, re-initializing.**

EXPLANATION

Loss of communication with the leader — go singleton.

USER RESPONSE

None.

2523-103 **Received an unknown message type, message = message_type.**

EXPLANATION

Message received is not recognized.

USER RESPONSE

None.

2523-104 **Malloc failed, size = size.**

EXPLANATION

Not able to satisfy the memory request, indicative of a memory leak.

USER RESPONSE

Start the daemon again.

2523-105 **Group Connectivity Message from outside the configuration, ID=ID, ignored.**

EXPLANATION

Received a group connectivity message from outside the configuration.

USER RESPONSE

None.

2523-106 **Open of css0 device failed.**

EXPLANATION

Unable to open css device.

USER RESPONSE

Check permissions on /dev/css0.

2523-107 **Css Ioctl Failed rc = return_code errno = errno.**

EXPLANATION

The ioctl of the css device failed with the indicated return code and errno value.

USER RESPONSE

Contact the IBM Support Center.

2523-108 **Undefined return from switch ioctl = return_value.**

EXPLANATION

The return value from ioctl of css device was unexpected.

USER RESPONSE

Contact the IBM Support Center.

2523-111 **Sending data to pid=PID would block, complete message not sent.**

EXPLANATION

Client socket not ready for sending.

USER RESPONSE

None.

2523-112 **Sending data to pid=PID interrupted and not restarted.**

EXPLANATION

Interrupted system call.

USER RESPONSE

None.

2523-113 **Sending data to pid=PID caused an IO ERROR.**

EXPLANATION

The client socket class caused an IO error.

USER RESPONSE

None.

2523-114 **Receiving data from pid=PID would block, complete message not received.**

EXPLANATION

The client socket is not read for reading.

USER RESPONSE

None.

2523-115 **Receiving data from pid=PID interrupted and not restarted.**

EXPLANATION

Interrupted system call.

USER RESPONSE

None.

2523-116 **Sending data to pid=PID caused an IO ERROR.**

EXPLANATION

The client socket class caused an IO error.

USER RESPONSE

None.

2523-117 **Client Socket Error ID = error_ID, not recognized.**

EXPLANATION

Unrecognized error from the client socket class.

USER RESPONSE

None.

2523-118 **sendto would block.**

EXPLANATION

The socket is not ready for sending.

USER RESPONSE

None.

2523-119 **recvfrom would block.**

EXPLANATION

Nothing to receive at this point.

USER RESPONSE

None.

2523-120 **recvfrom failure.**

EXPLANATION

Receive failure from the remote socket.

USER RESPONSE

None.

2523-121 **Late in sending Heartbeat by seconds.microseconds seconds.**

EXPLANATION

Heartbeats are sent out based on the time specified by either of two tunable fields: hbrate in the HACMPnim Global ODM or HbInterval in the HACMP topsvcs class, or defaults to 1 second. The daemon checks how much time has elapsed since the last heartbeat was sent. If the difference is greater than 2 times the hbrate/hbInterval tunable, then this message displays how many seconds late it was.

USER RESPONSE

None.

2523-122 **tunable_value, Not a valid value for TS_Frequency; positive integer expected.**

EXPLANATION

The tunable value is encoded as part of the configuration and is specified in the HACMPnim and HACMPtopsvcs Global ODM classes.

USER RESPONSE

Specify a correct value.

2523-123 **tunable_value, Not a valid value for TS_Sensitivity; positive integer expected.**

EXPLANATION

The tunable is encoded as part of the configuration and is and is specified in the HACMPnim and HACMPtopsvcs Global ODM classes.

USER RESPONSE

Specify a correct value.

2523-124 **tunable_value, Not a valid value for TS_FixedPriority; positive integer greater than 37 expected.**

EXPLANATION

The tunable value is encoded as part of the configuration and is specified in the HACMPtopsvcs Global ODM class.

USER RESPONSE

Specify a correct value.

2523-125 **tunable-value, Not a valid value for TS_LogLength; positive integer expected.**

EXPLANATION

Check the HACMPtopsvcs object in the Global ODM and specify a correct value for the tsloglength attribute.

USER RESPONSE

Specify a correct value.

2523-145 **Unable to lock queue.**

EXPLANATION

This messages indicates a condition that should never occur.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-146 **Unable to unlock queue.**

EXPLANATION

This messages indicates a condition that should never occur.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-147 **Unable to increment count of items of queue.**

EXPLANATION

This messages indicates a condition that should never occur.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-148 **Unable to decrement count of items of queue.**

EXPLANATION

This messages indicates a condition that should never occur.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-149 **Unable to initialize thread attributes.**

EXPLANATION

Could not initialize thread attributes. Therefore a new thread to be used to send or receive messages, or to handle client requests, cannot be created.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-150 **Unable to create the send thread for adapter.**

EXPLANATION

Could not create the thread responsible for sending messages.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-151 **Unable to create the receive thread for adapter.**

EXPLANATION

Could not create the thread responsible for receiving messages.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-152 **Incorrect machines list file, Instance Number header not found.**

EXPLANATION

Machines list has incorrect format. The instance number header was not found at the top of the file.

USER RESPONSE

Verify contents of machines.lst file. Start the cluster again. If the problem persists, contact the IBM Support Center.

2523-153 **Incorrect machines list file, configuration ID header not found.**

EXPLANATION

Machines list has incorrect format. The configuration ID header was not found at the top of the file.

USER RESPONSE

Verify contents of machines.lst file. Start the cluster again. If the problem persists, contact the IBM Support Center.

2523-154 **Incorrect machines list file, new ID=new-config-id current ID=current-config-id.**

EXPLANATION

The configuration id in the new machines.lst file generated in a refresh operation is different from that obtained previously.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-155 **Refresh dispatcher process pid hatsctrl/topsvcsctrl-process-pid has not finished.**

EXPLANATION

The process forked or executed by the daemon to rebuild the machines list file to refresh the configuration has not finished yet.

USER RESPONSE

If the problem persists, record the above information and contact the IBM Support Center.

2523-156 **Error errno-value in waitpid() system call.**

EXPLANATION

The waitpid() call issued by the daemon to obtain the exit status of the refresh dispatcher process returned with an error value.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-157 **Refresh dispatcher process terminated with signal signal-number.**

EXPLANATION

The refresh dispatcher process was terminated with the given signal.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-200 **Function not implemented yet.**

EXPLANATION

This messages indicates a condition that should never occur.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-201 **TMIOSTAT ioctl failed**

EXPLANATION

TMIOSTAT ioctl call issued for Target-mode device failed. The daemon will not be able to use the Target-mode device for heartbeating.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-202 **SCSI Adapter Error.**

EXPLANATION

Error related to obtaining status of Target-mode device operation.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-203 **General error in SCSI Card Status.**

EXPLANATION

Error related to obtaining status of Target-mode device operation.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-204 **Failed to obtain SCSI device sense.**

EXPLANATION

Failed to obtain Target-mode device sense data.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-205 **Sleeping too long without getting data.**

EXPLANATION

Not getting data from the Target-mode device.

USER RESPONSE

If problem persists, record the above information and contact the IBM Support Center.

2523-206 **Connection lost.**

EXPLANATION

Error while reading data from non-IP device.

USER RESPONSE

If problem persists, record the above information and contact the IBM Support Center.

2523-207 **Failed to open serial port on device-name.**

EXPLANATION

Cannot open serial port device for heartbeating. The daemon will not be able to use that device for heartbeating.

USER RESPONSE

Check if device actually exists and is of the correct type.

2523-208 **Could not get the flags - fcntl error**

EXPLANATION

Cannot get the flags associated with serial device.

USER RESPONSE

Check if serial device is of the correct type.

2523-209 **fcntl failure - could not set to non–blocking and no delay.**

EXPLANATION

Cannot configure serial device for non–blocking and no delay mode.

USER RESPONSE

Check if serial device is of the correct type. If it is and the problem still occurs, record the above information and contact the IBM Support Center.

2523-210 **Terminal device not OK.**

EXPLANATION

Cannot set current serial device state information

USER RESPONSE

Check if serial device is of the correct type. If it is and the problem still occurs, record the above information and contact the IBM Support Center.

2523-211 **Failed to set attributes for device.**

EXPLANATION

Cannot set serial device characteristics.

USER RESPONSE

Check if serial device is of the correct type. If it is and the problem still occurs, record the above information and contact the IBM Support Center.

2523-212 **Could not add rts to stream stack.**

EXPLANATION

Cannot set serial device to add rts to stream stack.

USER RESPONSE

Check if serial device is of the correct type. If it is and the problem still occurs, record the above information and contact the IBM Support Center.

2523-213 **Failed to flush input and output.**

EXPLANATION

Cannot flush input and output in serial device.

USER RESPONSE

Check if serial device is of the correct type. If it is and the problem still occurs, record the above information and contact the IBM Support Center.

2523-214 **No data read from device device-name.**

EXPLANATION

Not reading any data from given device. The connection may be broken or remote node may not be alive, or the daemon is not running on the remote node.

USER RESPONSE

Check if serial connection is OK. Check if the daemon is running on the remote node.

2523-215 **SCSI port device-name open failed.**

EXPLANATION

Cannot open Target-mode device for heartbeating. The daemon will not be able to use that device for heartbeating.

USER RESPONSE

Check if device actually exists and is of the correct type.

2523-216 **Trying to send to a device not opened.**

EXPLANATION

Trying to send a message to a Target-mode device that has not been opened.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-217 **No elements in the queue.**

EXPLANATION

There are no elements in the queue for a Target-mode device.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-218 **Packet size less than zero.**

EXPLANATION

Error while passing a packet from the Target-mode receive thread to the main thread.

USER RESPONSE

If problem persists, record the above information and contact the IBM Support Center.

2523-219 **Did not write all the bytes. Packet size = packet-size. Bytes sent= bytes-sent, device device-name.**

EXPLANATION

Error while writing to a serial or Target-mode device. Remote daemon may not be running, or connection may be broken.

USER RESPONSE

Check if remote daemon is running and device connection is OK.

2523-220 **Did not write all the bytes. Packet type = packet-type, device device-name.**

EXPLANATION

Error while writing to a serial or Target-mode device. Remote daemon may not be running, or connection may be broken.

USER RESPONSE

Check if remote daemon is running and device connection is OK.

2523-221 **Packet size too long!!. Packet size = packet-size.**

EXPLANATION

Trying to write a packet into a serial or Target-mode device that is larger than the maximum allowed.

USER RESPONSE

Record the above information and contact the IBM Support Center.

2523-222 **Packet check sum incorrect, device device-name.**

EXPLANATION

Received a packet with incorrect checksum through the serial device. May be caused by a noisy connection or another process trying to use the same serial device.

USER RESPONSE

Check the serial connection. Also check if there are other processes trying to write to the same serial device.

2523-223 **Errno errno when writing packet on device-name.**

EXPLANATION

Got given errno code when trying to write a packet to given device name. Connection may be broken or daemon on remote node may not be running.

USER RESPONSE

Check the device connection. Also check if daemon is running on remote node.

2523-224 **Received a Grace request from Adapter-ID with incorrect or missing data: 4 words of data from the auxiliary header.**

EXPLANATION

Details contained in the auxiliary header were not present. The message contains the data that was found.

USER RESPONSE

This message indicates corrupted message data. Record the above information and contact the IBM Support Center.

2523-225 **Cannot find network offset in ADAPTER_CONFIGURATION message. Gid group-id, address 0xaddress.**

EXPLANATION

Cannot find which network offset contains a given address in an ADAPTER_CONFIGURATION message.

USER RESPONSE

If message persists, record the above information and contact the IBM Support Center.

2523-226 **Cannot force boot-to-service transition for Service address**

EXPLANATION

Cannot force boot-to-service transition for boot address boot-addr1 since service address serv-addr is already being used by boot address boot-addr2

USER RESPONSE

There is a problem in the configuration either before or after migration. Force a dump of the Topology Services daemon (kill -6 <pid-of-hatsd-in-HACMP/ES>) and inform IBM Service Organization.

2523-227 **Cannot find adapter number for address when trying to force a boot-to-service transition**

EXPLANATION

Cannot find adapter number for adapter with address boot-addr. This is needed in the process of forcing a boot-to-service transition after a migration-refresh.

USER RESPONSE

There is a problem in the configuration either before or after migration. Force a dump of the Topology Services daemon (kill -6 <pid-of-hatsd-in-HACMP/ES>) and inform IBM Service Organization.

2523-228 **Address is already in use when trying to force a boot-to-service transition for address**

EXPLANATION

Address serv-addr is already in use by another adapter when trying to force a boot-to-service transition after a migration-refresh.

USER RESPONSE

There is a problem in the configuration either before or after migration. Force a dump of the Topology Services daemon (kill -6 <pid-of-hatsd-in-HACMP/ES>) and inform IBM Service Organization.

2523-229 **Cannot find service address when trying to force a boot-to-service transition**

EXPLANATION

After a migration-refresh, current address configured on the adapter does not match any of the possible service addresses for this network.

USER RESPONSE

There is a problem in the configuration either before or after migration. Force a dump of the Topology Services daemon (kill -6 <pid-of-hatsd-in-HACMP/ES>) and inform IBM Service Organization.

2523-230 **Interface name for adapter missing in machines.lst file. Discarding adapter.**

EXPLANATION

The interface name for the adapter was missing in the machines.lst file produced in a refresh.

USER RESPONSE

There may be a problem with either the HACMP/ES cluster manager or the Topology Services script. Inform IBM Service Organization.

2523-231 **Invalid slide of adapter on a node into network: migration-refresh.**

EXPLANATION

The machines.lst produced for a migration refresh has caused the Topology Services daemon to attempt an adapter slide. Adapter slides (adapters moving into a different Topology Services network)are not allowed in a migration-refresh, since adapters cannot be added or removed in a migration-refresh.

USER RESPONSE

There may be a problem with either the HACMP/ES cluster manager or the Topology Services script. Inform IBM Service Organization.

2523-240 **Time-to-Live expired for Daemon Routing message.**

EXPLANATION

There is a routing loop in the Daemon Routing messages. Different nodes have inconsistent views of the network topology.The present daemon routing message is going to be discarded.

USER RESPONSE

If problem persists inform IBM Service Organization.

2523-241 **No route to send Daemon Routing message to node.**

EXPLANATION

No route was found when trying to send a Daemon Routing message to the given node. It may be due to a temporary route outage. Usually Group Services does not even pass a Daemon Routing request to Topology Services if there is no route to the destination node.

USER RESPONSE

If problem persists inform IBM Service Organization.

2523-242 **Low priority queue is full.**

EXPLANATION

The queue for low priority packets has been filled, forcing a low priority packet to be dircarded. Low priority packets should correspond to PRM packets. These will be retried by the PRM client. This message generally indicates a burst of Daemon Routing messages. Keeping too many of these messages in the queue is not useful, since after some point, the time it takes for the message to be transmitted to the destination is more than enough for the client to consider the message lost.

USER RESPONSE

If problem persists inform IBM Service Organization.

2523-243 **Cannot find adapter offset to send Daemon Routing message.**

EXPLANATION

The Daemon Routing logic was unable to find an adapter offset that corresponds to a given destination address. (which is the first hop of the message route.) The problem could be caused by a changing network topology.

USER RESPONSE

If problem persists inform IBM Service Organization.

2523-244 **Invalid number of destination nodes in Daemon Routing**

EXPLANATION

The Topology Services daemon received a request to send a Daemon Routing message to an invalid number of nodes. The request is likely to contain invalid data.

USER RESPONSE

If problem persists inform IBM Service Organization.

2523-245 **Invalid node number in Daemon Routing message**

EXPLANATION

The Topology Services daemon received a request to send a Daemon Routing message to an invalid node number. The request is likely to contain invalid data.

USER RESPONSE

If problem persists inform IBM Service Organization.

2523-246 **Invalid size of PRM message in Daemon Routing message**

EXPLANATION

The Topology Services daemon received a request to send a Daemon Routing message with an invalid number of bytes. The request is likely to contain invalid data.

USER RESPONSE

If problem persists inform IBM Service Organization.

2523-300 **Refresh operation failed because of errors in machines.lst file**

EXPLANATION

The refresh operation failed because the new configuration has a problem. More details can be found in the Topology Services log file. One possible cause for the problem is an IP address that appears to belong to 2 different adapters.

USER RESPONSE

Correct the configuration problem and execute refresh again.

2525 – RS/6000 Cluster Technology Common Messages

2525-000 **Cannot display message from script: missing arguments (argument count).**

EXPLANATION

The named program is used by scripts to display informational or error messages. A script has called the named program with too few arguments. This indicates an error in the script.

USER RESPONSE

Contact the IBM Support Center with the name of the program you were executing at the time this error was displayed.

2525-001 **Cannot display message from script: set name too long (set name).**

EXPLANATION

The named program is used by scripts to display informational or error messages. A script has called the named program with a set name that was too long. This indicates an error in the script.

USER RESPONSE

Contact the IBM Support Center with the name of the program you were executing at the time this error was displayed.

2525-002 **Cannot display message from script: catalog name too long (catalog name).**

EXPLANATION

The named program is used by scripts to display informational or error messages. A script has called the named program with a catalog name that was too long. This indicates an error in the script.

USER RESPONSE

Contact the IBM Support Center with the name of the program you were executing at the time this error was displayed.

2525-003 **Cannot display message from script: message name too long (message name).**

EXPLANATION

The named program is used by scripts to display informational or error messages. A script has called the named program with a message name that was too long. This indicates an error in the script.

USER RESPONSE

Contact the IBM Support Center with the name of the program you were executing at the time this error was displayed.

2525-004 **Cannot display message from script: improperly formed catalog name (catalog name).**

EXPLANATION

The named program is used by scripts to display informational or error messages. A script has called the named program with a catalog name that was not in a proper format. This indicates an error in the script.

USER RESPONSE

Contact the IBM Support Center with the name of the program you were executing at the time this error was displayed.

2525-005 **Cannot display message from script: cannot open message map file message.**

EXPLANATION

The named program is used by scripts to display informational or error messages. The named program could not open the named message map file.

USER RESPONSE

Verify that the named message map file is present and readable. If the file is not present, reinstall the LPP that contains the program you were executing at the time this error was displayed. If this does not resolve the problem, reinstall prerequisite LPPs. If the named message map file is present but not readable, use the chmod command to make it readable. If the problem persists, record the above information and contact the IBM Support Center.

2525-006 **Cannot display message from script: error in message map file message map file name at line line number.**

EXPLANATION

The named program is used by scripts to display informational or error messages. The named program found an error in the named message map file at the indicated line number.

USER RESPONSE

Contact the IBM Support Center with the name of the program you were executing at the time this error was displayed, the name of the message map file and the line number in error.

Appendix H 7x24 Maintenance

The goal of high availability is to keep systems up and running, allowing continuous access to critical applications. In many enterprises it has become necessary to keep applications running seven days a week, 24 hours a day. With proper planning, customizing, and monitoring, an HACMP cluster can provide nearly continuous availability, interrupted only by scheduled, necessary maintenance.

This appendix is a collection of information describing the issues and procedures involved in keeping a cluster running on as close to a 7 X 24 basis as possible.

Overview

Throughout all stages of cluster administration—planning, configuration, maintenance, troubleshooting, and upgrading—here are tasks you can do and systems you can put in place to help ensure your cluster's nearly continuous availability.

Once you have configured the cluster and brought it online, it is very important to do maintenance tasks in as non-disruptive a way as possible. The HACMP cluster is a distributed operating system environment. Therefore maintaining an HACMP cluster requires attention to some issues that have different ramifications in the cluster environment compared to maintaining a single-server system.

Making changes to a cluster must be thoroughly planned, since changes to one component may have cascading effects. Changes on one node affect other nodes, but this is not often apparent until fallover occurs (or cannot occur due to a non-synchronized change to the cluster). Some of the "do's and don'ts" of cluster maintenance are explained in this appendix.

Finally, setting up and following regular preventive maintenance procedures helps alert you to any potential problems before they occur. Then you can take timely action or plan fallovers or cluster downtime at your convenience as necessary to deal with any impending issues.

The appendix is organized as follows:

- **Planning for 7 X 24 Maintenance**. This section reemphasizes the importance of careful planning and customization of the original installation of the cluster.
- **Run-Time Maintenance**. This section offers reminders and tips to help you avoid actions that endanger a stable, running cluster.
- **Hardware Maintenance**. This section contains procedures for changing or replacing certain hardware
- **Practicing Preventive Maintenance**. This sections reviews tools you can use to avoid problems or catch them early.

Planning for 7 X 24 Maintenance

Planning the original installation of your cluster carefully goes a long way toward making cluster maintenance easier. A well-configured and customized cluster is the first step for good preventive maintenance. Proper cluster configuration also makes it less likely you will have to make changes that affect cluster performance while users are accessing their applications.

Planning the cluster starts with a Single Point of Failure analysis. See Volume I, Part 1: Planning HACMP/ES Clusters for a detailed list of issues to consider. Once the cluster is installed and running, you need to handle any failures as quickly and/or as automatically as possible. Planning for run-time failure recovery will help ensure that HACMP for AIX does all that it is capable of doing to keep your critical resources online.

This section includes information on the following topics:

- Customizing the cluster, including setting up error notification to improve monitoring and management of the cluster
- Tuning the communications system—network and nameserving issues
- Planning disk and volume group layout
- General planning for hardware and software maintenance

Customizing the Cluster

Customizing the cluster enhances your ability to monitor the cluster and keep it running. You can define a pre-event, a post-event, and a notification method for every cluster event. Notification of events is critical to maintain service for any HACMP cluster. Although HACMP writes messages to the **hacmp.out** and **cluster.log** log files, it is very useful to include notifications to the console or mail to the system administrator when an event occurs that demands immediate attention.

You can include automatic recovery actions as well as notification in the cluster customization.

Use the HACMP and AIX tools available to customize some or all of the following:

- Hardware error notification
- Hardware failure
- Cluster event notification
- Pre- and post-event recovery actions
- Network failure escalation
- ARP cache refresh
- Application server scripts.

It is highly recommended that you maintain a test cluster as well as your production cluster. Thus before you make any major change to the production cluster, you can test the procedure on the test cluster. HACMP supplies event emulation utilities to aid in testing.

Customizing AIX Error Notification of Hardware Errors

Customizing notification when you configure the cluster is a good preventive measure. See Chapter 8, Cluster Events: Tailoring and Creating and Chapter 13, Tailoring AIX for HACMP/ES for complete information on customizing and setting up notification of cluster events.

See Chapter 13, Tailoring AIX for HACMP/ES for information on supporting AIX Error Notification, and for information on setting up automatic notification for hardware errors that do not cause cluster events.

Using the HACMP Automatic Error Notification SMIT screens, you can turn on automatic error notification for selected hard, non-recoverable error types: disk, disk adapter, and SP switch adapter errors. All disks defined as HACMP resources, and disks in the rootvg and HACMP volume groups and filesystems are included.

You may want to set up error notification for certain media or temporary errors. You may also want to customize the error notification for some devices rather than using one of the two automatic error notification methods.

Chapter 13, Tailoring AIX for HACMP/ES includes an example of how to promote a hardware error to a node failure, using the AIX Error Notification facility.

List of Hardware Errors to Monitor

The following list of hardware errors gives you a good idea of types of errors to monitor. The first list shows which errors are handled by the HACMP automatic error notification utility. The following lists show other types of errors you may want to address. You can decide whether you want to escalate any of these to node failure. For each device monitored, you can determine an additional action other than notification, such as:

- Initiate a graceful takeover
- Initiate a custom recovery action such as reconfiguration for a failed device using an alternative device.

Hardware Errors Handled by HACMP Auto-Error Notification	
DISK_ERR2	Permanent physical disk error (known error)
DISK_ERR3	Permanent physical disk error, adapter detected (known error)
SCSI_ERR1	Permanent SCSI adapter HW error (known error)
SCSI_ERR3	Permanent SCSI adapter microcode error (known error)
SCSI_ERR5	Temporary SCSI bus error
SCSI_ERR7	Permanent unknown system error
SCSI_ERR9	Potential Data loss condition
SDA_ERR1	Adapter hardware error condition
SDA_ERR3	Permanent unknown system error

SDC_ERR1	Controller/DASD link error
SDC_ERR2	Controller hardware error
SSA_HDW_ERR	SSA hardware error condition
SSA_DISK_ERR1	Permananent microcode program error
SSA_DISK_ERR4	Permanent disk operation error
DISK_ARRAY_ERR2	Permanent disk operation error (disk failure)
DISK_ARRAY_ERR3	Permanent disk operation error (disk failure)
DISK_ARRAY_ERR5	Permanent disk operation error (disk failure)
SCSI_ARRAY_ERR2	SCSI hardware error
HPS_FAULT4_ER	SP switch error
TB3_HARDWARE_ER	TB3 type switch hardware error
TB3_MICROCODE_ER	TB3 switch microcode error
TB3_SLIH_ER	

Disk and Adapter Errors Not Covered by HACMP Auto-Error Notification	
SSA_HDW_RECOVERED	Temporary adapter error
SSA_DISK_ERR3	Temporary disk operation error
SSA_DEGRADED_ERROR	Adapter performance degraded
SSA_LOGGING _ERROR	Permanent unable to log an error against a disk
SSA_LINK_OPEN	Permanent adapter detected open serial link
SSA_SOFTWARE_ERROR	Permanent software program error
SSA_LINK_ERROR	Temporary link error
SSA_DETECTED_ERROR	Permanent loss of redundant power/cooling
LVM_MISSPVADDED	PV defined as missing (unknown error)
LVM_SA_WRT	PV defined as missing (unknown error)
LVM_SA_PVMISS	Failed to write VGSA (unknown error)

Disk Array Errors not Covered by HACMP Auto-Error Notification	
DISK_ARRAY_ERR4	Temporary disk operation error (disk media failing)
DISK_ARRAY_ERR6	Permanent array subsystem degradation (disk media failure)

DISK_ARRAY_ERR7	Permanent array subsystem degradation (controller)
DISK_ARRAY_ERR8	Permanent array active controller switch (controller)
DISK_ARRAY_ERR9	Permanent array controller switch failure

Failed 64-port Adapter (tty device driver)	
COM_PERM_PIO	PIO exception, possible adapter failure

You may have additional devices critical to your operation that are not supported by HACMP for AIX. For example, you may have X.25 components that need monitoring. You can set up AIX error notification to monitor X.25 microcode errors or adapter time-outs.

Customizing Cluster Events

Customizing cluster events to send notification or to take recovery actions is another method you can use to help maintain the cluster running as smoothly as possible.

See Chapter 8, Cluster Events: Tailoring and Creating and Chapter 13, Tailoring AIX for HACMP/ES for complete information on customizing and setting up notification of cluster events.

See the section Sample Custom Scripts on page 18-44 for tips on writing scripts to make **cron** jobs and print queues highly available.

Customizing Application Server Scripts

See Appendix C, Applications and HACMP for tips on handling applications.

Some key things to keep in mind:

- Define an HACMP application server for each node that supports applications requiring recovery.
- Applications must be started up and shutdown in an orderly fashion. Some situations exist where the timing and control of starting and stopping applications needs to be handled based on pre/post event process. You may need to take into account the order in which applications assigned to the same node are started.
- Check for dependencies between nodes. For example, a process on node1 cannot start until a process that runs on node2 is up. Include a check for remote node/application availability before issuing the local startup command.
- You may need to perform some checks to make sure the application is not running and to clean up logs or roll back files before starting the application process.

See the section Sample Custom Scripts on page 18-44 for tips on writing scripts to make **cron** jobs and print queues highly available.

Application Monitoring

If you are running HACMP/ES 4.4, you can monitor a set of applications that you define through the SMIT interface. The monitoring is done in one of two ways:

Process application monitoring detects the death of one or more processes using RSCT Event Management. *User defined application monitoring* checks the health of an application at user-specified polling intervals.

In either case, when a problem is detected, HACMP/ES attempts to restart the application, and continues up to a specified retry count. You choose one of the following responses for HACMP/ES to take when an application cannot be restarted within the retry count:

- The **fallover** option causes the resource group containing the application to fall over to the node with the next-highest priority according to the resource policy.
- The **notify** option causes HACMP to generate a server_down event, similar to a network_down event, to inform the cluster of the failure.

You can customize the 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.

See Chapter 18, Configuring an HACMP/ES Cluster and Chapter 21, Monitoring an HACMP/ES Cluster, for complete information on application monitoring.

Network Configuration and Nameserving

Setting up and maintaining clear communication paths for the Cluster Manager is a key element for efficient cluster operation.

Setting up Serial Networks

It is crucial to have at least one serial network configured for the cluster. Without a serial network, you run the risk of a partitioned cluster if TCP/IP networks fail, since the nodes will be unable to maintain heartbeat communication.

If there is a problem with the serial network (or there is none configured and defined) you may also see the DGSP message in the AIX error log. (DGSP = Diagnostic Group Shutdown Partition.) This message is sent if a takeover is initiated after a node_down, and the downed node revives and sends heartbeats. The takeover nodes kill the reviving node to ensure non-contention over resource ownership within the cluster. This prevents that node from being able to rejoin the cluster normally.

Integrating HACMP with Network Services

HACMP requires IP address to name resolution. The three most commonly used methods include:

- Domain Name Service
- Network Information Service
- Flat file name resolution **(/etc/hosts**)

By default, a name request will look first for the DNS (**/etc/resolv.conf**), second for NIS, and last for **/etc/hosts** to resolve the name. Since DNS and NIS both require certain hosts as designated servers, it is necessary to maintain the **/etc/hosts** file in case the DNS or NIS name server is unavailable, and to identify hosts that are not known to the name server. It is required to have all cluster adapter IP labels in all cluster nodes' **/etc/hosts** tables.

To ensure the most rapid name resolution of cluster nodes, it is recommended that you change the default order for name serving so that **/etc/hosts** is used first (at least for cluster nodes).

To do this, edit the **/etc/netsvc.conf** file so that this line appears as follows:

```
Hosts=local,nis
```

Putting the local option first tells the system to use **/etc/hosts** first, then NIS.

You can also change the order for name resolution by changing the environment variable NSORDER so it looks like this:

```
NSORDER=local,bind
```

Note: If you are using NIS, it is recommended to have the NIS master server outside the cluster, and have the cluster nodes run as NIS slave servers.

If you are using DNS or NIS, this fact must be configured in the run-time parameters of all nodes in the cluster. The Cluster Manager then knows to stop and restart the **named** when it is changing adapter IP addresses (in an adapter_swap event).

See Using HACMP/ES with NIS and DNS on page 4-14 for more information on possible NIS and DNS issues such as changing a host name. See Chapter 13, Tailoring AIX for HACMP/ES for information on editing the **/etc/hosts** file, and also for notes on NIS and **cron** considerations.

Warning: You cannot use DHCP to allocate IP addresses to HACMP cluster nodes. Clients may use this method, but cluster nodes cannot.

Tuning Networks for Best Performance

HACMP 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.

See Planning for Cluster Performance on page 4-17 for information on setting AIX and HACMP parameters for best performance.

If you are running a cluster on an SP, also consult your SP manual set for instructions on tuning SP switch networks.

Planning Disks and Volume Groups

Planning the disk layout is crucial for the protection of your critical data in an HACMP cluster. Follow the guidelines carefully, and keep in mind these issues:

- All operating system files should reside in the root volume group (**rootvg**) and all user data should reside outside that group. This makes updating or reinstalling the operating system and backing up data more manageable.
- A node whose resources are not designed to be taken over should not own critical volume groups.
- When using copies, each physical volume using a mirror copy should get its power from a UPS system.
- Volume groups that contain at least three physical volumes provide the maximum availability when implementing mirroring (one mirrored copy for each physical volume).
- **auto-varyon** must be set to **false**. HACMP will be managing the disks and varying them on and off as needed to handle cluster events.

Quorum Issues

Setting up quorum correctly when laying out a volume group is very important. Quorum *must* be enabled on concurrent volume groups. With quorum enabled, a two-disk non-concurrent volume group puts you at risk for losing quorum and data access. The failure of a single adapter or cable would cause half the disks to be inaccessible.

Either build three-disk volume groups or disable quorum on non-concurrent volume groups. You can also use a quorum buster disk that contains no critical data to ensure maintenance of quorum.

Maintaining volume groups, logical volumes, and filesystems shared by cluster nodes require that you keep the ODM definitions of these components synchronized across all cluster nodes. It is recommended to use the C-SPOC commands HACMP provides for maintaining these components, thus updating configuration changes from one node to all the others.

See the detailed section Quorum on page 6-5.

Planning Hardware Maintenance

Good maintenance practice in general dictates that you

- Check cluster power supplies periodically.
- Check the **errlog** and/or any other logs where you have redirected information of interest and attend to all notifications in a timely manner.
- Be prepared to replace any failed or outdated cluster hardware.

If possible, you should have replacement parts readily available. If the cluster has no single points of failure, it will continue to function even though a part has failed. However, now a single point of failure may exist. If you have set up notification for hardware errors, you have an early warning system in place.

This Guide contains procedures detailing how to replace the following cluster components while keeping the cluster running:

- Network
- Adapter
- Disk
- Node.

See Hardware Maintenance on page H-14 for more information.

Planning Software Maintenance

Planning for software maintenance includes:

- Customizing notification of software problems
- Periodically checking and cleaning up log files
- Taking cluster snapshots when making any change to the cluster configuration
- Preparing for upgrading AIX, applications, and HACMP for AIX.

See Preventive Maintenance on page H-19 for more details.

Run-Time Maintenance

Once you have configured the cluster and brought it online, it is very important to do maintenance tasks in as non-disruptive a way as possible. Maintaining an HACMP cluster requires attention to some issues that have different ramifications in the cluster environment compared to maintaining a single system.

This section discusses the following issues:

- Tasks that require stopping the cluster
- Warnings about the cascading effects caused by making certain types of changes to a stable, running cluster

Tasks that Require Stopping the Cluster

HACMP allows you to do many tasks without stopping the cluster; you can do many tasks dynamically using the DARE and C-SPOC utilities. However, in order to do the following tasks, you must stop the cluster:

- Change the name of a cluster component: network module, cluster node, network adapter, application server, or resource group. Once you configure the cluster, you should not need to change these names,
- Change the cluster ID. This is also a rare occurrence.
- Maintain RSCT on clusters running HACMP/ES. HACMP/ES does not support the use of a forced down.

- Change auto error notification.
- Change SSA fence registers.

Changing the Cluster Configuration—Cascading Effects on Cluster Behavior

Installing HACMP makes changes to several AIX files (see Chapter 19, Maintaining an HACMP/ES Cluster). All the components of the cluster are under HACMP control once you configure, synchronize, and run the cluster software. Using AIX to change any cluster component, instead of using the HACMP menus and synchronizing the topology and/or the cluster resources, will interfere with the proper behavior of the HACMP cluster software and thus affect critical cluster services.

This section contains warnings about actions that will endanger the proper behavior of an HACMP cluster. It also includes some reminders about proper maintenance procedures.

Stopping and Starting Cluster Services

Do not start or stop daemons or services that are running under the control of HACMP. Any such action will affect cluster communication and behavior. You can choose to run certain daemons (Clinfo) but others are required to run under HACMP control.

Most important, never use the **kill – 9** command to stop the Cluster Manager or any RSCT daemons. This causes an abnormal exit. SRC will run the **clexit.rc** script and halt the system immediately. This causes the other nodes to initiate a fallover.

TCP/IP services are required for cluster communication. Do not stop this service on a cluster node. If you need to stop HACMP or TCP/IP to maintain a node, use the proper procedure to move the node’s resources to another cluster node, then bring down the node. Follow the instructions inChapter 24, Changing the Cluster Configuration to make changes to cluster topology or resources.

Node, Network and Network Adapter Issues

The HACMP configuration of the nodes and IP addresses of network adapters is crucial to the communication system of the cluster. Any change in the definitions of these elements must be updated in the cluster configuration and resynchronized.

Do not change the configuration of a cluster node, network, or adapter using AIX SMIT menus or commands, individually on a cluster node, outside of HACMP. See Chapter 24, Changing the Cluster Configuration, for instructions on changing the configuration dynamically following the proper HACMP cluster procedures.

Do not start or stop daemons or services that are running under the control of HACMP. Any such action will affect cluster communication and behavior.

Be sure to follow proper procedures for the following types of changes:

- Changing boot, service, or standby IP address of any adapter defined to HACMP. Changes to IP addresses must be updated in the HACMP cluster definition and the cluster must then be resynchronized. Any change to network adapter attributes normally requires stopping cluster services, making the change, and restarting cluster services.

Note that in some circumstances you can use the HACMP facility to swap a network service or boot adapter IP address dynamically, to an active standby adapter on the same node and network, without shutting down cluster services on the node. See Swapping a Network Adapter Dynamically on page 24-8 for more information.

- Changing netmasks of network adapters. Service and standby adapters on the same network must have the same netmask on all cluster nodes. Changes made outside the cluster definition will affect the ability of the Cluster Manager to send heartbeat messages across the network.
- Enabling an alternate Hardware Address for a service adapter using AIX SMIT.
- Taking down adapters: Do not take down all adapters on the same network if network_down is set up to perform a graceful down with takeover. If the cluster is customized to perform a graceful down with takeover when all communications on a specific network fail, and you take down all adapters, this will force the graceful down with takeover event to occur whether you intend it or not.
- Taking down adapters: Do not bring all adapters down on the same network if there is only one network and no serial point-to-point network is defined. Doing this will cause system contention between cluster nodes and fallover attempts made by each node. The cluster will remain unstable until you reboot all cluster nodes on an HACMP/ES system (no forced down). On an HACMP system, you must force down all the nodes and then bring them back up. In either case, the cluster is unavailable while you fix the situation.

Making Changes to Network Interfaces

In some circumstances, you can use the HACMP facility to swap a network service or boot adapter IP address dynamically, to an active standby adapter on the same node and network, without shutting down cluster services on the node. See Swapping a Network Adapter Dynamically on page 24-8 for more information.

You must normally stop the cluster to make any change to network interfaces. If you must change the IP address of an adapter, or if you change the adapter label, make sure to make the changes to both DNS or NIS *and* the **/etc/hosts** file. If DNS or NIS and **/etc/hosts** are not updated, you will be unable to synchronize the cluster nodes or do any DARE operations. If DNS or NIS services are interrupted, the **/etc/hosts** file is used for name resolution. You must also fix the **/.rhosts** file or redo **cl_setup kerberos** if you are using enhanced security.

Handling Network Load/Error rates

Dropped packets due to network loads may cause false fallovers. Also, high throughput may cause the deadman switch to time-out. If either of these conditions occurs, check the AIX network options and the Failure Detection Rate you have set for the cluster. These parameters are contained in the Advanced Performance Tuning option of SMIT.

See Changing the Configuration of a Network Module on page 24-11 for information on tuning the Failure Detection Rate.

HACMP/ES 4.4 has new RSCT logging to help with the tuning of networks. See Chapter 18, Configuring an HACMP/ES Cluster.

Network Maintenance/Reconfiguration

Moving Ethernet ports on a running cluster will result in adapter swap or node failure. Even a brief outage will result in a cluster event.

Shared Disk, Volume Group, and Filesystem Issues

Do not change the configuration of a shared volume group or filesystem using AIX, outside of HACMP. Any such action will affect cluster behavior. The Cluster Manager and the cluster event scripts assume the shared volume groups and filesystems are under HACMP control. If you change the environment, the event scripts will not be able to complete properly and you will get unexpected results. HACMP is very sensitive to filesystem mounting and the way volume groups are varied on. (HACMP is very sensitive to all configuration information.)

Disk Issues

Disks should always be mirrored (or use a disk array), to protect against loss of data. Once they are defined and configured within the HACMP cluster, you should always use the HACMP C-SPOC utility (`smitty cl_admin`) to add or remove disks from a volume group with the cluster running. The cluster needs to be made aware of disks being added to or removed from a shared volume group. If you add or remove disks using the conventional method, the cluster will not be aware that these changes have occurred.

Volume Group and Filesystems Issues

Use the C-SPOC utility (`smitty cl_admin`) for common maintenance tasks like creating, extending, changing, or removing a shared filesystem. See Chapter 5, Planning Shared Disk Devices, in this Guide.

See Chapter 28, Additional Tasks: NFS and Run-Time Parameters for information on NFS and HACMP.

- Do not set filesystems to automount; HACMP handles the mounts at startup and during cluster events.
- Do not set volume groups to autovaryon; HACMP executes the varying on and off as needed.
- If you are testing something when the cluster is not running and you varyon a volume group or mount a filesystem, remember to unmount the filesystem and vary off the volume group before you start HACMP.
- Do not have any processes running that would point to a shared filesystem when a graceful down occurs on the node that currently owns that filesystem. If a graceful down occurs and the application stop script fails to terminate the processes that are using the filesystem, that filesystem will be unable to unmount and the fallover will not occur. The cluster will go into a “reconfiguration too long” condition.

One of the more common reasons for a filesystem to fail being unmounted during a graceful down is because the filesystem is busy. In order to successfully unmount a filesystem, no processes or users can be accessing it at the time. If a user or process is holding it, the filesystem will be “busy” and will not unmount.

This is easy to overlook when you write stop application scripts. The script to stop an application should also include a check to make sure that the shared filesystems are not in use. You can do this by using the **fuser** command. The script should use the **fuser** command to see what processes or users are accessing the filesystems in question. The pids of these processes can then be acquired and killed. This will free the filesystem so it can be unmounted.

Refer to the AIX man pages for complete information on this command.

Expanding Filesystems

Use C-SPOC to increase the size of a filesystem.

- Log in as root.
- Enter `smitty cl_fs`
- Select the option to change a cluster filesystem
- Select the filesystem to change
- Enter the new size for the filesystem.
- Synchronize the new definition to all cluster nodes via Sync Cluster Resources.

General Filesystems Issues

The following are some **/etc/filesystem** concerns:

- Changes are made to stanzas in various files during startups, fallover, and other cluster events. You can set up a **crontab** job to monitor filesystem size to help avoid trouble (for example, the **hacmp.out** file can get quite large).
- Shared filesystems must have the *mount* option set to false, so that HACMP can mount and unmount them as needed to handle cluster events.
- Full filesystems in the root volume group may cause cluster events to fail. You should monitor this volume group and clean it up periodically.

Also consider these **NFS Filesystem** concerns:

HACMP 4.4 includes improvements to the way NFS filesystems are handled. See *Chapter 28, Additional Tasks: NFS and Run-Time Parameters.*

Application Issues

Appendix C, Applications and HACMP gives many pointers on planning and maintaining applications in the HACMP environment. You can monitor applications if you are running HACMP/ES 4.4. Some key points to remember:

- Application maintenance will require downtime for resource groups due to binaries residing on a shared disk.
- Upgrades should be tested prior to implementation to anticipate effects on the production cluster.
- Changes to application start and stop procedures should be thoroughly tested prior to going into production.
- Do not have shared applications already running when you start the cluster. A second attempt at starting already running applications may cause a problem.
- Do not manually execute the application stop script for any reason on a running cluster without starting the application back up again. Problems may occur if an attempt is made to stop the application that is already down. This could potentially cause a fallover attempt to be unsuccessful.

Hardware Maintenance

Hardware failures must be dealt with promptly to maintain no single point of failure in the cluster. If you have carefully set up error notification and event customization as recommended, you receive quick notification via email of any problems. You should also periodically do error log analysis. See Viewing HACMP/ES Cluster Log Files on page 29-1 for details on error log analysis.

Some issues to be aware of in a high availability environment include:

- Shared disks connect to both systems, thus open loops and failed disks can result in fragmented SSA loops and the loss of access to one mirror set.
- Set up mirroring so that the mirrored disk copy is accessible by a different controller. This will prevent loss of data access in the event of a disk controller failure. In the event that a disk controller fails, the mirror disk will be accessible through the other controller.

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.

Replacing Topology Hardware

Nodes, networks, and network adapters comprise the topology hardware. Changes to the cluster topology often involves down time on one or more nodes if changes to cabling or adding/removing adapters is involved. In most situations, you can use the DARE utilities to add a topology resource without down time.

The following sections indicate the conditions under which you can use DARE and the conditions under which you must plan cluster downtime.

See Chapter 18, Configuring an HACMP/ES Cluster for full documentation on all procedures.

Replacing Nodes

Using the DARE utility, you can add or remove a node while the cluster is running

Replacing a Node or Node Component

If you are replacing a cluster node keep this list in mind:

- The new node must typically have the same amount of RAM (or more) as the original cluster node
- The new node must typically be same type of system if your applications are optimized for a particular processor
- The new node's slot capacity typically must be the same or better than the old node
- Adapter placement is important – use the same slots as originally assigned

- If you have a concurrent environment, you must reinstall the CRM software. This is also a consideration if you are cloning nodes.
- Get the new license key from the application vendor for the new CPU ID if necessary.

If you are replacing a component of the node:

- Be aware of CPU ID issues
- For SCSI adapter replacement – reset external bus SCSI ID to original SCSI ID
- Adapter replacement – use the same slots as originally assigned.

Procedure for Adding or Removing a Node

Also see Changing the Configuration of Cluster Nodes on page 24-4 for more complete information.

The basic procedure is outlined here:

- Install AIX, HACMP and LPPs on new node
- Connect networks and SSA cabling and test
- Configure TCP/IP
- Import volume group definitions
- Connect serial network and test
- Change HACMP ODM configuration
- Synchronize and verify.

Replacing Networks and Network Adapters

HACMP does not provide highly available networks. You can only protect your applications from downtime due to a network failure if you configure more than one IP network. You should also have a serial network. If no backup network is configured, the cluster will be inaccessible to all but directly connected clients.

You can replace network cabling without taking HACMP off line. You can also replace hubs, routers, and bridges while HACMP is running. Be sure to use the correct IP addresses when reconfiguring a router.

You can use the DARE swap_adapter function to swap the IP address from a boot or service adapter to an active standby adapter on the same node and network. Then you can service the failed adapter without stopping the node.

Procedure for Replacing a LAN adapter

If the hardware supports hot-pluggable network adapters, no cluster downtime is required for this procedure.

A general procedure, if you cannot use the swap_adapter function:

- Move resource groups to another node using DARE
- Stop and detach the device driver for the failed adapter
- Remove the device from the ODM
- Shut down the node containing the failed adapter
- Boot into Diagnostics and test the failed adapter – the problem may be a faulty cable, for example.

- Replace the adapter if necessary (reconnecting drop cables)
- Reboot in normal mode (**cfgmgr** runs)
- Start the device driver for the adapter
- Assign IP addresses and netmasks (using **smitty chinet)** for interfaces if they were undefined
- Test IP communications.

Handling Disk Failures

Handling shared disk failures differs depending on the type of disk and whether it is a concurrent access configuration or not.

- SCSI non-RAID - you will have to shut down the nodes sharing the disks
- SCSI RAID - perhaps no downtime, depends on the capabilities of the array
- SSA non-RAID – requires manual intervention – you can replace disks with no system downtime
- SSA RAID – no downtime necessary.

See Replacing a Failed Drive in a Concurrent Access Volume Group on page 23-11 for information on that procedure.

Replacing a Failed SSA non-RAID Disk

Note: AIX 4.3.3 has a new utility, **replacepv**. See the AIX manpage to see if you can use this utility to replace your failed disk.

This section describes the process of replacing a mirrored disk drive, using C-SPOC commands. The sample cluster is set up as follows:

- Two nodes, NodeB and NodeA
- Two cascading resource groups, each node has top priority for one of the resource groups.
  - NodeArg: NodeA, NodeB
  - NodeBrg: NodeB, NodeA.
- One shared Volume Group: vg1 (includes hdisk1 and hdisk2)
  - vg1 is included in the resource group NodeArg
  - vg1 has one mirrored logical volume (lv01) and one filesystem: (/fs1)
- The logical volume is mirrored on the two disks in the volume group:

```
>lslv -m lv01
lv01:/fs1
LP    PP1  PV1               PP2  PV2               PP3  PV3
0001  0111 hdisk2            0056 hdisk1
```

- Initial disk configuration on both nodes:

```
hdisk#                pdisk#          vg defined?
hdisk1                pdisk0          vg1
hdisk2                pdisk1          vg1
hdisk3                pdisk2          None
```

When HACMP starts on the cluster, NodeA varies on vg1

```
>lsvg -o
   NodeB:
   NodeA:   vg1
```

Procedure to Replace Failed Disk

1. Unmirror a shared vg

   a. From NodeB, execute:

   ```
   smitty cl_admin / Cluster LVM/ Shared vg / Unmirror a shared vg
   ```

   b. Select: `NodeArg    vg1`

   the choices:

   ```
   None
   NodeA   hdisk1
   NodeA   hdisk2
   ```

   c. Select:

   ```
      NodeA hdisk2
   ```

   d. lslv on NodeA shows logical volume is no longer mirrored:

   ```
   >lslv -m lv01
   lv01:/fs1
   LP    PP1  PV1               PP2  PV2               PP3  PV3
   0001  0056 hdisk1
   ```

2. Remove a physical volume from a shared vg

   a. From NodeB execute:

```
smitty cl_admin /Cluster LVM/Shared vg /set vg chars/remove a pv from vg
```

   b. Select: NodeArg   vg1

   the choices:

   ```
   NodeA   hdisk1
   NodeA   hdisk2
   ```

   c. Select:

   ```
   NodeA hdisk2
   ```

   d. Force deallocation of physical partitions = yes

   e. On both nodes, lspv shows hdisk2 is removed from vg1

   ```
     hdisk1   vg1
     hdisk2   none
     hdisk3   none
   ```

3. Remove a disk from the cluster

   a. From NodeB execute:

   ```
   smitty cl_admin / Cluster Physical VM/ remove a disk from the cluster
   ```

   b. Select:

   ```
     NodeA
     NodeB
   ```

   the choices:

   ```
   0cc6
   0578
   ```

   c. Select:

   ```
   0578   (hdisk2)
   ```

d. Do not keep definition in database

e. On both nodes, lspv shows hdisk1 is removed.

```
hdisk1   vg1
hdisk3   none
```

f. On both nodes, ssaxlate shows pdisk1 is still present and unpaired:

```
hdisk1: pdisk0
hdisk3: pdisk2

pdisk0: hdisk1
pdisk1:
pdisk2: hdisk3
```

4. Pull failed disk and replace with a new disk.
5. Add a disk to the cluster:

a. From NodeB, execute:

```
smitty cl_admin / Cluster Physical VM/ add a disk to the cluster
```

b. Select:

```
   NodeA
   NodeB
```

c. Select:

```
  hdisk   ssar   SSA logical disk drive
connection address: (from F4)
   0004ac7c036500d
```

d. On both nodes, lspv shows new disk added as hdisk2

```
hdisk1   vg1
hdisk2   none
hdisk3   none
```

e. On both nodes, ssaxlate shows hdisk2 is unpaired with a pdisk.

```
hdisk1: pdisk0
hdisk2:
hdisk3: pdisk2

pdisk0: hdisk1
pdisk1:
pdisk2: hdisk3
```

6. Add a disk to a shared vg

a. Execute:

```
smitty cl_admin /Cluster LVM/shared vg/set chars of vg/add pv to vg
```

b. select:

```
   NodeArg   vg1
```

the choices:

```
   NodeA hdisk2
   NodeA hdisk3
```

c. select:

```
   NodeA hdisk2
```

d. All data on physical volume will be destroyed. Continue? yes

e. On both nodes, lspv shows hdisk2 added to vg1

```
hdisk1   vg1
hdisk2   vg1
hdisk3   none
```

7. Mirror a shared vg:

a. Execute:

```
smitty cl_admin / Cluster LVM/ shared vg/ mirror a shared vg:
```

b. select:

```
  NodeArg  vg1
```

the choice:

```
  None
  NodeA hdisk1
  NodeA hdisk2
```

c. select:

```
  NodeA hdisk2
```

d. On NodeA (where vg is varied on) lslv shows lv is mirrored

```
lslv -m lv01
lv01:/fs1
LP    PP1  PV1               PP2  PV2               PP3  PV3
0001  0056 hdisk1            0056 hdisk2
```

Preventive Maintenance

It is highly recommended that you maintain a test cluster as well as your production cluster. Thus before you make any major change to the production cluster, you can test the procedure on the test cluster.

HACMP also supplies event emulation utilities to aid in testing.

Cluster Snapshots

It is highly recommended to take periodic snapshots of the cluster in case you need to reapply a configuration. You should take a snapshot any time you change the configuration. Keep a copy of the snapshot on another system, off the cluster, as protection against loss of the cluster. You can use the snapshot to rebuild the cluster quickly in case of an emergency. See Chapter 26, Saving and Restoring Cluster Configurations in this Guide for complete information on cluster snapshots. You might want to consider setting up a **cron** job to do this on a regular basis.

Backups

HACMP does not provide tools for backing up the system. You should plan for periodic backups just as you do for a single system. You should do backups of **rootvg** and shared volume groups.

Backups of shared volume groups should be done more frequently.

Some applications have their own online backup methods.

You can use any of the following:

- **mksysb** backups
- Online backups (**sysback, splitlvcopy**)

Using mksysb

You should do a **mksysb** on each node prior to and following any changes to the node environment. Such changes include:

- Applying PTFs
- Upgrading AIX or HACMP software
- Adding new applications
- Adding new device drivers
- Changing TCP/IP configuration
- Changing cluster topology or resources
- Changing LVM components of **rootvg** (paging space, filesystem sizes)
- Changing AIX parameters (including the tuning parameters: I/O pacing, **syncd**)

Shared Volume Group Backup—Using splitlvcopy

You can use the **splitlvcopy** method on raw logical volumes and filesystems to do a backup while the application is still running. This method is only possible for LVM mirrored logical volumes.

By taking advantage of the LVM's mirroring capability, you can stop the application briefly to split off a copy of the data using the AIX **splitlvcopy** command. Stopping the application gives the application its checkpoint. Then restart the application so it continues processing while you do a backup of the copy.

You can do the backup using **tar**, **cpio**, or any other AIX backup command that operates on a logical volume or a filesystem. Using **cron**, you can automate this type of backup.

Using cron

Use the AIX **cron** utility to automate scheduled maintenance and to monitor the system.

Using cron To Automate Maintenance Of Log Files

Use this utility to automate some of the administrative functions that need to be done on a regular basis. Some of the HACMP log files need **cron** jobs to ensure that they do not use up too much space.

Use **crontab –e** to edit **/var/spool/cron/crontabs/root**.

Cron will recognize the change without need for rebooting.

You might establish a policy for each log, depending how long you want to keep the log, and what size you will allow it to grow. **hacmp.out** is already set to expire every seven days.

The RSCT logs are stored in the **/var/ha/log** directory. These logs are trimmed regularly. If you want to save information for a longer period of time you can either redirect the logging to a different directory, or change the maximum size file parameter (using SMIT). See HACMP/ES Log Files on page 21-38.

Using cron To Set Up An Early Warning System

You can set up **clverify** to run as a **cron** job, sending email notification to the system administrator if **clverify** fails. The output from **clverify** then serves to notify where you might have configuration problems.

Use **cron** to set up jobs to proactively check out the system.

- Run a custom verification daily and send a report to the system administrator
- Check for full filesystems (and take action if necessary)
- Check that certain processes are running
- Run event emulation and send a report to the system administrator

Do Regular Testing

It is recommended to regularly schedule a testing window where a failure is conducted in a controlled environment. That way you can evaluate a fallover before anything happens in your production cluster and people are anxiously awaiting the fix. It should include fallovers of all nodes and full verification of tested protected apps. This is strongly encouraged if you are changing or evolving your cluster environment.

Upgrading Software (both AIX and HACMP)

Take a cluster snapshot and save it somewhere off the cluster.

Back up the operating system and data before performing any upgrade. Prepare a backout plan in case you encounter problems with the upgrade.

Whenever possible, plan and do an initial run through on a test cluster.

AIX patches need to be applied according to the HACMP operations guide:

- Apply patch to standby node.
- Fallover (graceful shutdown with takeover) to standby machine.
- Apply patch to primary node.

Follow this same general rule for patches to the application; follow specific instructions for the application.

See Chapter 14, Installing the HACMP/ES Software for general installation procedures and for instructions on node-by-node migration from HACMP 4.3.1 or HACMP 4.4.to HACMP/ES 4.4. See Chapter 15, Upgrading an HACMP/ES Cluster for instructions on upgrading to HACMP/ES 4.4.

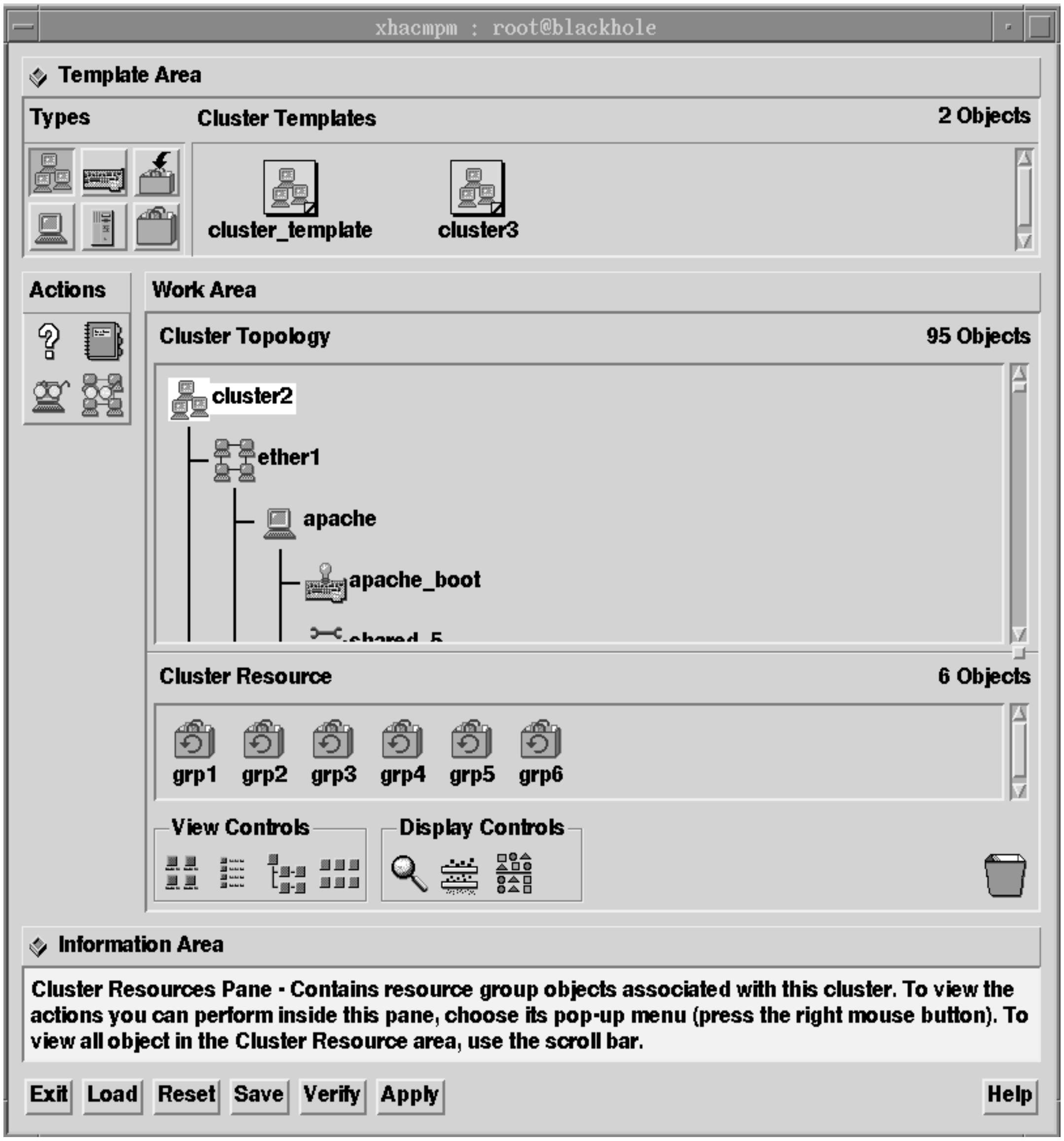
Appendix I VSM Graphical Configuration Application

This appendix describes the **xhacmpm** graphical interface tool, found in the samples directory, for configuring cluster environments and components.

Overview

Through the graphical interface **xhacmpm,** the HACMP for AIX Visual System Management (VSM) application, you can perform many configuration tasks. Using **xhacmpm**, a Common Desktop Environment (CDE) application, you can define attributes for clusters and cluster objects like nodes, resource groups, or application servers, and simultaneously and dynamically configure your cluster environment. You also can save snapshots of cluster configurations and restore them when needed. By clicking, or by dragging and dropping objects onto other cluster objects (or into a Work Area pane) in VSM, you can configure cluster components and the cluster environment.

For example, to configure a cluster, click on the cluster icon in the Types window and then drag and drop a cluster_template into the Cluster Topology pane. A dialogue window appears to let you define the cluster name and ID. Drag and drop to add nodes, adapters, resource groups, and other components to the cluster. The following figure illustrates a cluster configuration represented in the **xhacmpm** window.



Cluster Resources Pane - Contains resource group objects associated with this cluster. To view the actions you can perform inside this pane, choose its pop-up menu (press the right mouse button). To view all object in the Cluster Resource area, use the scroll bar.

The HACMP for AIX xhacmpm Application

Starting xhacmpm

To start **xhacmpm**, enter the following command:

```
/usr/es/lpp/cluster/samples/xhacmpm &
```

When started, **xhacmpm** displays its default configuration. If you want to cancel a configuration you are working on but have not applied, click the Reset button to return **xhacmpm** to its startup display, canceling all changes. To save a snapshot of the current configuration, click the Save button. You can restore the snapshot by clicking the Load button and by specifying a PATH and file name in the dialogue window.

The xhacmpm Window

The **xhacmpm** window contains several distinct areas:

Template Area

In the Template Area, at the top of the window, you define characteristics of cluster objects.

The **xhacmpm** application includes templates for:

- Clusters
- Nodes
- Adapters
- Application servers
- Resources
- Resource groups.

Each template defines characteristics of a cluster object. For example, the cluster template allows you to specify the cluster name and ID. To define the characteristics of a cluster object, drag its icon from the Template Area into the Work Area.

Work Area

In the Work Area, the middle section of the window, you create a cluster configuration by combining cluster object templates. For example, to configure a cluster to include nodes, drag and drop a node_template onto a cluster object in the Work Area’s Cluster Topology pane. The **xhacmpm** application displays a dialog box where you can specify each node’s attributes. After configuring a node object, you can keep a permanent record of its definition by dragging the icon up into the Node Templates pane in the Template Area. The node object becomes a template object with the same name.

Information Area

The Information Area at the bottom of the window displays brief descriptions of whatever part of the **xhacmpm** interface is under the mouse pointer. For example, when you move the pointer over the wastebasket icon in the lower-right corner of the window, the Information Area displays a description of how to use it.

The **xhacmpm** interface includes other buttons that allow you to customize the view of the cluster you are configuring, search for particular cluster objects, and determine the current values of their characteristics.

Using xhacmpm Online Help

For more information about using **xhacmpm** to configure an HACMP cluster, start the application and view the online help that comes with it. To access help, drag the question mark (?) Action icon onto any icon representing a cluster object and **xhacmpm** displays a window providing information about the object and any related configuration tasks.

You must perform the cluster configuration steps in the order specified in the online help:

1. Define the cluster topology
2. Configure application servers
3. Configure cluster resources
4. Verify a cluster topology
5. Save a cluster configuration
6. Activate the configuration.

Note: You can customize many aspects of **xhacmpm**'s window appearance, such as the background color. Specify customizations using command-line arguments, or by defining resources in the **.Xdefaults** file in your $HOME directory to override the defaults in the **xhacmpm** resource file.

Index

Vos remarques sur ce document / Technical publication remark form

Titre / Title : Bull HACMP 4.4 Enhanced Scalability Installation and Administration Guide
Volume 2/2

Nº Reférence / Reference Nº : 86 A2 89KX 01

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.

PLACE BAR CODE IN LOWER
LEFT CORNER

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

ORDER REFERENCE
86 A2 89KX 01

Bull

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 2/2

86 A2 89KX 01

AIX

HACMP 4.4
Enhanced Scalability
Install. & Admin.
Guide

Volume 2/2

86 A2 89KX 01

AIX

HACMP 4.4
Enhanced Scalability
Install. & Admin.
Guide

Volume 2/2

86 A2 89KX 01