Bull ESCALA EPC Series

EPC & HA Solutions
Setup Guide

AIX

ORDER REFERENCE
86 A2 79HX 05

Bull ESCALA EPC Series

EPC & HA Solutions Setup Guide

AIX

Software

September 1999

BULL ELECTRONICS ANGERS
CEDOC
34 Rue du Nid de Pie – BP 428
49004 ANGERS CEDEX 01
FRANCE

ORDER REFERENCE
86 A2 79HX 05

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.

About This Book

This guide is intended for system administrators and customer engineers who want to set up a Bull Escala Powercluster or Escala HA Solution. The guide discusses both the hardware and software aspects.

This guide also describes the features of the BullCluster software package, which comes with Powercluster and Escala HA Solutions.

If you are about to implement a Powercluster or an HA Solution, begin with this guide,as an entry point, before referring to other documents.

Terminology

Throughout this guide:

- the generic term *cluster* refers to any cluster type, whether it is a Powercluster or a HA Solution.
- Ethernet refers to Standard Ethernet (10 Mbps) and Fast Ethernet (10/100 Mbps).
- Disaster Recovery is defined as the capability to offer a site to site recovery within a limited radius, following a failure provoking loss of node and storage access. For example, in a campus environment.

Remote Monitoring

An optional remote monitoring facility is available for use with the Integrated System Management (ISM) product. ISM is not provided in this software package.

Overview of Contents

This guide is divided into the following chapters.

- **Chapter 1, "Concepts and Components Overview,"** provides an overview of the software and hardware components of Powercluster and HA Solutions.
- **Chapter 2, "Installation Overview and Planning,"** is the entry point to the overall planning/installation/configuration procedure.
- **Chapter 3, "Supplemental Planning Information,"** supplements the *HACMP Planning Guide*. It includes information relating to cabling concerns and shared disk drives.
- **Chapters 4 to 8** give step-by-step guidance for hardware and software setup.
- **Chapter 9, "The BullCluster Software Package,"** provides a roadmap to information related to the various utilities of the BullCluster software package.
- **Chapter 10, "Managing Cluster Resources and Shared LVM Components,"** explains how to change cluster resources (LVM components, application servers, resource groups...) using the menus of the BullCluster software package. These menus are integrated into the standard HACMP menus.
- **Chapter 11, "System Management Tools and Tips,"** discusses BullCluster management tools (diagnostics tool, event customization manager...) as well as special issues and tools (timed, dsmit, iFOR/LS, ArrayGUIde, Performance Toolbox...) useful when carrying out system management tasks in a cluster environment.
- **Chapter 12, "Remote Monitoring a Cluster with ISM,"** explains how to monitor remotely a cluster from an ISM (Integrated System Management) manager station. ISM is a Bull product designed to monitor and manage distributed systems and networks, in a user-friendly way, through a configurable graphical interface.

- **Chapter 13, “Integration of a PCI Drawer,”** uses a typical configuration to show how to integrate a PCI drawer in an Escala EPC rack.
- **Appendix A, “Implementing Console Solutions,”** discusses the different basic solutions that can be implemented to handle the consoles of the cluster nodes. Introduces the PowerConsole, System Console, Graphics Display, Cluster Console and associated devices with the Console concentrator and explains how to use the pwcons BullCluster utility.
- **Appendix B. ”Implementing the Cluster Console Solution”**, explains the installation of the X-terminal.

Related Publications

Powercluster & HA Solutions

The following documents apply to Powercluster and HA Solutions:

- *Escala EPC Series – EPC & HA Solutions Setup Guide*, Order Number 86 A2 79HX, this present guide.
- *Powercluster & HA Solutions: Using the Sample HA Scripts*, Order Number 86 A7 82HX, is mainly intended for Bull technical personnel. It gives hints on using the reference start and stop scripts that come with the BullCluster software package (see “Planning Specific Applications”, on page 3-1, for details).
- *Powercluster - Cluster Management Tools: ClusterWatch*, Order Number 86 A2 80HX, discusses hints and Powercluster-specific management tool ”ClusterWatch” which comes with the BullCluster software package.
- *Escala EPC Series – Site Preparation for Rack Systems*, Order Number 86 A1 30PX, provides site preparation information for Escala Racks and the Escala Powercluster (EPC) family.
- *Escala EPC Series – EPC Connecting Guide*, Order Number 86 A1 65JX, provides cabling and configuration information for the Escala EPC family. See also Other Publications, on page vi.

The following document applies only to Powercluster:

- *Escala EPC Series – PowerConsole & ClusterAssistant Setup Guide*, Order Number 86 A2 81HX, comes with the PowerConsole, which is optional and is proposed only with the Powercluster product (not with HA Solutions). The PowerConsole is a workstation that provides a single point of control for managing clusters. ClusterAssistant is a set of applications that run on the PowerConsole for configuring, managing, and monitoring clusters. ClusterAssistant is integrated in CDE (Common Desktop Environment), which provides a windows/icons/mouse-based framework.

Software Release Bulletin

Please read the *Software Release Bulletins* (SRB) that come with the software. For HA Solutions, Order Number 86 A2 81WF and for Powercluster, Order Number 86 A2 79WF. They include environment requirements and restrictions, the procedure to install the software, as well as late-breaking news.

HACMP for AIX Software Documentation

HACMP for AIX (High Availability Cluster MultiProcessing) is the core software used for clustering. The HACMP documentation set is listed below.

Release Notes

The *Release Notes* for HACMP are found in the **/usr/lpp/cluster/doc** directory (available once HACMP is installed).

HACMP Documentation for Administrators

These HACMP books are intended for administrators, and for anyone who has to understand and to deal with clusters:

HACMP 4.2

- *HACMP Concepts and Facilities,* Order Number 86 A2 54KX
- *HACMP Planning Guide,* Order Number 86 A2 55KX
- *HACMP Installation Guide,* Order Number 86 A2 56KX
- *HACMP Administration Guide,* Order Number 86 A2 57KX
- *HACMP Troubleshooting Guide,* Order Number 86 A2 58KX
- *HANFS Installation and Administration,* Order Number 86 A2 61KX
- *Enhanced Scalabilty Installation and Administration Guide*, Order Number 86 A2 62KX
- *Event Management Programming Guide and Reference,* Order Number 86 A2 63KX
- *Group Services Programming Guide and Reference,* Order Number 86 A2 64KX
- *HACMP Master Index and Glossary,* Order Number 86 A2 65KX
- *Bull System Management Guide Operating System and Devices for AIX,* Order Number 86 A2 53AP.

HACMP 4.3

- *HACMP Concepts and Facilities,* Order Number 86 A2 54KX
- *HACMP Planning Guide,* Order Number 86 A2 55KX
- *HACMP Installation Guide,* Order Number 86 A2 56KX
- *HACMP Administration Guide,* Order Number 86 A2 57KX
- *HACMP Troubleshooting Guide,* Order Number 86 A2 58KX
- *HANFS Installation and Administration,* Order Number 86 A2 61KX
- *Enhanced Scalabilty Installation and Administration Guide*, Order Number 86 A2 62KX
- *Event Management Programming Guide and Reference,* Order Number 86 A2 63KX
- *Group Services Programming Guide and Reference,* Order Number 86 A2 64KX
- *HACMP Master Index and Glossary,* Order Number 86 A2 65KX
- *Bull System Management Guide Operating System and Devices for AIX,* Order Number 86 A2 53AP.

HACMP Documentation for Programmers

Two additional books are intended for application developers who want to write specific highly available applications that run in an HACMP clustered environment:

HACMP 4.2

- *HACMP Programming Client Applications,* Order Number 86 A2 60KX
  This book describes the client application programming interfaces (APIs), which notably provide interfaces to the Cluster Information Program (clinfo). In addition, the book includes a listing of the HACMP MIB (since this MIB comes with the software, you can directly examine the MIB file to see its contents).
- *HACMP Locking Applications,* Order Number 86 A2 59KX
  This book describes the Cluster Lock Manager (CLM) application programming interface (API).

HACMP 4.3

- *HACMP Programming Client Applications,* Order Number 86 A2 60KX
  This book describes the client application programming interfaces (APIs), which notably provide interfaces to the Cluster Information Program (clinfo). In addition, the book includes a listing of the HACMP MIB (since this MIB comes with the software, you can directly examine the MIB file to see its contents).
- *HACMP Locking Applications,* Order Number 86 A2 59KX
  This book describes the Cluster Lock Manager (CLM) application programming interface (API).

Disk Storage Systems

These books provide information on installing, managing and configuring disk storage systems.

Escala JBOD Storage Subsystem – Setup & Operator Guide, Order Number 86 A1 79GX.

DAS 5700 Series Disk array Storage System, Installation & Service for Rackmount Models, Order Number 86 A1 43KX.

DAS 5700 Series Disk array Storage System, Installation & Service for Deskside Models, Order Number 86 A1 44KX.

DAS 5300 Disk array Storage System, Installation & Service for Rackmount Models, Order Number 86 A1 24KX.

DAS 5300 Disk array Storage System, Installation & Service for Deskside Models, Order Number 86 A1 25KX.

DAS 5000 Disk array Storage System, Installation & Service for Rackmount Models, Order Number 86 A1 45KX.

DAS 5000 Disk array Storage System, Installation & Service for Deskside Models, Order Number 86 A1 46KX.

DAS 3500 Disk array Storage System, Installation & Service for Rackmount Models, Order Number 86 A1 47JX.

DAS 3500 Disk array Storage System, Installation & Service for Deskside Models, Order Number 86 A1 48JX.

7133 SSA Disk Subsystems Service Guide, Order Number 86 A1 94GX.

PCI Fibre Channel Adapters, Installation & Configuration Guide, Order Number 86 A1 95HX.

Escala Disk Expansion Unit Service Guide, Order Number 86 A1 13PX.

Planning a DAS Installation (Fibre Channel Environment), Order Number 86 A1 94JX.

Navisphere Setup and Operation, Order Number 86 A2 47KX.

Configuring and Managing a DAS in an AIX Environment, Order Number 86 A2 20PN.

Other Publications

The AIX document set, as well as manuals accompanying machines, disks, and other hardware involved in the cluster, also provide indispensable information: have them handy.

Cabling Guide, Order Number 86 A1 87AQ.

Ordering Publications

To order additional copies of this guide, use Order Number 86 A2 79HX.

Table of Contents

Chapter 1. Concepts and Components Overview

This chapter includes the following sections:

Cluster Concepts

Note: For a more detailed introduction to cluster concepts, refer to the *HACMP Concepts and Facilities* guide.

Clustering technology is an efficient solution for building mission-critical platforms. It is able to meet two major goals: high availability and high performance.

High Availability

The clustering technology consists of coupling several networked servers. In such an environment, each server backs the others. The servers and the applications they run are made **highly available**. In case of failure or maintenance operation, this provides the ability to failover a current application workload from the failing server to another server, and to have that application become available to the users with **minimal disruption**.

Note that in most cluster configurations (such as "mutual takeover"), the backup servers are not required to stay idle: they are available for use during normal operation.

Concerned Applications

High availability is an essential feature for many enterprises that run business-critical applications. Any application benefits from this feature and can be made highly available.

Terminology

Clusters designed for high availability are sometimes referred as "fault-resilient" clusters (or even "near fault-tolerant" clusters). "HA" is often used as an abbreviation for "high availability" or "highly available".

Disaster Recovery

There are two levels of disaster recovery:

- **Extended high availability** (Bull's solution). In this case the cluster is shared between two sites with common data storage. This shared storage limits the inter-site cabling distance. The local HA feature is maintained if certain precautions are observed. Details, see page 10-22.
- **Geographical disaster recovery**. In this case the sites work independantly with no shared storage. Mechanisms are use to assure replication of data on the backup site and High Availability is limited to mutual surveillance and site takeover.

High Performance

In addition to **high availability**, the clustering technology provides the ability to achieve high performance and scalability.

Concurrent Access Mode

A cluster can be configured to work in **concurrent access** mode, where the different servers access concurrently the same data on shared disks. Concurrent access configurations offer the following advantages:

- Outstanding performance and scalability:
  - The load of a single application can be split across multiple nodes.
  - Simply by adding servers or other resources to an existing configuration, the cluster's performance grows with your needs.
- Enhanced availability:
  - Failover time is reduced, notably because disk takeover is not required when a cluster node fails.
  - If a cluster node is brought down, access to the data from the other nodes is not interrupted. Applications can switch to another node immediately.

Only applications that have been specifically adapted for concurrent access are able to benefit from this feature.

- A typical application is the Oracle Parallel Server (OPS), an extension to the Oracle RDBMS.
- Other RDBMS and OLTP applications will be available for running in concurrent access mode.

Check with your Bull representative for availability of RDBMS and OLTP applications adapted to concurrent access.

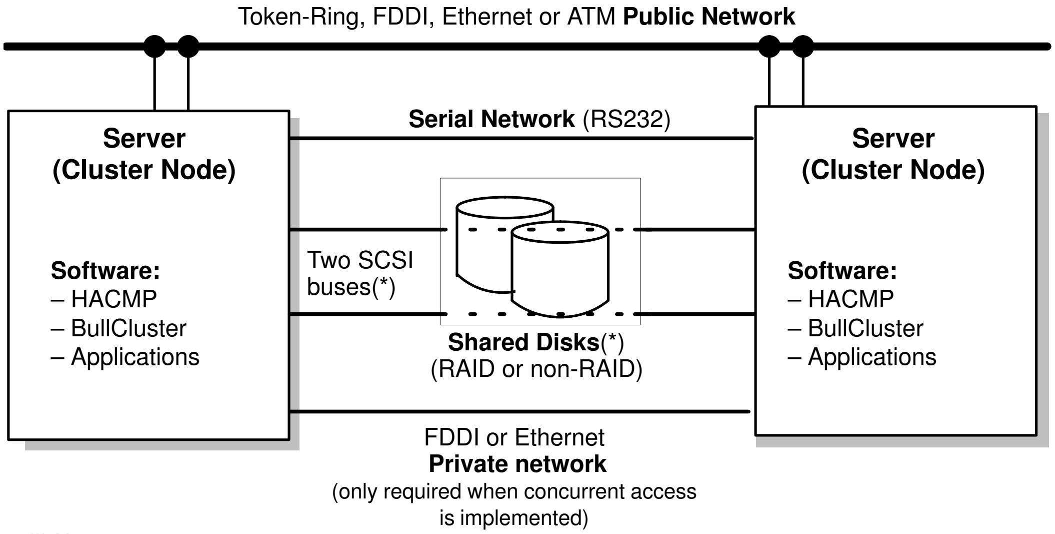
Note: All of the expressions "concurrent access", "concurrent multi–processing", "cluster multi–processing" and "CMP" refer to the same concept. Also note that concurrent access mode is sometimes referred as "mode 3" in the HACMP documentation.

Overview of Powercluster and HA Solution Components

Powerclusters and HA Solutions are clusters built by combining software and hardware components. Both products provide great power and availability.

Cluster Components at a Glance

The illustration below shows the main hardware components of a sample 2-node cluster.



(*) Notes:

- Most configurations use two SCSI buses. However, low-cost configurations, having a single SCSI bus, can be implemented.
- In the so-called “HA Native” configuration, available for E and T series, internal disks can be used instead of external disks for shared storage.
- Instead of SCSI-based shared disks and buses, SSA (Serial Storage Architecture) disk subsystems and connection loops can be used.
- Instead of SCSI-based shared disks and buses, Fibre Channel disk subsystems, FC-AL loops and Fibre Channel Adapters can be used.

Software Components

The software components of a cluster are:

- HACMP for AIX, the core software for clustering,
- BullCluster, a Bull software package comprised of various utilities and productivity tools,
- any application to be made highly available, running on the servers.

Hardware Components

The main hardware components of a cluster are:

- servers (the nodes of the cluster),
- shared disk devices (either RAID or non-RAID; supported disk devices are described later in this chapter),
- hardware connections: SCSI (or SSA) adapters, fibre channel adapters, adapters for public and private networks, RS232 serial node-to-node connections.

Note: Notions related to networks (public, private, and serial) are explained on page 3-4.

Console Solutions

Various hardware solutions for console handling are available with the Powercluster product. The **PowerConsole**, notably, is a workstation that provides a single point of control for managing clusters. It serves both as the access point (via a concentrator) to the system consoles of the different nodes, and as a powerful workstation able to run any administration utility.

Hardware Redundancies

High availability requires some hardware redundancies in order to eliminate most of the potential "single points of failure" (SPOFs).

For example, the sample configuration shown in the illustration above includes two nodes, two network adapters per node, and two distinct SCSI buses for access to external shared disks. Data on disks should be redundant, either through AIX mirroring features (for non-RAID disks), or through the use of RAID disks (RAID disks provide their own data redundancy).

Servers (Cluster Nodes)

HA Solutions are clusters built on Escala D, T, M and E series servers.

The Powercluster offer is made up of Escala rack-mountable servers. Three uni-node server models are available:

- EPC400, EPC430 and EPC440 with a PCI bus
- EPC800, with an MCA bus
- EPC1200, EPC1200A and EPC2400 with a PCI bus and consisting of two racks (a computing rack with a CPU drawer and an expansion rack with an I/O drawer).

  In addition, multiple node configurations can be defined from these models by adding more nodes.

HACMP

HACMP (High Availability Cluster Multi–Processing) for AIX is the core software for clustering. It provides mechanisms for controlling the cluster.

The main HACMP module, “**High Availability Subsystem**”, is in charge of:

- recognizing changes within the cluster (among these changes are notably hardware failures, and intentional server shutdown for maintenance),
- starting and coordinating the actions needed, when a failure occurs, for re–configuring the environment and assuring resources takeover.

In addition, HACMP features an optional module, the “**Concurrent Resource Manager**” (CRM). It adds concurrent shared access management for the disks. This module is required for implementing concurrent access, and is provided as an option.

Note: For a more detailed description of HACMP features, refer to the *HACMP Concepts and Facilities* guide.

BullCluster Software Package

BullCluster is a software package developed by Bull. It includes various utilities and productivity tools useful when dealing with a cluster environment.

The features of BullCluster are summarized in “General Tools” and “Powercluster-Specific Tools” below.

General Tools

Quick Initial Configuration Tool

BullCluster includes an easy-to-use quick initial configuration tool, which is implemented as an extension to the SMIT menus of HACMP (SMIT, System Management Interface Tool, is the standard menu-driven AIX facility dedicated to system management).

The quick initial configuration tool allows you to configure the cluster topology with only a few menus. In addition, it provides options for customizing HACMP event processing.

Easy-to-use Menus for Managing Cluster Resources

BullCluster adds a menu to the original HACMP menus that facilitates these administrative tasks:

- creation and modification of cluster resources for use by HACMP
- set up and management of shared LVM components (volume groups, logical volumes, file systems)

Reference Takeover Scripts

BullCluster includes start and stop scripts useful for implementing Informix, Oracle, OPS, Tuxedo, HVX, X25 and Internet services. These scripts are provided for reference only, and are intended to be customized to fit customer requirements (for details, see “Planning Specific Applications”, on page 3-1).

Utilities for Supporting Various Shared Storage Configurations

BullCluster provides additions to HACMP scripts, in order to support the DAS (Disk Array Storage) RAID subsystems, the SSA (Serial Storage Architecture) disk subsystems,the Disaster Recovery (mirrored shared disk subsystems), and the so-called “HA Native” configuration (the latter consists of two Escala servers that mutually share their internal disks, as described on page 3-7).

BullCluster Diagnostic Tool

BullCluster provides a diagnostic tool that allows you to generate diagnostic information. The diagnostic tool analyzes the state and consistency of cluster resources, and scans HACMP and system log files for major cluster events and hardware errors. This tool is provided as UNIX commands.

Utilities for Remote Monitoring a Cluster (and OPS) through ISM

BullCluster includes utilities that facilitate the use of ISM (Integrated System Management) manager station for:

- remote monitoring the cluster state
- remote monitoring and tuning OPS (Oracle Parallel Server)

BullCluster includes also utilities that generate SNMP traps when errors related to shared disks are detected. These SNMP traps can be monitored through ISM.

Note: ISM is a Bull product designed to monitor and manage distributed systems and networks. ISM offers a user-friendly and configurable graphical interface, through which administrators can visualize and control the networked systems.

ClusterWatch

ClusterWatch is a Web-based facility for monitoring cluster operations. With ClusterWatch, the system administrators instantly have access to all the critical information they need to manage clusters and maximize up–time.

Note: A SMIT-based version of ClusterWatch is also provided.

Other Utilities

BullCluster also includes the **pwcons** and **clusterview** utilities, provided as UNIX commands:

- **pwcons** is a utility that enables the operator to gain control of the system consoles of nodes that are not equipped with an attached terminal.
- **clusterview** is a utility that reports the status of clusters. It is comparable to **clstat**, which comes with the HACMP software.

Disk Devices

This section concerns the shared disks, i.e. the disks that can be accessed by the different cluster nodes.

Note on Internal Disks

The servers come with internal disk devices. These disks are used only to store system files and applications (and the **rootvg** system volume group). They are not suitable to store the data accessed by the applications you want to be highly available. In other words, internal disk devices cannot be used as shared disks in a cluster (an exception to this rule is the "HA Native" configuration, discussed on page 1-7).

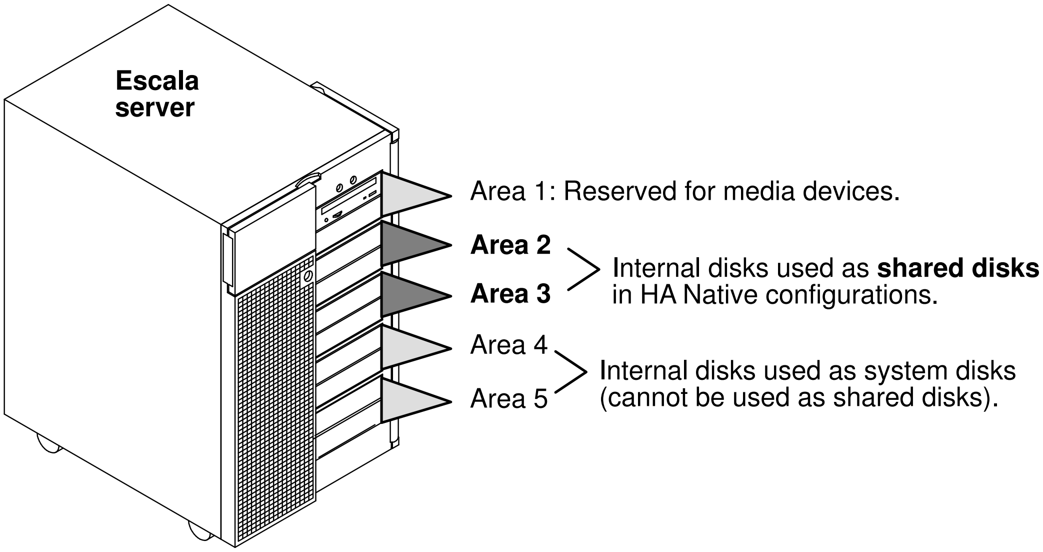
Note that internal system disks (notably the **rootvg** system volume group) can be mirrored to increase node availability. The mirroring function is implemented through the standard features of the AIX system.

"HA Native" Shared Storage

For HA Solutions based on Escala E or T servers, it is possible to implement the so-called "HA Native" configuration.

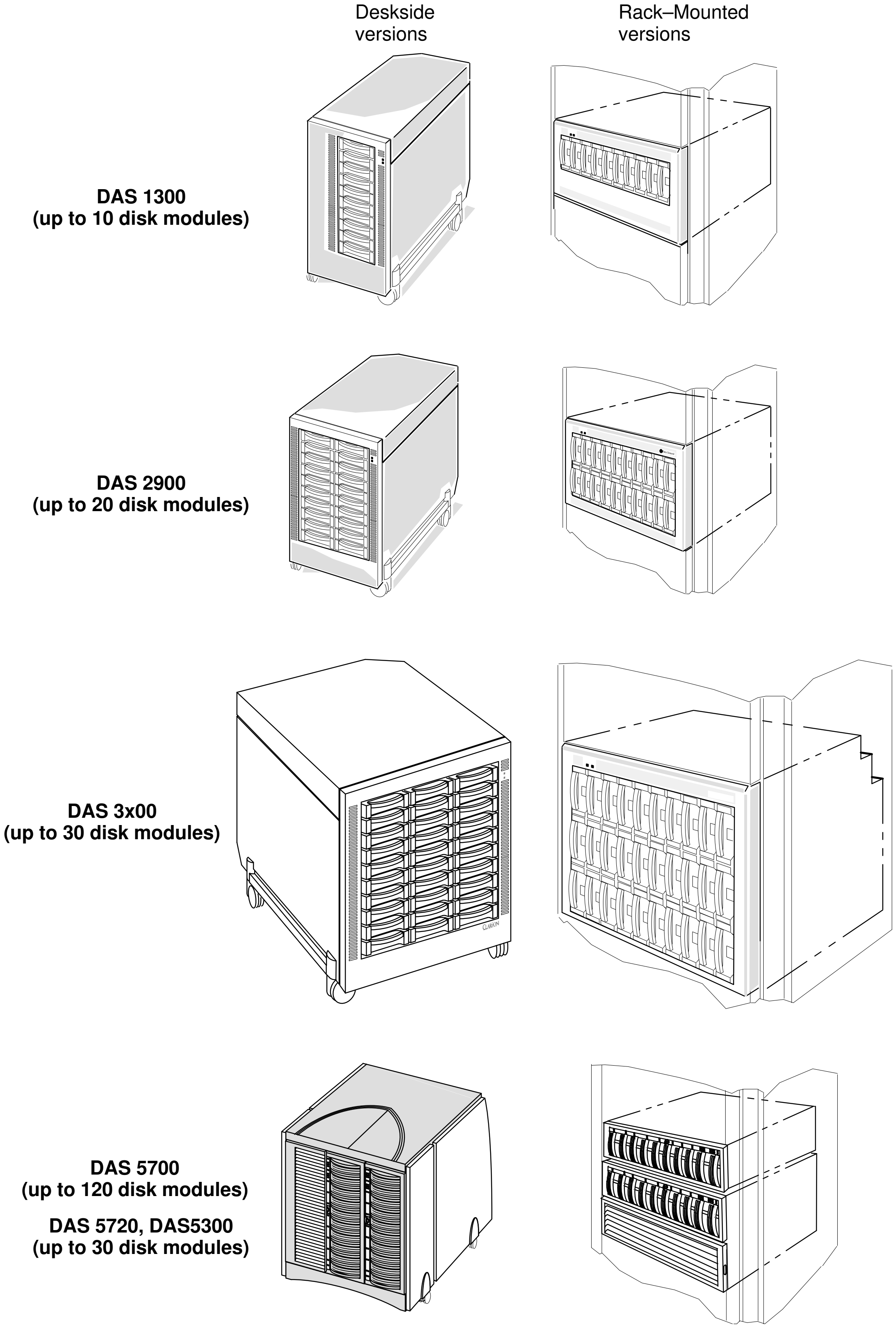
A HA Native configuration consists of two Escala servers that mutually share their internal disks. HA Native configurations provide a cost-effective yet powerful alternative to configurations that use external subsystems for shared storage.

Refer to page 3-7 for details on HA Native configurations.


Escala server
Area 1: Reserved for media devices.
Area 2
Area 3
Internal disks used as **shared disks** in HA Native configurations.
Area 4
Area 5
Internal disks used as system disks (cannot be used as shared disks).

DAS RAID Subsystems

The DAS (Disk Array Storage) subsystems offer RAID functionality, allowing transparent data recovery in case of disk failure. They are perfect for high availability and clusters that require large disk storage capacity. They are available for Powerclusters and any HA Solutions based on Escala servers.



DAS 1300/2900/3200 Features

These DAS subsystems feature great “no-single-point-of-failure” capabilities, through dual SCSI bus configurations, as well as redundant storage processors and power supplies. These DAS subsystems use SCSI-2 DE F/W connections. In order to maximize data availability, all major components, including disk drives, storage processors and power supplies, are replaceable when under power.

These DAS subsystems are modular and provide high-level flexibility and an optimum balance between reliability and performance (high-speed disks, cache-write). DAS 2900 and DAS 3200 offer improved performance, because they are equipped with new, enhanced, storage processors.

Note: Storage Processors (SPs) are integrated units that control the disk modules through a SCSI bus.

DAS 3500 (Fibre Channel) Features

DAS subsystems feature great “no-single-point-of-failure” capabilities, through dual fibre channel bus configurations, as well as redundant storage processors and power supplies. In order to maximize data availability, all major components, including disk drives, storage processors and power supplies, are replaceable when under power.

A Fibre Channel (FC) cable connects the storage system to a fibre channel Host Bus Adapter (HBA) in the server. Commands and data circulate between the storage system and its server(s) in a Fibre channel Arbitrated Loop (FC-AL).

DAS 5700/5720/5300 (Full Fibre Channel) Features

Fibre Channel disk-array storage systems provide terabytes of disk storage capacity, high transfer rates, flexible configurations, and highly available data. Full Fibre Channel implementation, includes the use of fibre channel disk drives (unlike DAS 3500 systems, that use SCSI disks).

High avaibility design: all the components are designed for nondisruptive removal and insertion; they can be replaced under power.

RAID Levels

DAS subsystems support RAID levels 0, 0/1, 1, 3, and 5.

Note: For details, on DAS-based configurations, refer to page 3-11.

SSA (Serial Storage Architecture) Disk Subsystems

SSA disk subsystems are available for Powerclusters and any HA Solutions based on Escala servers.

Serial Storage Architecture in Brief

The Serial Storage Architecture (SSA) is a point-to-point architecture for storage devices. SSA, which is now an industry standard, can be an alternative to SCSI-based storage configurations.

SSA delivers high performance:

- SSA provides full duplex communication and offers a total bandwidth of 80 MBytes/s (two read channels and two write channels are used simultaneously, each at 20 MBytes/s).

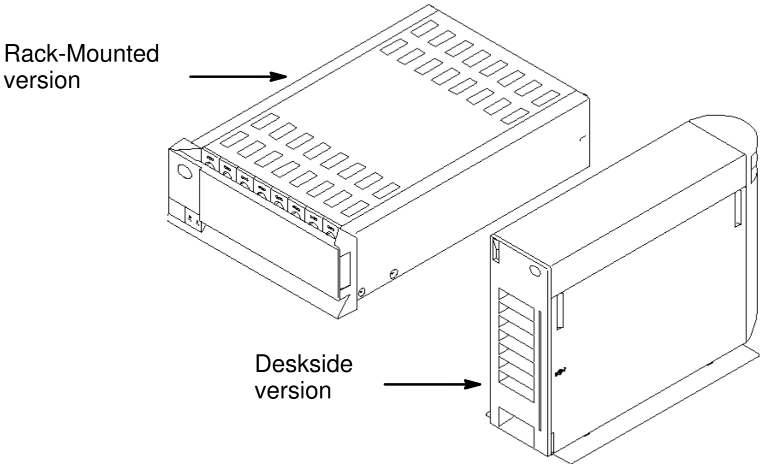
Significant availability features are inherent in the SSA technology:

- In a cluster, SSA links are implemented as loops, in order to eliminate data paths as SPOFs (single points of failure). With such a loop topology, nodes can get access to the disk subsystems through two distinct data paths. If a data path becomes unavailable, an alternate path is automatically used. Thus, availability is enhanced, and hot pluggability is made possible.

Supported SSA Disk Subsystems

SSA disk subsystems are available in both rack-mounted and deskside versions. The first are designed to fit in the racks of the Escala Rack-Mounted servers.

SSA Disk Subsystems



Here are some key features of these SSA disk subsystems:

- A single subsystem can house from 4 up to 16 disk drives. The storage capacity is easily augmented by interconnecting multiple subsystems: up to 48 disks can be connected in one loop, and up to 96 disks can be connected to one SSA adapter.
- Most of the components of the SSA disk subsystems (disk drives, power supplies, SSA cables...) can be added, removed or replaced without stopping the operation of the subsystem and the using nodes. Redundant power and cooling is available as an option.

Data Availability Concerns

The supported SSA disk subsystems do not provide inherent RAID data availability functions. Therefore, the administrator must implement mirroring (using the standard AIX features) to implement data redundancy.

Note: For details on SSA-based configurations, refer to page 3-22.

EMC’s Symmetrix CDA Storage Systems

Powerclusters are usually equipped with DAS RAID subsystems, or with SSA disk subsystems as shared disk devices. Optionally, they can be equipped with Symmetrix CDA storage systems.

The Symmetrix CDA systems (from the EMC manufacturer) provide RAID data protection, including RAID-1, RAID-S (an EMC developed RAID strategy), and dynamic spare disk volumes.

The setup of these systems is not described in the present document, and must be carried out by specialized field personnel. Check with your Bull representative for details.

Tape Libraries

High performance, multiple magazine tape drives of type DLT4000/DLT7000, from Overland, provide data backup facilities. The drives, housed in rackmounted units make up the LibraryXpress LXB4000/LXB7000 libraries.

An LXB library can be attached to a single node. The library can only be shared for high-availability by two nodes of the same type with the HA NetBackup application. The Master Server or Slave Server can be redundant implementing hot standby or mutual recovery mode.

Chapter 2. Installation Overview and Planning

This chapter is the entry point to the overall procedure for planning, installing and configuring your cluster.

Section “Procedure Overview”, summarizes the steps involved.

Once you have completed the planning process, proceed with the next chapter to carry out the installation. Simply follow the steps in the order they are listed. When needed, the steps instruct you to refer to other chapters in this document or to other appropriate documents.

Procedure Overview

This section summarizes the steps of the planning/installation/configuration procedure.

The involved steps are:

1. **Plan the cluster installation**
2. **Set up the console and the basic hardware components**
   - Attach system consoles and network connections. Do not connect now the external shared disk devices.
3. **Boot each node and check for proper operation**
4. **On each node, configure TCP/IP boot address for the service and standby network adapter**
   - Use `smit mktcpip`
5. **Configure the adapter SCSI IDs** (not applicable if you use SSA or fibre disk subsystems)
   - Then reboot the nodes to make the change effective.
6. **Check the software installation**
7. **Install the software** (this step is usually not needed because software is pre-loaded at factory)
8. **Shutdown and power off the nodes**
9. **Interconnect nodes via RS232 serial networks**
10. **Prepare the SCSI adapters for later cabling** (not applicable if you use SSA-based shared disk storage)

    If you use SCSI-based shared disk storage, this step applies only if you are adding a supplementary server to an existing cluster; this is not needed if you are installing a new, complete, cluster, that has been set up at the factory.
11. Prepare the fibre channel loops.

    If using fibre channel adapter with DAS (optionally FC-AL Hub) establish the different FC-AL loops between the DAS subsystems and fibre channel adapters, directly or indirectly, using the FC-AL Hubs if needed.
12. **Set Up the shared disk devices** (non-RAID or RAID disks)
    - Cable the disk devices (and set their SCSI IDs, if applicable).

13. **Start up the disks and the nodes, configure the DAS RAID subsystem**
    - Create and set up LUNs on DAS RAID subsystems.
    - Do not create shared volume groups now.
    - If using FC-AL Hubs, power on Hubs and Disk subsystems. When the subsystem is ready, the service light (yellow LED) turns OFF. then Power on the nodes, and configure the subsystem.
14. **Determine which method to use for configuring the cluster**
    - Either the standard HACMP-based configuration method, or the quick initial configuration method (BullCluster-specific).
15. **Configure the cluster topology for HACMP**
    - using the quick initial configuration tool,
    - OR using the standard HACMP-based configuration method
16. **Complete the configuration** as explained in chapter “Completing the Installation”, starting on page 8-1.

Where You Go From Here

Once you have completed the first step, which consists in planning the cluster installation, go to the chapter “Setting Up Basic Hardware”, on page 4-1, and proceed with the step-by-step instructions to complete the installation and configuration tasks.

Chapter 3. Supplemental Planning Information

This chapter supplements the *HACMP Planning Guide*. It includes information related to cabling concerns, shared disk drives, consoles, and specific applications.

- Planning Specific Applications
- Planning Serial Ports, on page 3-3.
- Planning Networks, on page 3-4.
- Planning “HA Native” Shared Storage for Escala HA Solutions, on page 3-7.
- Planning Shared DAS Subsystems, on page 3-11.
- Incorporating Disk-Array Enclosures (DAE), on page 3-19.
- Planning Extended HA Shared Storage for Escala HA Solutions, on page 3-20.
- Planning Shared SSA Disk Subsystems (Serial Storage Architecture), on page 3-22.
- Planning Shared EMC Symmetrix Subsystem Disks, on page 3-25.

For further information about cables, refer to the *Cabling Guide*, which comes with the hardware.

Planning Specific Applications

The BullCluster software package includes start and stop scripts useful for implementing various applications in your cluster as highly available applications. However, it should be stressed that these scripts are provided as reference only (not as supported softwares), and must be customized to match customer needs.

- In accordance with your specific environment and requirements, you may want to use the scripts as a basis to write appropriate custom scripts, or you may decide to write your scripts entirely on your own, without making use of the provided scripts.
- In any case, have in mind that writing reliable start and stop scripts requires thorough technical knowledge about HACMP and the application to be implemented. We recommend you get assistance from Bull technical services.

Sample Scripts for Major Applications

Sample scripts for the applications listed below are located under the `/usr/sbin/bullcluster/install/ha_conf` directory:

- Oracle
- OPS (Oracle Parallel Server)
- Informix
- Tuxedo
- Internet services (Netscape Enterprise Server, Netscape Proxy Server, NetWall)
- HVX
- X25

In addition, note that the **Create and Add an Application Server to a Resource Group** submenu of the BullCluster **Cluster Resources Modification** SMIT menu allows you to easily define for HACMP an application server for any of the applications listed above.

Sample Scripts for Miscellaneous Applications

The BullCluster software package comes with other sample scripts for other applications. These samples and their associated README informative files are located under the **/usr/sbin/bullcluster/install/samples** directory. Look at the files in this directory for additional information.

Documentation for Using the Reference Scripts

The *Powercluster & HA Solutions: Using the Sample HA Scripts* document, mainly intended for Bull technical personnel, provides hints for using reference start and stop scripts that come with the BullCluster software package.

Planning Serial Ports

Each node requires several RS232 serial ports to handle the hardware of a cluster. This section will help you to determine the total number of serial ports needed for the nodes of your cluster.

Note on the S2 Serial Port

Special considerations apply to the S2 serial port on the Escala servers.

EPC 1200/1200A/2400 Server

The S2 port of the Escala EPC 1200/1200A and EPC2400 must only be used for remote maintenance. It cannot be used for any other purpose. The 8–port (16-port) must be used for any serial connection needed in the HACMP configuration.

All other ESCALA Servers

EPC400/430/440 and ESCALA E or T use COM2 (external modem) or COM3 (UPS) and not 8-port for heartbeat and DAS connection.

(A) Serial Ports Needed for System Console and DAS SP Connection

For each cluster node:

- One serial port (the S1 serial port) is needed to provide for system console connection. This rule applies whether or not the node has a console directly attached.
- If DAS subsystems are used, one supplementary serial port (usually the S3 or COM3 serial port) is needed per DAS subsystem in order to connect the node to a DAS storage processor (SP).

Thus, for system consoles and DAS SP connections, the total number of serial ports needed per node is:

- 1 serial port if the cluster does not include a DAS subsystem
- 2 serial ports if the cluster includes one DAS subsystem (one SP)
- 3 serial ports if the cluster includes two chained DAS subsystems

(B) Serial Ports Needed for Serial Network Connections

Furthermore, each node requires additional serial ports to connect serial networks:

- 1 serial port if the cluster includes two nodes
- 2 serial ports if the cluster includes more than two nodes

(C) Serial Ports Needed for UPS

If UPS (Un–interruptible Power Supply) are used (optional), each node requires one additional serial port.

Planning Networks

This section supplements the network-related discussions given in the *HACMP Planning Guide*. It discusses some additional specificities and hints that apply to the Powercluster and HA Solutions offerings.

Public, Private and Serial Networks

According to the HACMP terminology:

- A *public* network is a network that connect multiple nodes and allows clients to access the cluster nodes.
- A *private* network is a network that provides point-to-point communication between cluster nodes. A private network is mainly used for lock traffic between nodes. It is required to implement concurrent access. It is also used for implementing 3-tier architecture.
- A *serial* network consists in RS232 links that connect nodes together, in order to transport "keep-alive" traffic generated by HACMP.

Public Network

The public network can be implemented through Ethernet, FDDI, Token-Ring or ATM.

Ethernet (Including Fast Ethernet)

In typical configurations that rely on Ethernet, each node is equipped with two Ethernet 10/100 Mbps or 1Gb adapters (one service adapter and one standby adapter) to connect to the public network. Some complex cluster configurations, that implement two public ethernet networks for redundancy, require four ethernet adapters per node.

FDDI

When the public network is implemented through FDDI, it is recommended to use dual-ring configurations (since dual-rings provide a higher availability than single-rings).

In typical configurations that rely on FDDI, each node is equipped with two dual-ring adapters (one service adapter and one standby adapter) to connect to the public network. Some complex cluster configurations, that implement two public FDDI networks for redundancy, require four dual-ring adapters per node.

ATM

ATM (Asynchronous Transmission Mode) can be used to implement the public network, providing a throughput of 155 mega–bits per second.

In typical configurations, an ATM switch is used and each node is equipped with two ATM adapters (one service adapter and one standby adapter).

Private Network

Private networks concern only clusters that implement concurrent access or clusters implementing 3-tier architecture nodes having a role of application servers and other nodes with a role of database servers.

The private network is implemented through Ethernet or dual-ring FDDI.

- For information about Ethernet:
  - single or double interconnect with a standard hub (> 2 nodes).
  - single or double interconnect with a high performance switch (> 2 nodes).

  See your Customer Service Representative for information concerning the planning of a Fast Ethernet switch.
- For further advice on FDDI, see "Planning FDDI Topology for Private Networks" below.

FDDI Hub and Ethernet Hub: Special mounting kits (one for the FDDI hub, one for the Ethernet hub) can be ordered to fit these hubs in a rack.

Planning FDDI Topology for Private Networks

Note: This section applies only to clusters to be equipped with private networks.

Typical implementations make use of only one private network to interconnect cluster nodes. However, for even higher availability, a cluster may include two private networks: such complex configurations are not discussed here.

When planning a FDDI private network, you must have in mind that its availability degree depends on the topology you choose. Some guidelines are given below.

Clusters Having 2 or 3 Nodes

If the cluster has two or three nodes, you can implement the FDDI private network by the means of a simple dual ring, without any other specific FDDI device.

Indeed, even in the event of a ring wrap (if a ring is broken due to a node powered off or to a faulty FDDI section), no fragmentation occurs: each node is still able to communicate with the others.

Notes:

1. From a scalability standpoint, an FDDI private network is constructed with a hub (or a switch, for higher performance).
2. A FDDI private network based on a single ring is not suitable for high availability purpose. A FDDI dual ring is made of two separate loops (rings), and thus provides higher availability (one ring is the primary ring and the other, the secondary ring, providing backup to the primary ring).
3. As explained above, from a technical viewpoint, a 3-node cluster does not require the use of special FDDI devices (an ordinary dual-ring is sufficient). However, note that 3-node Powercluster models can be delivered with 2 FDDI hubs, so that you can implement a dual-homing FDDI configuration, as explained in “Clusters having 3 or More Nodes” below. Adopting the dual-homing configuration for a 3-node cluster facilitates subsequent upgrade operations (addition of supplementary nodes to the 3-node cluster).

Clusters Having 3 or More Nodes

If the cluster has three or more nodes, implementing the private network as a simple FDDI dual ring may be regarded as unsatisfactory from the availability standpoint. Indeed:

- In the event of a ring wrap (if a ring is broken due to a node powered off or to a faulty FDDI section), the ring continues to operate normally. Each node is still able to communicate with the others.
- The problem is when a second ring wrap occurs. In this event, the FDDI network becomes fragmented into two independent rings. The nodes in a fragment are thus unable to communicate with the nodes that are in the other fragment.

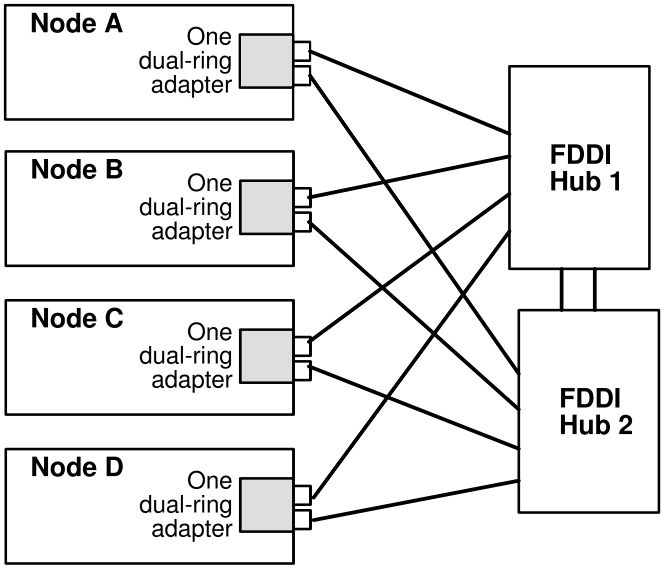
Dual-Homing FDDI Configuration

To prevent the private FDDI from becoming fragmented, it is necessary to connect the nodes to a FDDI hub or concentrator. With a FDDI hub or concentrator, a node can be powered off or a a FDDI adapter can fail without fragmenting the network.

However, if you have strong requirements concerning the private network availability, this may be insufficient. Indeed, the hub or concentrator constitutes a SPOF (Single Point of Failure): if it fails, the FDDI network goes down. To eliminate this SPOF, the solution is to implement a dual-homing configuration, which relies on two FDDI hubs or concentrators. In such a configuration, nodes are dual-homed, i.e. they are connected to two distinct hubs or concentrators to provide an alternate path should the primary path fail.

In the suggested dual-homing configuration of the Powercluster offering, each node is equipped with a dual-ring FDDI adapter and relies on two FDDI hubs (Linkbuilder Model from the 3COM vendor).

This dual-homing configuration is illustrated below (4-node cluster example):



PRIVATE NETWORK: FDDI Implementation with Dual-Homing

Planning “HA Native” Shared Storage for Escala HA Solutions

Prior Knowledge

- Please read first the “Planning Shared Disk Devices” chapter included in the *HACMP Planning Guide*.

“HA Native” Configuration Overview

For HA Solutions based on Escala E and T servers, it is possible to implement the so-called “HA Native” configuration.

Shared SCSI Buses and Mirroring Concerns

Two configurations are possible:

- using internal disks only
- using internal disks and a disk expansion cabinet.

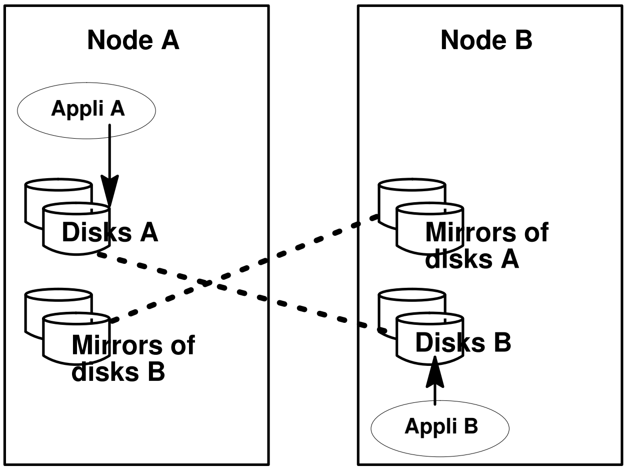
In an HA Native two shared SCSI buses are implemented with non-RAID disks.

Thus for high availability purposes, disk mirroring must be implemented. The mirroring function is implemented through the standard features of the AIX system. This is explained in section “LVM Mirroring” of the *HACMP Planning Guide*.

Mirroring Scenario with Simple Native HA

Mirroring must be implemented across the two nodes. That means that the internal disks of one node (that are accessed through one shared SCSI bus) must be mirrored on the internal disks of the other node (that are accessed through the other shared SCSI bus).

The following figures illustrates a sample mutual takeover configuration in normal operation:

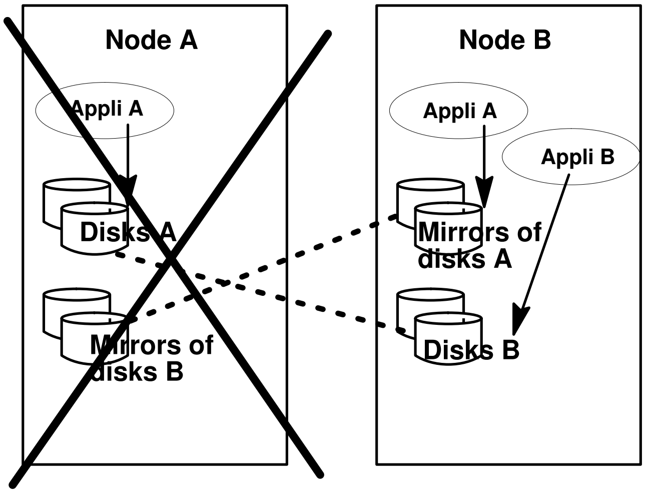


Normal operation:

• Node A runs the Appli A application which uses disks A for data storage. Disks A are mirrored on disks of Node B. Disks A and their mirror are on a different SCSI bus, and then, can be accessed by both nodes.

• On Node B, a similar, symmetrical, environment has been set up. Another application, Appli B, uses disks B that are mirrored on the other node.

Assume that Node A fails. In that case, HACMP carries out a failover process: it starts the Appli A application on the surviving Node B. Then, Appli A works by using the mirrors of disks A, as shown in the illustration below.



Operation after failover:

- Node A is down.
- Node B runs both Appli A and B applications. The Appli A application now uses the mirrors of disks A.

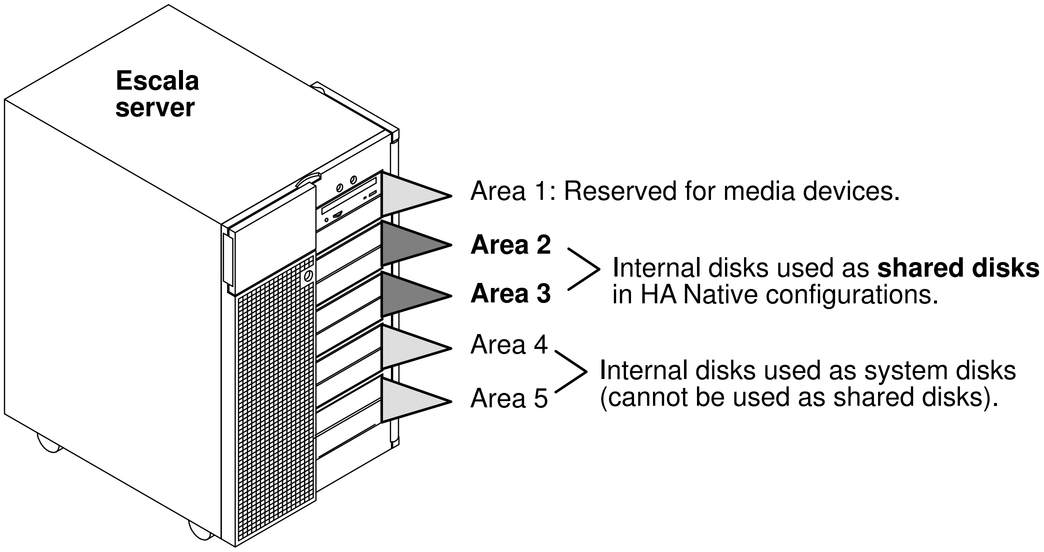
Mirroring Scenario with Disk Expansion Cabinet

When using the Disk Expansion Cabinet, the mirror of disks internal on Node A is made using disks in the Expansion Cabinet. Similarly, those internal disks of Node B are also mirrored in the Expansion Cabinet.

For further information, refer to the Disk Expansion Unit Service Guide and the PCI Expansion Drawer Quick Set Up.

Internal Disks to be Used

In HA Native configurations, the disks used as shared storage are the internal disks located in device area 3, and optionally device area 2, of the server base unit.


Escala server
Area 1: Reserved for media devices.
Area 2
Area 3
Internal disks used as **shared disks** in HA Native configurations.
Area 4
Area 5
Internal disks used as system disks (cannot be used as shared disks).

Note the following:

- Device areas 3 and 2 use the same SCSI bus. Each area can house up to two 1.6 inch high disks or three 1 inch high disks.
- For mirroring purposes, the configuration should be as symmetrical as possible. Shared device areas (area 3 and optionally 2) of one node should be populated in the same way as shared device areas of the other node: identical disks, same capacity, same number and same location.

SCSI Adapters and Bus Cabling

Note the following:

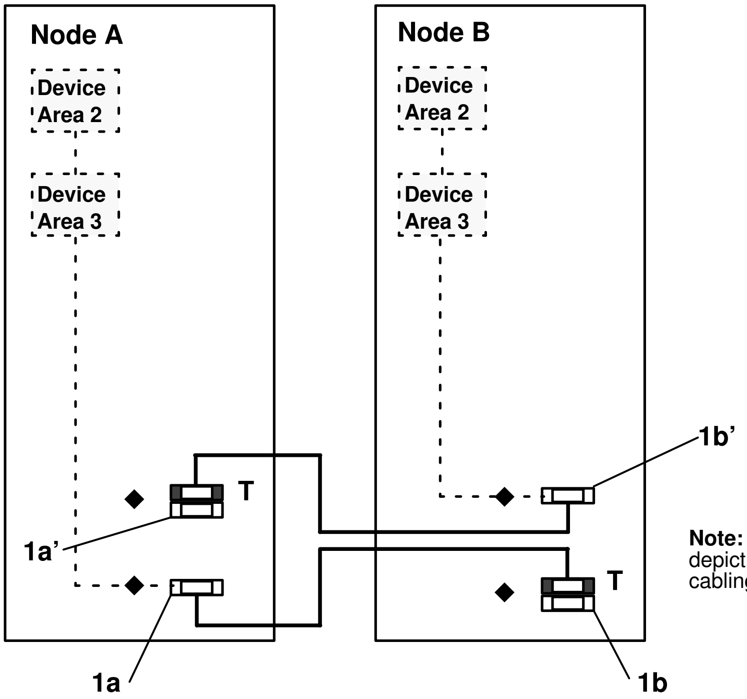
- To implement the two SCSI buses, each node must be equipped with two Ultra SCSI, single-ended, adapters. These adapters are labeled with the “B4-5” code. (Four adapters are needed with Native HA and the Disk Expansion Cabinet.
- Increased-availability is implemented by using cables with integrated terminators. Implementing such a cabling method facilitates repairs. It allows you to unplug an Y-cable without breaking the continuity of the concerned SCSI bus.

Two SCSI shared buses are to be implemented. The illustrations below show how to cable the first shared SCSI bus (rear side of Escala servers):

CAUTION:
All servers are supplied with standard internal cabling and connectors which must be adapted to suit the configurations shown.

Simple Native HA

This Simple Native HA configuration provides a low–cost HA disk extension solution with mirroring/sharing of internal disks over two SCSI chains.



Note: Thick dotted lines depict internal SCSI cabling.

Total length of cable bus must not exceed 1.5 metres.

Component References:
Internal SCSI cabling, Ref: CBLG178
(P/N: 78172894-001)

SCSI Adapters (type B4-5) ▭

SCSI Cable + 2 wrap plugs (90982001-001)
+1 pass-thru terminator (91076001-001) ▬

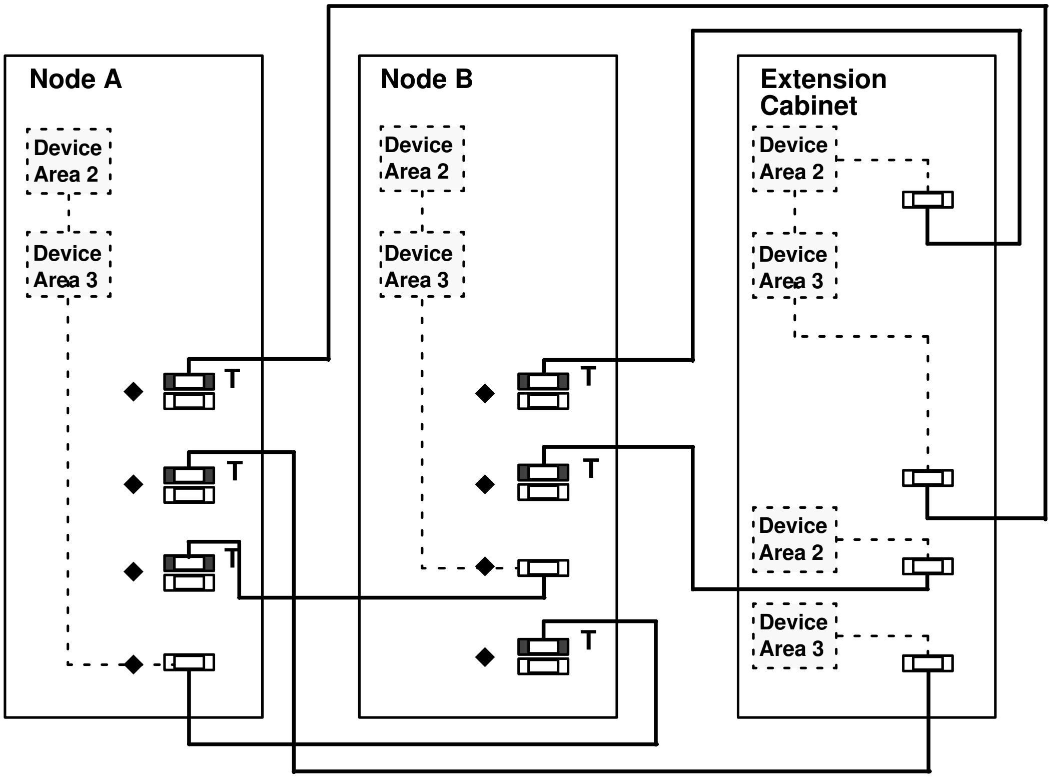
These references are liable to change.

Legend for Examples

1a, 1a', 1b, 1b'	Each node is equipped with two single-ended Ultra **SCSI adapters** (type B4-5), installed in PCI slots. Within Node A, an internal SCSI path (thick dotted line in the illustration) goes from adapter **1a** to the disks in area 3 (and possibly in area2), via the internal bulkhead (not shown in the illustration). Similarly, within Node B, an internal SCSI path goes from adapter **1b'** to the disks in area 3 (and possibly in area2), via the internal bulkhead.
T	Pass-thru terminators located at one end of external cables and connected to SCSI adapters **1a'** and **1b'** (which have no internal disks attached to them).
◆	**Wrap-plug** concerns (wrap-plugs are denoted by the ◆ symbol): • The two internal SCSI paths discussed above must be equipped with a wrap-plug to invalidate SCSI termination at the adapter level. The wrap-plug must be plugged onto the intermediate connector of the internal SCSI cable that goes from the SCSI adapter (**1a** and **1b**) to the internal bulkhead. • The two SCSI adapters **1a'** and **1b'** must also be equipped with a wrap-plug to invalidate SCSI termination at the adapter level. The wrap-plug must be plugged into the edge connector of the SCSI adapter. • For new systems, the wrap-plugs have been installed at the factory, so you do not have to worry about wrap-plug concerns. If you are upgrading existing Escala servers, you must install the wrap-plugs that come with the adapters in the add-on kit.

Native HA with Shared Expansion Cabinet

This Native HA with shared Disk Expansion Cabinet, provides a low–cost HA disk sharing solution between two hosts, with mirroring of internal and external disks over two SCSI chains.



Planning Shared DAS Subsystems

Prior Knowledge

- Please read first the “Planning Shared Disk Devices” chapter included in the *HACMP Planning Guide*.
- For an overview of the supported disk devices, also refer to “Disk Devices” on page 1-7 of the present guide.

Possible DAS Configurations (Split Bus, Dual-initiator/Dual-adapter)

Overview

Several configurations are possible for a DAS subsystem in a cluster environment. They differ in the way the buses are implemented.

In a cluster environment, two configurations (SCSI or Fibre Channel) are of particular interest:

- split-bus configuration: includes two SCSI buses, but only one bus per node.
- dual-initiator/dual-adapter configuration: includes two SCSI buses, each bus being shared by each node

Actually, **the “dual-initiator/dual-adapter” configuration provides the highest availability, and thus, is recommended**. This is because it implements two distinct shared buses.

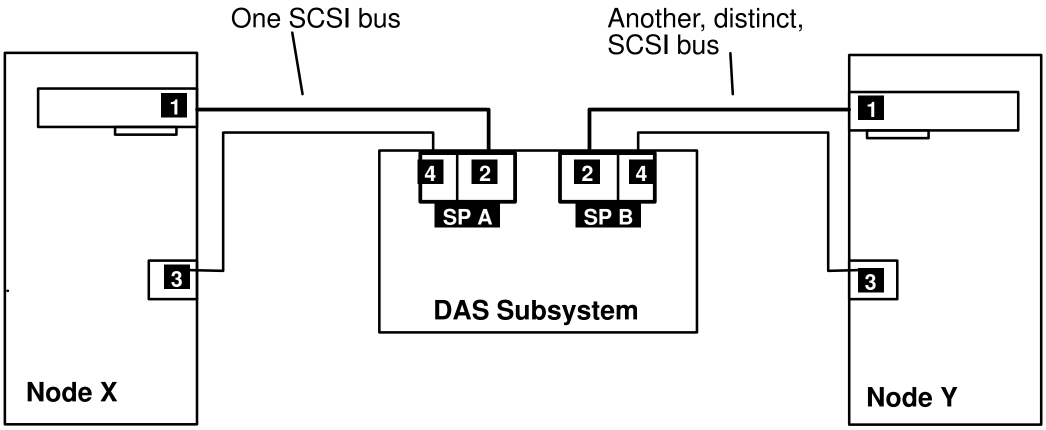
To help you in your planning task, hints are given below, that highlight important points to consider, and show some of the most typical configurations. However, the given indications should be regarded as guidelines only.

For detailed information, you must refer to the documentation that comes with the DAS subsystem. In particular, pay attention to chapter “Understanding configurations” in the book *Configuring and Managing a DAS and Using ATF*, and to chapter “About the disk-array storage system” in the book *Installing and Maintaining a DAS*.

Note: The illustrations below focus on the split-bus and dual-initiator/dual-adapter configurations, which are the most frequently implemented. Other kinds of configurations are however possible: refer to the DAS documentation.

Understanding the Split Bus Configuration (SCSI)

The figure below illustrates a split-bus configuration:



SP A, SP B	The DAS subsystem is equipped with two SPs (Storage Processors).
1, 2	Each node is equipped with a single SCSI-2 Differential Fast/Wide adapter (1), which is attached to the SCSI-2 adapter (2) of one SP. The two distinct SCSI buses are not shared by the nodes (each bus starts at a node and ends at a SP).
3, 4	Each node is connected to a DAS storage processor (SP) via serial RS-232 ports (marked 3 and 4 on the figure). This serial line is required to access DAS management and configuration functions from the node.

Highlights

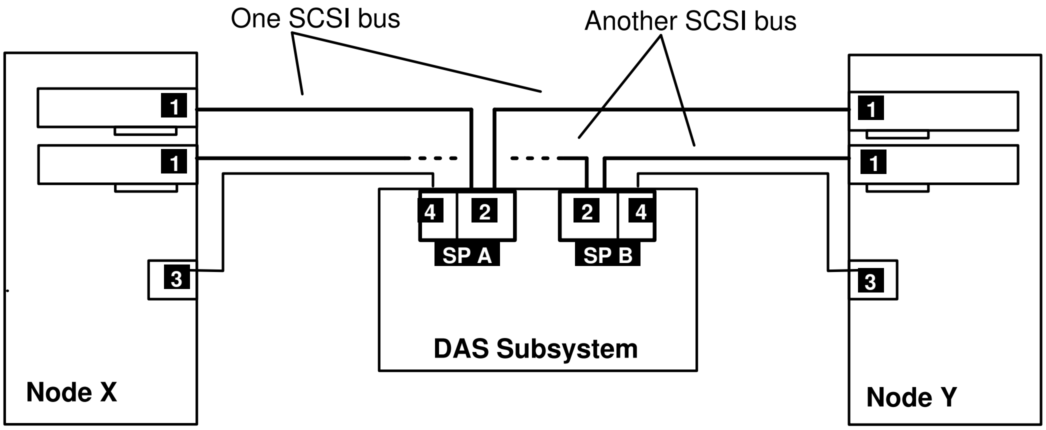
Each node has one SCSI adapter connected by a SCSI bus to an SP in the DAS. Either node can access any of the physical disk units in the DAS, but only one node at a time can access a given physical disk unit.

With the split-bus configuration, if a SCSI adapter, or SP fails, the concerned node cannot continue after failure, but the other node can continue. Thus, one node can take over the other's disks if the other node fails: the working node, through the accessible SP, is able to gain control of all shared physical disk units.

The split-bus configuration may be of interest at certain sites. However, most sites implement the dual-initiator/dual-adapter configuration which suits the highest availability requirements. See below.

Understanding the Dual-initiator/Dual-adapter Configuration (SCSI)

The figure below illustrates a dual-initiator/dual-adapter configuration:



SP A, SP B	The DAS subsystem is equipped with two SPs (Storage Processors).
1, 2	Each node is equipped with two differential Corvette SCSI adapters (1). Each is attached to the SCSI-2 adapter (2) of one SP. The two SCSI buses are shared by each node (each bus starts at one node, goes to an SP, then continues to another node).
3, 4	Each node is connected to a DAS storage processor (SP) via serial RS-232 ports (marked 3 and 4 on the figure). This serial line is required to access DAS management functions from the node.

Note: Up to two DAS subsystems can be daisy-chained in order to provide greater storage capacity.

Highlights

Each node has two SCSI adapters, each connecting by a separate SCSI bus to a separate SP in the DAS. As in the split-bus configuration, either node can access any of the physical disk units in the DAS, but only one node at a time can access a given physical disk unit.

This configuration provides very high availability, because it provides two independent routes to the DAS. It protects against any failure that may occur anywhere on a SCSI bus.

If one node, SCSI adapter, or SP fails, other node can take over failed node's disk unit. The working node is able to transfer control of all shared physical disk units to the accessible SP.

This dual-initiator/dual-adapter configuration is recommended since it provides the highest availability. This configuration requires you implement the ATF software, discussed below.

ATF Software

The ATF (Application-Transparent Failover) software must be ordered with the DAS subsystems. It must be installed on the two nodes. In the event of a failure in one SCSI path (adapters, cables, connections, or storage processor), ATF routes traffic from the failed SCSI path to the second SCSI path. ATF acts automatically and transparently, without impact on applications and operations in progress.

Note: The ATF function is totally independent, and has nothing to do with HACMP failover mechanisms. (The ATF software is a product of the CLARiiON Business Unit of Data General Corporation.)

Planning the Physical Disk Units Setup (RAID Levels, LUNs)

DAS subsystems support RAID levels 0, 0/1, 1, 3, and 5. Groups of different RAID levels can be mixed within a DAS subsystem.

- RAID levels 3 and 5 are usually the most suitable for high availability purposes.
- It is up to the system administrators, to determine which RAID level best suits their needs.

In addition to these considerations, you must plan for the LUNs configuration. A LUN (Logical Unit) is one or more disk modules bound into a single entity (note that the terms "LUN" and "physical disk unit", as used in the DAS documentation, have the same meaning). A maximum of 8 LUNs can be created.

Refer to the DAS documentation for detailed planning information and worksheets.

Planning for Cables

You should plan for the cables needed to install DAS subsystems. All the cabling elements must support the **SCSI–2 Differential Fast/Wide** standard. On any SCSI bus, the total cable length must not exceed 18 meters (25 meters being the theoretical maximum).

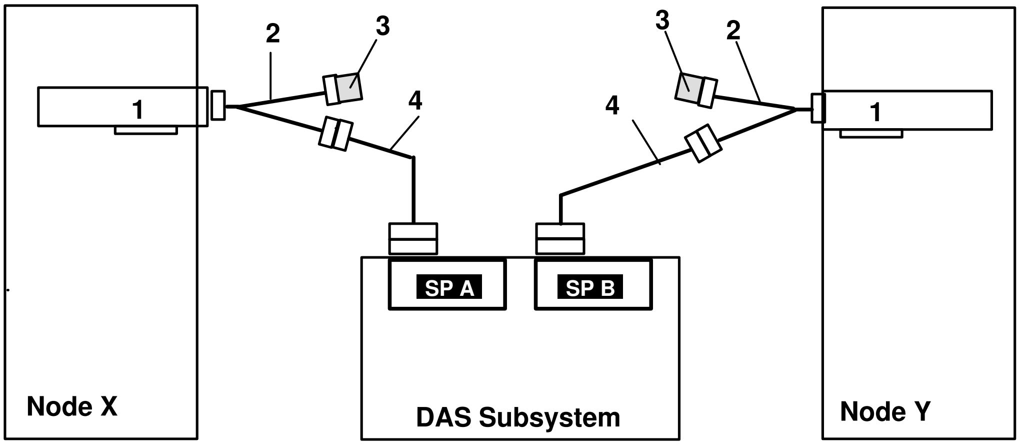
Increased-availability is implemented by using "Y" cables. Implementing such a cabling method facilitates repairs. It allows you to unplug an Y-cable without breaking the continuity of the concerned SCSI bus.

Below are presented four configurations with comments focusing on SCSI cabling concerns (RS-232 node-to-SP cables are not shown):

- cables for a split-bus configuration
- cables for a dual-initiator/dual-adapter configuration
- cables needed for chaining two DAS subsystems
- other cables needed for clusters having more than 2 nodes

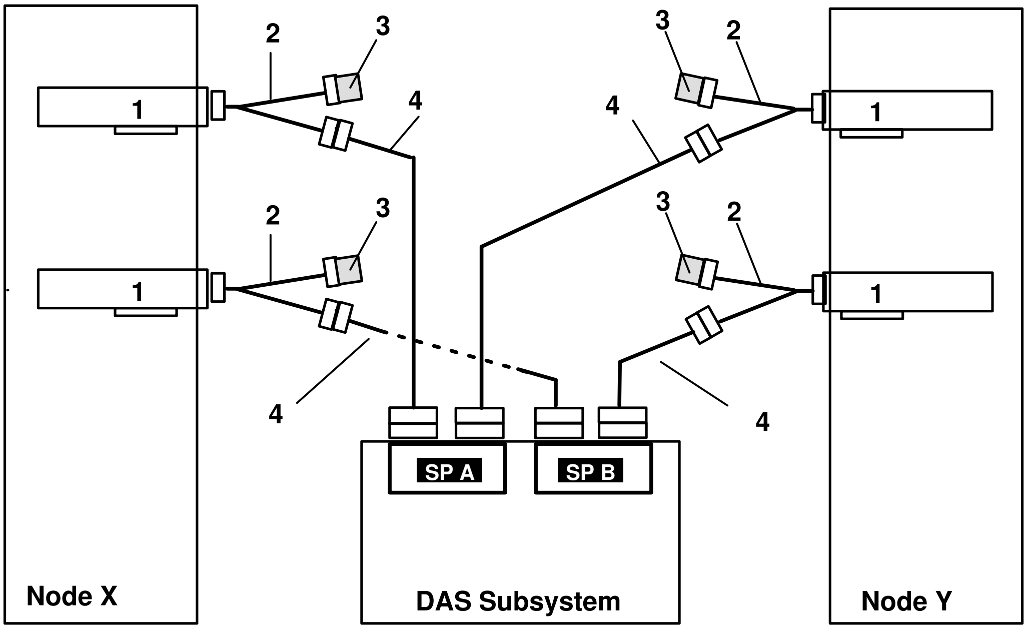
These indications should be regarded as an illustration only. For reference information (including cable length limitations), refer to the DAS documentation.

Cables for a Split-bus Configuration



1	One SCSI-2 Differential Fast/Wide adapter per node. Their internal terminator must be removed.
2	Y-cables (differential).
3	Terminators (differential). They are attached to the shorter legs of the Y-cables.
4	Controller-to-DAS cables (differential).
*	Here, two RS-232 cables (not shown on the figure) are needed to connect each node to a SP. For serial port concerns, refer also to page 3-3.

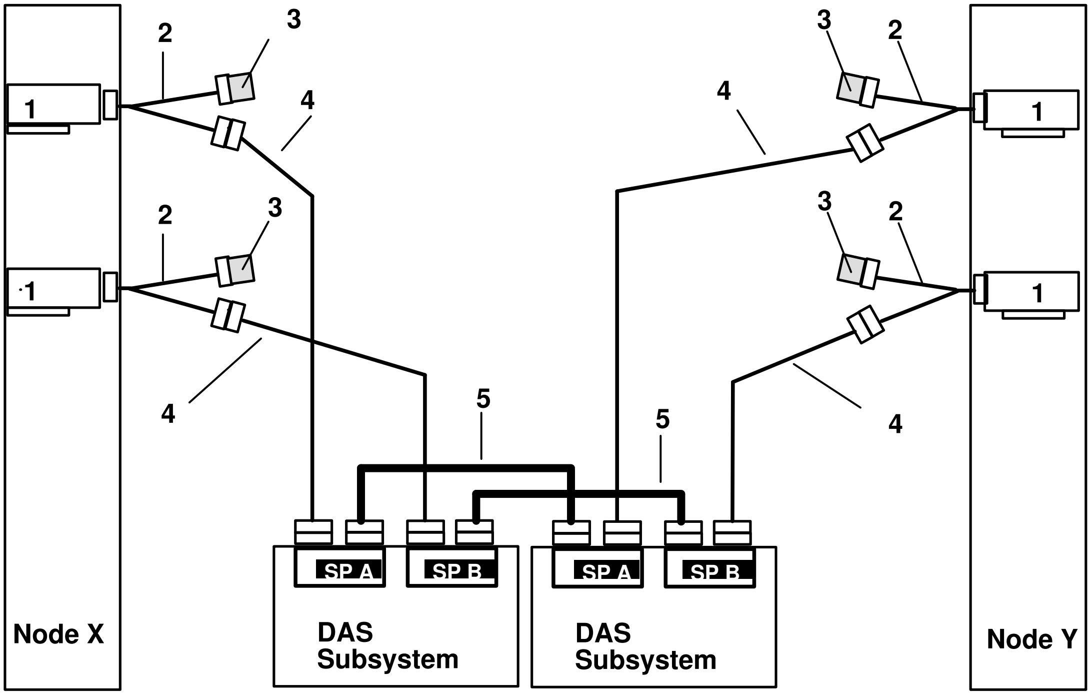
Cables for a Dual-initiator/Dual-adapter Configuration



1	Two SCSI-2 Differential Fast/Wide adapters per node in order to provide the two distinct SCSI shared buses. Their internal terminator must be removed.
2	Y-cables (differential).
3	Terminators (differential). They are attached to the shorter legs of the Y-cables.
4	Y-cable-to-DAS cables (differential).
*	Here, two RS-232 cables (not shown on the figure) are needed to connect each node to a SP. For serial port concerns, refer also to page 3-3.

Cables Needed for Chaining Two DAS Subsystems

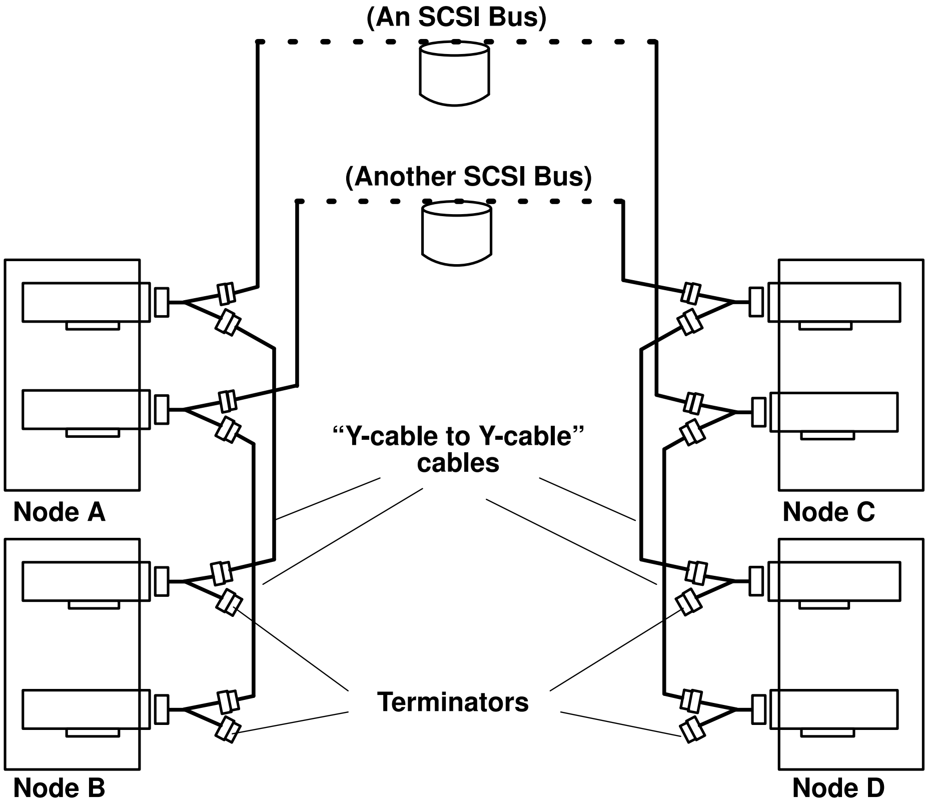
Up to two DAS subsystems can be daisy-chained in order to provide greater storage capacity. The figure below illustrates a dual-initiator/dual-adapter configuration with two chained DAS subsystems (refer to the DAS documentation for further information).



1	Two SCSI-2 Differential Fast/Wide adapters per node in order to provide the two distinct SCSI shared buses. Their internal terminator must be removed.
2	Y-cables (differential).
3	Terminators (differential). They are attached to the shorter legs of the Y-cables.
4	Y-cable-to-DAS cables (differential).
5	DAS-to-DAS cables (differential), represented with thick lines on the figure. They are used to daisy-chain the two DAS subsystems.
*	Here, four RS-232 cables (not shown on the figure) are needed. Each node is connected to two SPs, one on each DAS. For serial port concerns, refer also to page 3-3.

Clusters having more than 2 nodes

If the cluster is equipped with more than 2 nodes, the same cables are used. In addition, however, “Y-cable to Y-cable” cables are required to cable the SCSI path between two neighboring nodes. See the figure on page 3-17, which depicts a sample 4-node cluster, focusing on the needed “Y-cable to Y-cable” cables and on the appropriate placement for the terminators.

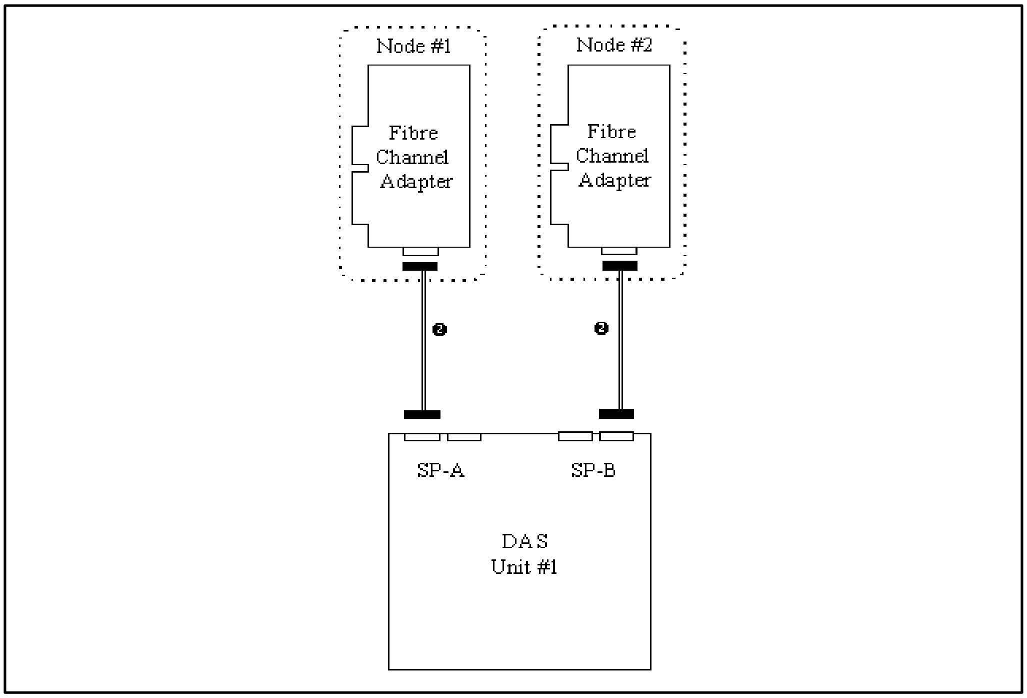


The figure above depicts the SCSI cabling for a sample 4-node cluster. It focuses on the needed "Y-cable to Y-cable" cables, and on the appropriate placement for the terminators. (Regarding these issues, the scheme applies whether the shared disks are DAS subsystems or ordinary SCSI disks, thus the disk devices are not represented).

Possible DAS Configurations with Fibre Channel

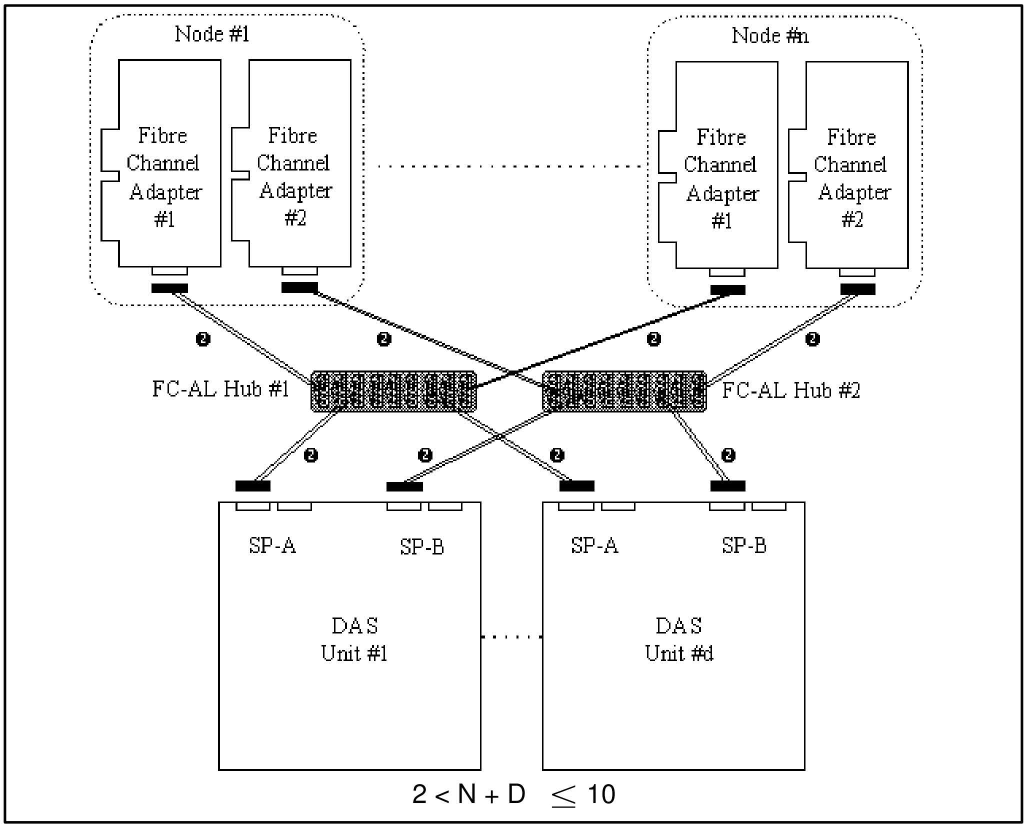
Dual Loop

With dual initiators (2 nodes) using 1 DAS with 2 SPs.



More Than 2 Nodes

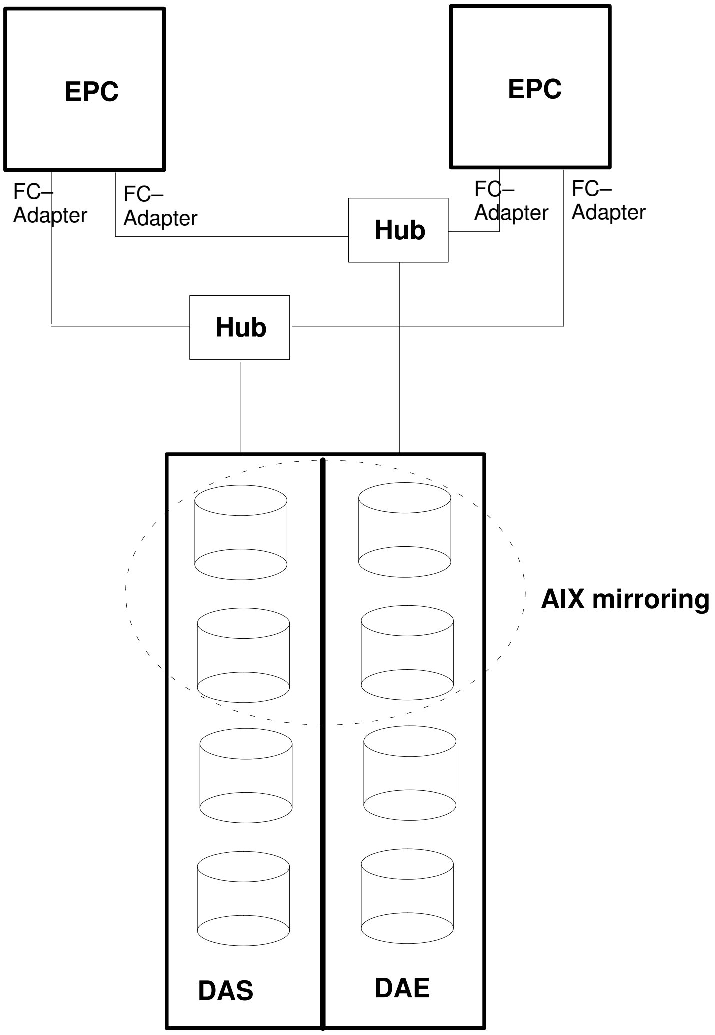
With dual loops, 2 hubs, N nodes, using D DAS with 2 SPs.



Incorporating Disk-Array Enclosures (DAE)

The high performance, high capacity disk–array storage system using a Fibre Channel Arbitrated Loop (FC-AL) can be used to provide High Availability features between two Escala Power Cluster Nodes.

The figure below shows the configuration using two EPC nodes.



With this configuration, data is mirrored between the two DAEs to ensure high availability.

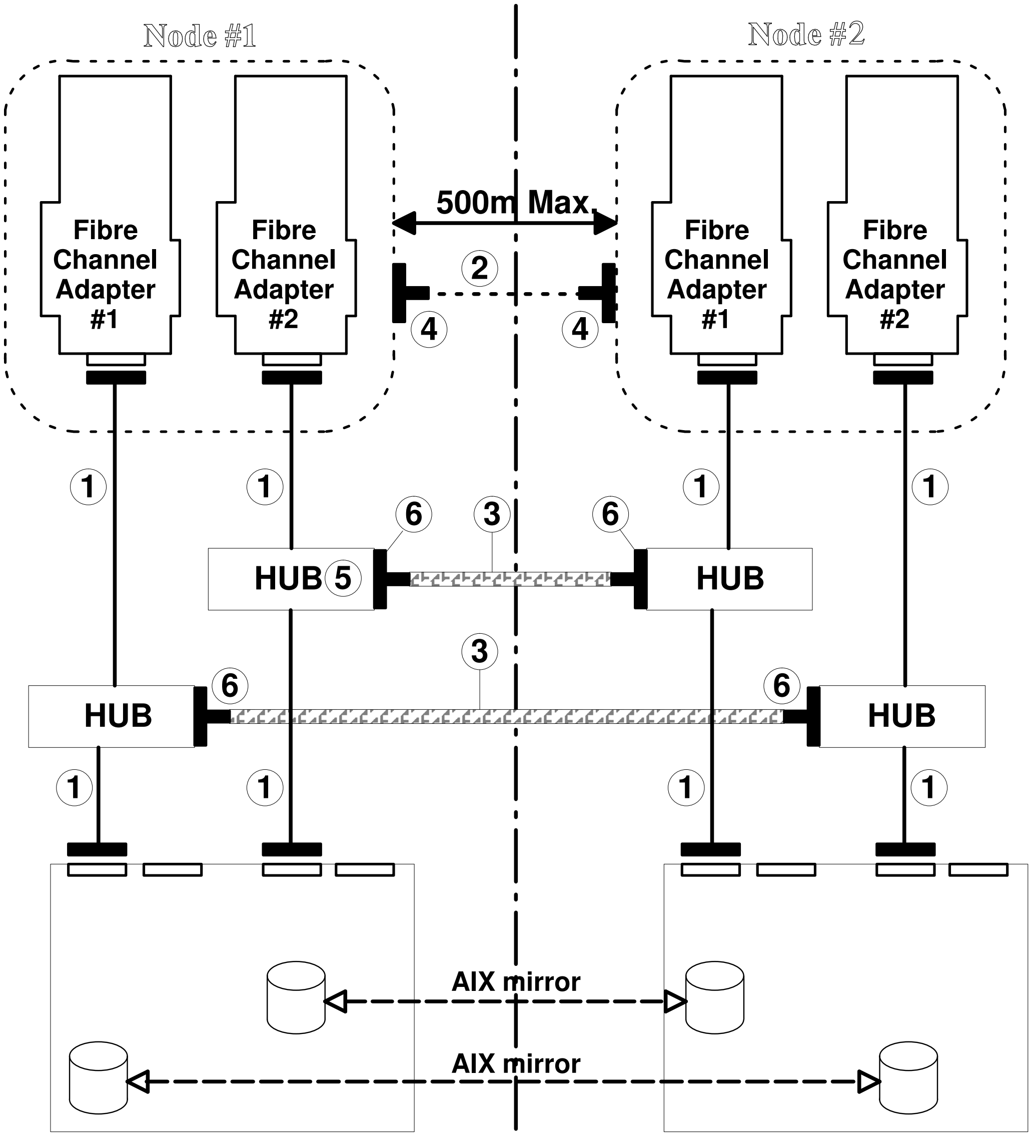
Planning Extended HA Shared Storage for Escala HA Solutions

Extended High Availability is the Disaster Recovery solution proposed by Bull. It is based on fiber subsystem technology to allow shared disk connectivity over 500 meters. AIX mirroring capability is used to assure mirroring of data.

The HA features are the same as for local clusters, with some precautions, see page 10-22.

Cabling

Set up the FC–AL loops and the extended serial link (HACMP keep-alive heartbeat) as shown in the figure. No interconnect is available for this configuration.



1	Copper Cable
2	RS232 Extended Cable (keep-alive heartbeat)
3	Optical Fiber Cable
4	Micro-modem for RS232
5	FC-AL Hub
6	MIA Media Interface Adapter (copper to optical fiber)

Shared FC–AL Loops Concern

Set up loop ID on each SP to be unique in the two fibre DAS systems.

Power on the HUBs and the two DAS systems, and wait until that the service LED on each DAS comes on and goes off, before you power on your Escala servers.

Planning the Physical Disk Units Setup

Configure the two DAS systems as described in the DAS documentation: bind Luns and make the disks available to the systems.

Note: With Fiber DAS systems 32 LUNs can be created and accessed by each SP.

Planning Shared SSA Disk Subsystems (Serial Storage Architecture)

Note: For an overview of SSA (Serial Storage Architecture), refer to page 1-10 of the present guide.

When planning for SSA Disk Subsystems, refer to the following reference documents:

- the "Planning Shared Disk Devices" chapter included in the *HACMP Planning Guide*,
- the documentation of the SSA adapters and SSA disk subsystems.

In addition, you may find useful the hints below, which supplement these reference documents.

About SSA Adapters

MCA Bus

Two SSA adapter models are available:

SSA 4-Port Adapter (Type 4-D) (MI = MSCG021)

When this adapter model (labeled with the "4-D" code) is used, no more than two SSA adapters can be connected in an SSA loop.

Enhanced SSA 4-Port Adapter (Type 4-G) (MI = MSCU-101)

When this adapter model (labeled with the "4-G" code) is used, up to 8 SSA adapters can be connected in an SSA loop. Thus:

- The enhanced adapter is required if the cluster includes more than 2 nodes on one SSA loop.
- In addition, for scalability, always consider the use of enhanced adapters (even if the cluster includes only 2 nodes). Indeed, this makes provision for possible future cluster upgrades.

SSA Multi-Initiator / RAID Enhanced Loop Adapter (4-M) (MI = MSCU-038)

No more than two SSA adapters can be connected in an SSA loop. However, in RAID configuration, this adapter works in mono-initiator mode only.

PCI Bus

One SSA adapter model is available:

SSA Multi-Initiator / RAID Enhanced Loop Adapter (4-N) (MI = MSCG-039)

No more than two SSA adapters can be connected in an SSA loop. However, in RAID configuration, this adapter works in mono-initiator mode only.

Note: The mix of PCI and MCA adapters on a unique SSA loop is allowed between PCI SSA Multi-Initiator (4-N) and MCA SSA 4-Port (Type 4-M) adapters only.

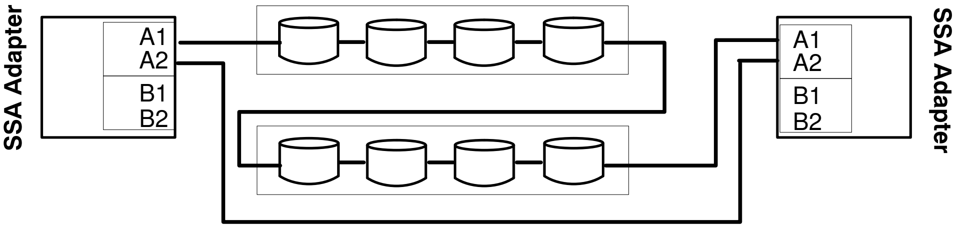
About SSA Topology

When planning for SSA cabling, keep in mind the hints below:

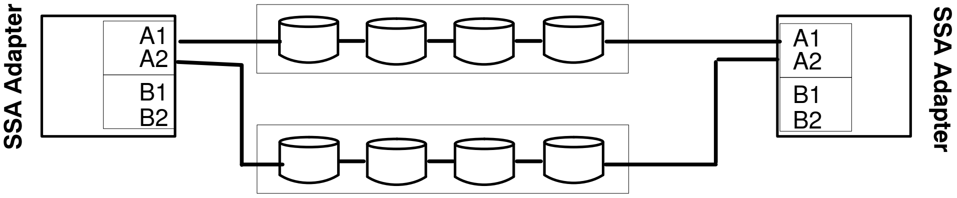
- The loop topology is best suited to a cluster environment. Point-to-point string topology is not appropriate.
- Each SSA adapter has 4 ports, and thus, can be connected to one or two distinct loops.
- For high availability purposes, disk mirroring must be implemented. The mirroring function is to be implemented through the standard features of the AIX system.

  If you decide to cable two SSA loops (this provides better availability), mirroring should be implemented across the two loops (disks on one loop should be mirrored to disks on the other loop). In addition, availability is further increased if the mirror copy is accessed through two adapters.

- For performance purposes, avoid connecting adapters adjacent to each other in a loop. The figure below illustrates this recommendation. In the topology shown at the top, two adapters are adjacent in the loop. At the bottom, is shown the recommended topology, where no adapter-to-adapter connection is made. This second topology offers slightly better performances, because shorter paths from adapters to disks are provided.



The topology above works, but reduces the possible throughput.
The topology below offers slightly better performances.



Sample SSA Topology

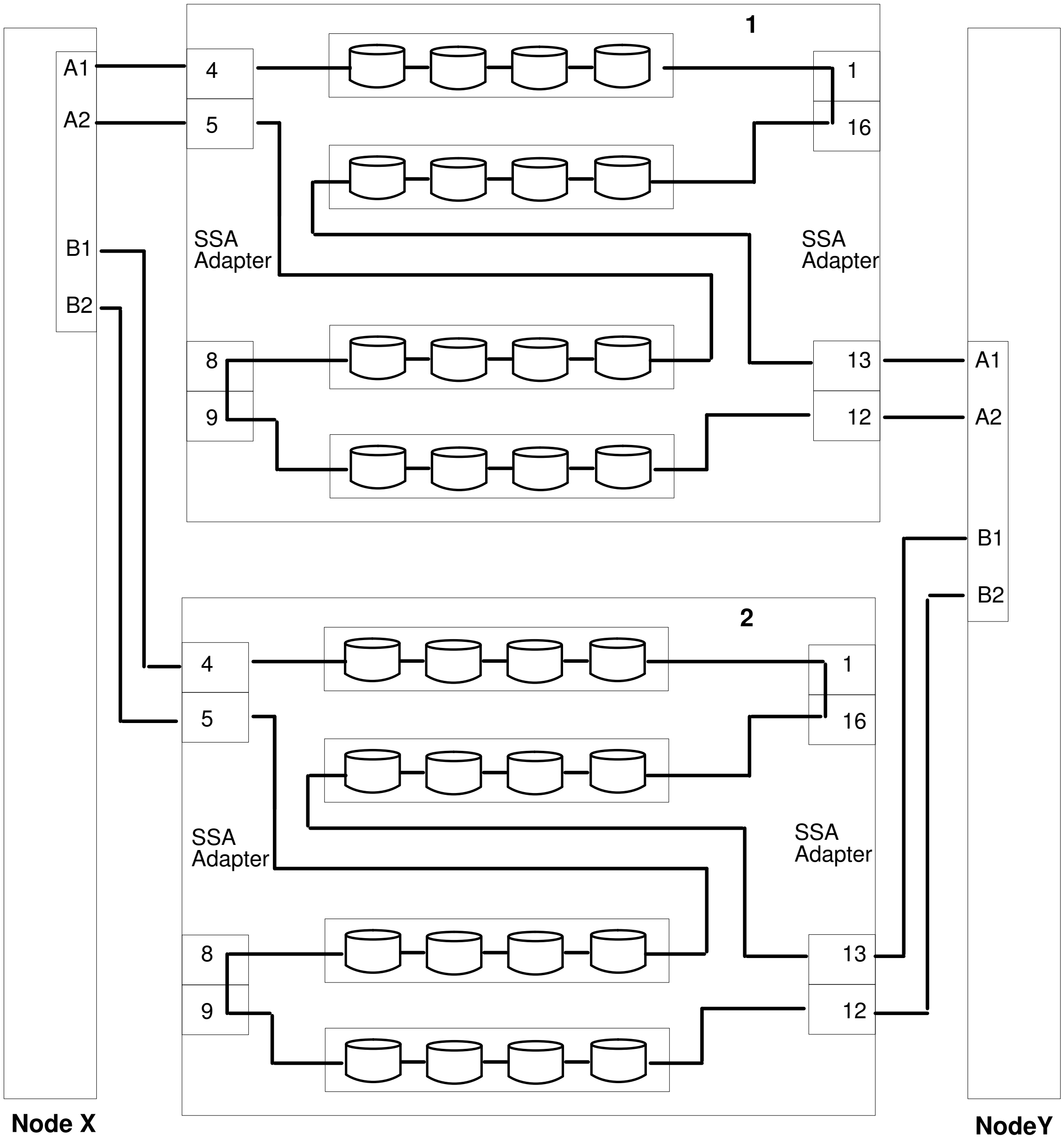
As said above, for configurations with two SSA loops, mirroring should be (as far as possible) implemented across the two loops. In practice, however, designing an appropriate topology requires proper planning, and thought needs to be given to the problem. As an illustration, we present below a case which is not very easy to grasp.

Sample Topology with One SSA Adapter per Node

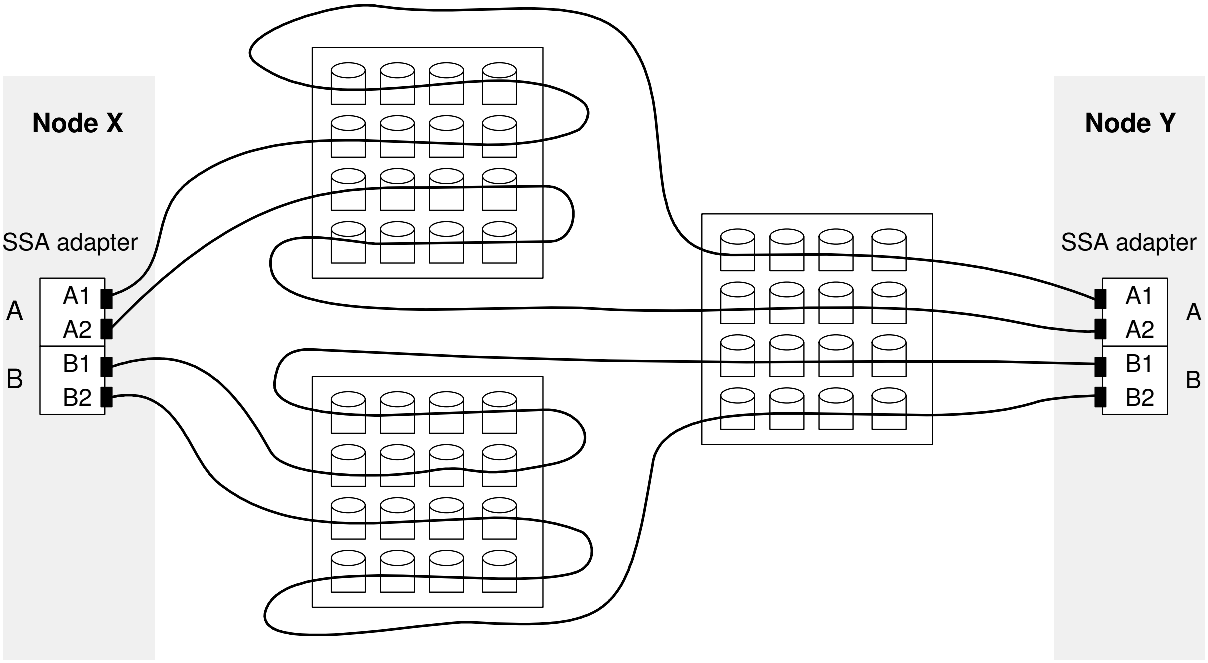
Suppose that you want to implement cross-loop mirroring, assuming that you have the following equipment:

- the cluster consists of two nodes,
- each node is equipped with a single SSA adapter,
- SSA subsystem models can accommodate up to 16 disks, 32 disks can be connected in a dual–loop configuration (16 disks on one loop are mirrored on 16 disks on the other loop),
- 8 adapter–to–subsystem SSA cables.

The figure below shows a typical example.



2 Nodes with 2 Loops



Topology Variant with Two SSA Adapters per Node

If the nodes are each equipped with two SSA adapters, even higher availability can be achieved by implementing not only cross–loop mirroring, but also cross–adapter mirroring (a loop being implemented using one adapter on each node, and the other loop using the two other adapters).

Planning Shared EMC Symmetrix Subsystem Disks

The EMC Symmetrix is a disk subsystem that houses all storage control functions in a single cabinet. It is composed of:

- a dual internal bus,
- Channel Directors (CD) that manage the host links and storage control functions,
- cache memory cards,
- Disk Directors (DD) that handle the data storage functions,
- power, battery subsystems and cooling modules,
- integrated service processor (lap top PC); downloads the Symmetrix configuration to the directors and provides diagnostics and maintenance utilities for Symmetrix.

The channel directors are available in two versions which can be mixed in the same Symmetrix subsystem:

- Ultra Wide SCSI directors containing four host ports
- Fibre Channel directors containing two host ports

The disk director provides an interface between cache and the disk devices, and manages four Fast Wide SCSI buses to the internal disks. Depending on the model, either four or six disks are connected to each each internal SCSI bus.

The attachment of a Symmetrix Fast Wide Differential SCSI port to an AIX server is made using standard MCA or PCI adapters.

The attachment of a Fibre Channel port or a Fibre Channel director to an AIX/PCI server is made through a Bull Fibre Channel adapter. The Bull Fibre Channel AIX driver supports Symmetrix FC-AL subsystem. The Symmetrix Fibre Channel port uses a fiber optic connector and therefore, connections to Symmetrix must be made using a fibre optic cable with an MIA installed at the other end of the cable. The copper side of the MIA is connected to an Emulex adapter.

The following disk features are provided by Symmetrix:

- RAID-1
- RAID-S which is a high performance RAID level 5 protection for data
- Dynamic sparing

Base Configuration with HACMP

The usual HA configuration with Symmetrix subsystems is to have a point to point connection (called split bus configuration in this document) for each server, and to configure the Symmetrix subsystem in order to make the data volumes available to both servers through the two separate host ports.

For the Oracle Parallel Server model, the storage volumes must be shared by all the nodes of the cluster.

The usual CMP configuration with Symmetrix subsystems is to have a point to point connection between all servers and the Symmetrix subsystem, and to configure the Symmetrix subsystem in order to make the data volumes available to both servers through all the separate host ports.

EMC Disk Subsystems in Concurrent Mode

In EMC Symmetrix subsystems disks can be used in concurrent mode if you perform the following command on each hdisk used in this mode:

```
chdev -l hdisk<x> -a 'reserve_lock=no'
```

Chapter 4. Setting Up the Basic Hardware

This chapter explains these steps:

- Physically Install the Servers and the Disk Units
- Set up the Hardware for System Console Handling, on page 4-2.
- Attach Network Connections, on page 4-2.
- Boot Each Node and Check for Proper Operation, on page 4-4.
- Configure TCP/IP For Boot/Service Network Adapters, on page 4-4.
- Configure the Adapter SCSI IDs (and Reboot the Concerned Nodes), on page 4-7.

Follow the instructions step by step, then proceed with the next chapter.

Prior Knowledge

- This chapter gives only guidelines. For instructions that are not specific to cluster configurations, rely on the documentation that comes with your hardware.
- Before proceeding with the hardware installation, we recommend you review the *HACMP Planning Guide* as well as the chapter “Supplemental Planning Information”, starting on page 3-1 in the present guide.

Physically Install the Servers (Nodes) and the Disk Units

General Instructions

- Unpack the servers and the external disk units (if any).
- Determine their physical location on the installation site, keeping in mind the various physical constraints: adequate space for maintaining and servicing the hardware, cable length limitations, paths of power cords, and so on.
- Place the units on the designated location.

CAUTION:
Do not connect now the external shared disk devices to the nodes!
(This is because SCSI IDs and hardware configuration of external SCSI adapters are not necessarily set appropriately).

- For the moment, do not power on the nodes and the devices.

Instructions Specific to Powercluster Models

If you are installing a Powercluster (i.e. a cluster based on Escala Rack models), physically install the devices in the rack now. However, as stated above, do not connect now the shared DAS RAID subsystems to the nodes.

Depending on the Powercluster model you are setting up, one or two racks are to be installed. If you are installing a configuration that comprises several racks, consider that, from the power supply stand point, the racks are totally independent. That is, you need one distinct AC wall outlet per rack.

Set Up the Hardware for System Console Handling

To handle the consoles of the different cluster nodes, several solutions are offered, based on different hardware components.

- The ordinary solution consists in attaching one terminal to each node. This is the solution proposed for HA Solutions. If you are about to implement this console solution, proceed with the "Procedure Using Standard Consoles" below.
- If you are about to implement another console solution (a single terminal for all the nodes, or solutions based on X terminals or the PowerConsole), please refer to appendix "Implementing Console Solutions", starting on page A-1.

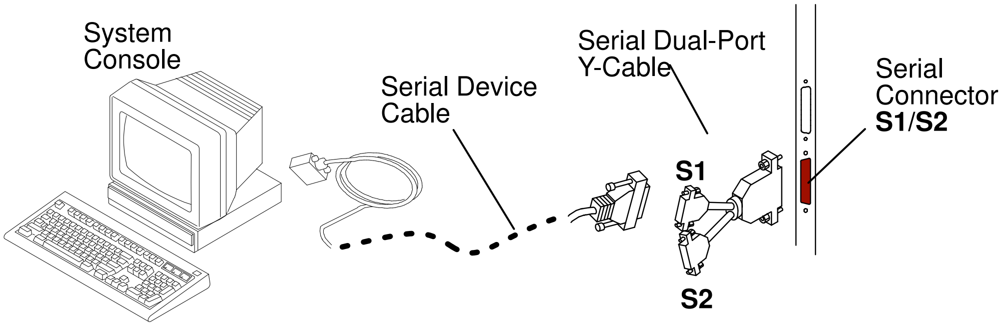
Procedure Using Standard Consoles

This procedure applies to traditional configurations that make use of one system console (one terminal device) per cluster node.

Attach a system console to each node. Use the first serial port of the node, which is usually labeled "S1" or "async L1".

Notes:

- Escala Minitower models are equipped with a serial dual-port. In this case, a single connector, usually labeled "S1/S2", provides two ports through a serial dual-port Y-cable. Use this cable as shown in the figure below.



Attach Network Connections

Attaching Public Network Connections

Each node is equipped with two network adapters for each connected network: one adapter serves as the service adapter, and the other as the standby adapter. One service/standby pair should be connected to each distinct network: see the illustration below.

Connect the adapters to the network(s). If needed, refer to your hardware documentation for further explanation of network (Ethernet, FDDI, or ATM) connection concerns.

Attaching Private Network Connections

Private networks concern only clusters that implement concurrent access or client/server architectures.

If you implement a private network, which may be Ethernet or FDDI, make the appropriate connections.

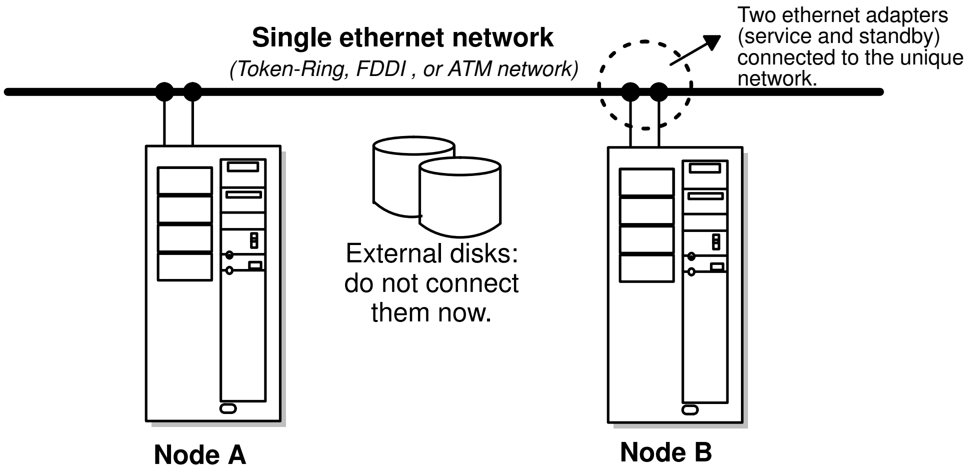
Notes:

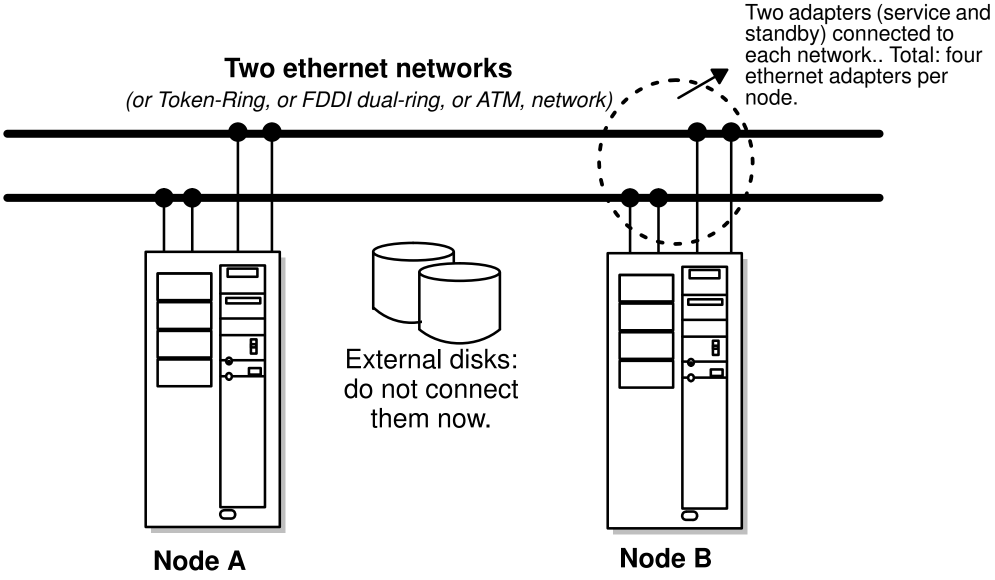
- Refer to section “Planning Networks”, on page 3-4, for background information related to public and private networks. As a reminder note that:
  - Concerning the public network, basic cluster configurations use only one Ethernet network: this is a common configuration. However, for a higher availability, the nodes in the cluster can be connected by multiple networks (two networks, typically).
  - If you use FDDI for the private network, you may want to implement a dual-homing configuration.
- For details on Ethernet and FDDI network cabling concerns, also refer to your hardware documentation.
- For implementing a private network based on Fibre Channel, please contact your service representative.

Your Basic Hardware Configuration Now Resembles the Following

The figure below depicts two cases. At the top, a single Ethernet network is used, and each node has two network adapters (this is a common configuration). At the bottom, two distinct Ethernet networks are used, and each node has four network adapters (this configuration offers higher availability).

Note: Ethernet is just used as an example to illustrate the instructions (if you use FDDI or ATM to implement your networks, adapt the examples to your situation).


Single ethernet network
(Token-Ring, FDDI , or ATM network)
Two ethernet adapters (service and standby) connected to the unique network.
External disks: do not connect them now.
Node A
Node B


Two ethernet networks
(or Token-Ring, or FDDI dual-ring, or ATM, network)
Two adapters (service and standby) connected to each network.. Total: four ethernet adapters per node.
External disks: do not connect them now.
Node A
Node B

Boot Each Node and Check for Proper Operation

Power on each node and check that they boot properly.

Notes:

- For the moment, the current configuration is not a cluster. This step is just a basic check to see if the basic hardware seems OK before proceeding with subsequent cluster installation steps.
- If the cluster is equipped with a limited number of system consoles, it is more practical to first boot the node(s) which has(have) a console attached. Then, from here, you can use special commands to connect to the S1 serial ports (which are dedicated to system console handling) of nodes that have not a console directly attached. These special commands are described in appendix “Implementing Console Solutions”, starting on page A-1.

Configure TCP/IP for Boot/Service Interfaces

Note: The instructions below concern the adapters connected to the public network.

This step consists in using **mktcpip** on each node to define the IP label, IP address, and network mask for the network adapter that will act as the boot and service interface. For the moment, it is useless to deal with the standby adapter, which will be configured through the quick configuration tool.

This step assumes you are familiar with concerns that relate to HACMP and networking. In particular, it assumes you are familiar with concepts such as boot/service/standby adapters, IP labels, IP addresses, network masks, and so on. For a reminder, refer to “Prior Knowledge” below; otherwise, directly skip to “Naming Conventions”.

Prior Knowledge: IPAT / Boot, Standby, and Service Addresses

For detailed information on the topics discussed here, refer to the HACMP documentation.

Discussion on IP Address Takeover

Generally, you want to enable your cluster configuration for IP Address Takeover (sometimes referred as IPAT). IP Address Takeover consists in moving the service IP address of a node that fails to a surviving node. With this feature enabled, clients reconnect to the same IP address to communicate with their application (assuming that the concerned application has also been taken-over by the surviving node).

Note, however, that IP Address Takeover is not necessarily a required feature, and you may want or not to enable it, depending on your specific needs. If you do not enable IP Address Takeover, then, if a node fails, its resources are taken-over by a surviving node, but its IP service address is not taken over. Thus, the clients can no longer use this IP address to communicate with their application. They must use another address, namely the service IP address of the node that has taken over the concerned application.

Boot Standby and Service IP Addresses

Cluster nodes are usually equipped with two network adapters: a service adapter and a standby adapter (we do not discuss here configurations where each node is connected to two distinct networks, and is equipped with four adapters). Below are given some reminders related to the involved IP addresses.

- If you choose to not enable IP Address Takeover for a node (if you do not require that its IP address can be taken over by another node), then the service adapter of the node always assume the service IP address (i.e. the notion of boot IP address makes no sense for this node).

- If you choose to enable IP Address Takeover for a node, then the notion of boot IP address intervenes. Boot and service IP addresses are assumed by the same network adapter, called the service adapter:
  - At boot time, and until the cluster services are not started, the service adapter assumes the boot IP address. Thus, the address is used only temporarily.
  - When the cluster services are started, the IP address on the service adapter is switched from the boot address to the service address. From now on, the service adapter assumes the IP service address. Thus, the IP service address is the "normal" address that is used by clients to communicate with the applications when HACMP is running.
- The IP standby address is assumed by a second network adapter, called the standby adapter. The standby adapter is here as a back up when the service adapter on this node fails or when another node fails. In such an event, the standby adapter is re–configured to assume the IP service address involved in the failure.

  Note that **you should consider that the standby adapter is reserved for HACMP takeover operation**. It is not intended for handling connections of applications that do not participate in cluster resources (i.e. that do not pertain to a HACMP resource group).

Naming Conventions

According to the naming rules used by the *Bullcluster quick initial configuration* tool, the `_boot` and `_stby` strings are always used as the standard suffixes for the names of the boot, and standby adapters.

For example, assume you plan to implement a cluster with two nodes having the names `foo` and `bar`. The *Bullcluster quick initial configuration* tool will carry out automatically many setup tasks, and in the resulting configuration:

- The `foo` node will be configured has having the three following interfaces:

foo_boot	*(boot)*
foo	*(service*)
foo_stby	*(standby)*

- Similarly, the `bar` node will be configured has having the three following interfaces:

bar_boot	*(boot)*
bar	*(service)*
bar_stby	*(standby)*

Note that you can use the *Bullcluster quick initial configuration* tool whether or not you plan to enable IP Address Takeover. If you do not enable IP Address Takeover, the nodes are configured has having only two interfaces (service and standby), for example `foo` and `foo_stby`, since boot IP addresses do not make sense.

Applying the naming conventions is:

- mandatory if you plan to use the *Bullcluster quick initial configuration* tool
- optional if you plan to not use the *Bullcluster quick initial configuration* tool

Procedure

For each node, follow the procedure below:

1. Being root, enter:

   ```
   smit mktcpip
   ```

   The **Available Network Interfaces** screen appears, listing interfaces on the node available for network communications.

2. From the list, select the adapter you plan to use as the service adapter.

   The **Minimum Configuration & Startup** screen appears. Its contents depends on the type of network adapter you selected (ethernet or FDDI).

3. Go to the **HOSTNAME** field. If you intend to use the *Bullcluster quick initial configuration* tool for setting up your cluster, you must comply with naming conventions explained above (see "Naming Conventions" above).

   – If you plan to enable IP Address Takeover, enter as the **HOSTNAME** the name of the boot interface, such as:

     `nodename_boot`

     Where:
     – `nodename` is the name of the node, i.e. the name you will subsequently specify when configuring the cluster.
     – The `_boot` string is, by convention, a suffix appropriate for use by the *Bullcluster quick initial configuration* tool.

   – If you plan to not enable IP Address Takeover, enter as the **HOSTNAME** the name of the service interface, such as:

     `nodename`

4. In the **Internet ADDRESS** field, specify the IP address to be assigned to this interface (this is the boot or the service interface, depending on the case).

5. If needed, set the **Network MASK** field in accordance with your configuration.

6. Specify the **Your CABLE** type field as appropriate.

7. Press Enter to apply these settings.

8. Do not forget to perform this procedure **on each node** you are setting up.

Notes

- When you apply the settings, the default host name of the node becomes the name you have specified in the **HOSTNAME** field. Thus, if you have specified `foo_boot`, the host name of the node is now `foo_boot`.
- In addition, an entry is automatically added to the /**etc**/**hosts** for the specified host name.
- It is important to realize that a hostname (in the usual AIX sense) is not the same as a node name. A hostname is the same as a node name in HACMP only.

Important Remark About Hostnames

In a subsequent configuration step, when using the *Bullcluster Quick Initial Configuration* tool, this tool automatically configures other interfaces. If you choose to enable IP Address Takeover, the tool carries out configuration tasks so that the service IP address and the service adapter are set up, as appropriate. **The quick configuration process also changes the hostname, so that it matches the service interface**. For example, the hostname `foo_boot`, is changed in `foo`. This scheme is usually desirable (you rarely want a hostname that matches the boot interface).

If you choose to not enable IP Address Takeover, the quick configuration process does not change the hostname (for example, the initial name `foo` remains `foo`).

Configure the Adapter SCSI IDs

Note: This section applies only if you are installing a cluster where shared disk storage is provided by SCSI-based disk devices. If you use SSA disk subsystems, skip to the next chapter.

Prior Knowledge

Understanding the SCSI ID Concerns

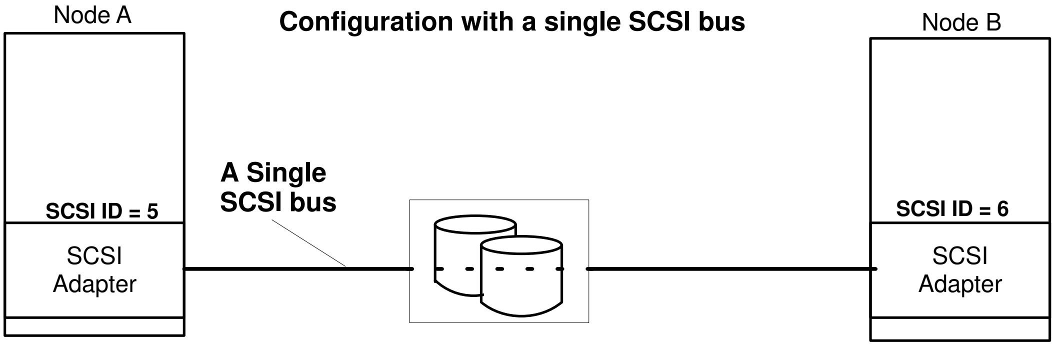
In a cluster configuration, nodes share disk devices via external SCSI buses. Each shared SCSI bus has a node adapter at both ends. Such a configuration, where two servers (initiators) are attached to the same SCSI bus, is sometimes referred as “dual head of chain”.

Regarding the SCSI ID concerns, consider the following:

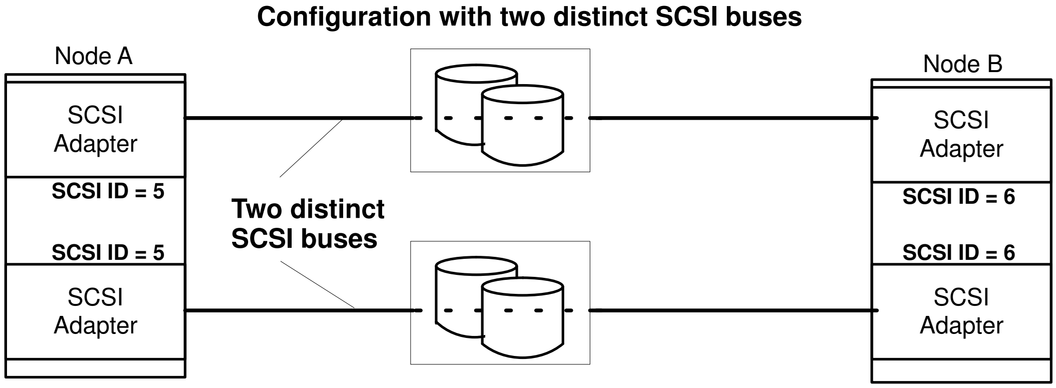
- Each device on an SCSI bus must have a unique SCSI ID.
- The ID 7 must not be assigned to SCSI adapters that equip the nodes (this is because, if diagnostics are run, they force the SCSI ID of the adapter to 7, and thus, in this event, an address conflict would occur). Unfortunately, the default SCSI ID for the node adapters is 7.

Consequently, to avoid duplicate SCSI IDs on the same bus, and to ensure that the ID 7 is not used by any node, you must change on all nodes the SCSI ID of the external SCSI adapter(s).

Note: In some configurations, the cluster is configured with a single shared SCSI bus. For a higher availability, however, typical configurations include two shared SCSI buses. The figure below illustrates these two cases.



Configuration with a single SCSI bus



Configuration with two distinct SCSI buses

To check or to set up the SCSI IDs, perform the procedure below.

Set Appropriately the Adapter SCSI IDs

Note: If you have followed the installation procedure explained in this guide, the external shared disk devices are still not connected to the nodes. SCSI cabling is to be performed in a subsequent step, as explained later in this guide.

Configuration Procedure

To configure the adapter SCSI IDs, do the following:

1. On one node, log on as root and enter:

   ```
   smit chgscsi
   ```

2. The list of installed adapters is displayed. From the list, choose an adapter whose SCSI ID is to be changed.

   a. If you are setting up a "Corvette" adapter (type "4-6"), choose the appropriate adapter name (which begins with the `ascsi` prefix, and is labeled "Wide SCSI I/O Controller Adapter").

   b. If you are setting up a "WSA" adapter (type "WSA1"), choose the appropriate driver name (which begins with the `scsi` prefix, and is labeled "Wide SCSI Adapter SCSI Driver").

   c. If you are setting up a PCI-based SCSI adapter (type "B4-5" and type "B4-6" adapters used on Escala servers), choose the appropriate adapter name (which begins with the `scsi` prefix, and is labeled "Wide/Fast-20 SCSI I/O Controller").

3. A new menu is displayed. Examine the valued indicated in the **External SCSI ID** (or **Adapter card SCSI ID**, depending on the adapter type) field. If the value is not appropriate (this is not necessary the case, because some configurations are appropriately set up at the factory), specify the desired SCSI ID, keeping in mind that:

   – The ID 7 must not be used for the node adapters.

   – Each device on the SCSI bus must have a unique SCSI ID.

   – It is recommended to reserve higher IDs for the node adapters, and use lower IDs for the external devices.

4. Specify **yes** in the **Apply change to database only** field. (Leave the other fields as they are).

5. Press Enter to apply the specified settings.

6. If your cluster uses two SCSI buses, repeat all the steps above (on the same node) to set the SCSI ID of the other SCSI adapter on the other bus.

7. Repeat all the steps above on the other cluster nodes.

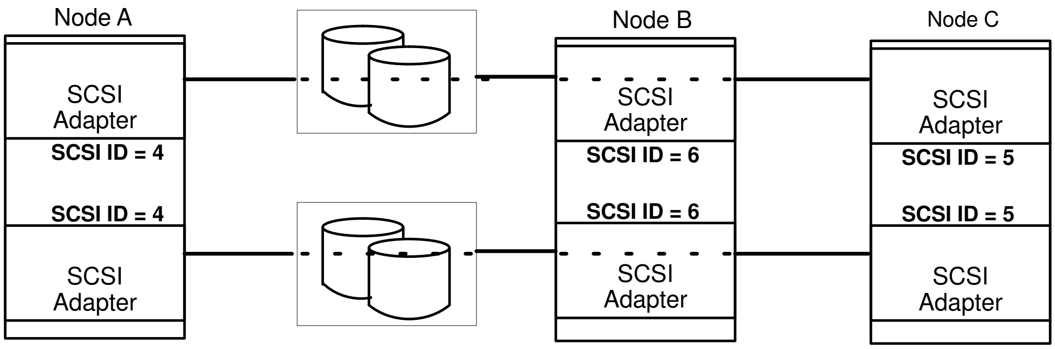
Note about clusters that include more than two nodes:

If your cluster includes more than two nodes, the considerations above apply (SCSI IDs must be unique on each SCSI bus). This situation requires you to change adapter SCSI IDs appropriately on several nodes (see the figure below).

In this case, repeat as needed the whole procedure above (**smit chgscsi**) for other cluster nodes that share SCSI buses with several nodes. Since SCSI IDs 6 and 5 are already in use (assigned to the adapters on the two first nodes), choose 4 for the next concerned node (and so on, choose the highest available unique ID).

CAUTION:
Do not connect now the external shared disk devices to the nodes!
(This is because SCSI IDs and hardware configuration of external SCSI adapters are not necessarily set appropriately).

Sample configuration with three nodes and two distinct SCSI buses



Reboot the Concerned Nodes

If you have changed adapter SCSI IDs on a node (or on several nodes), shutdown and reboot the node(s) in order to make the new SCSI ID effective for the adapter.

Where You Go From Here

Continue with the next chapter to check the software installation.

Chapter 5. Checking Software Installation

This chapter explains these steps:

- Check the Software Installation
- Read the Software Release Bulletin.
- Check that Both HACMP and BullCluster are Pre-loaded.

Follow the instructions step by step, then proceed with the next chapter.

Check the Software Installation

Before proceeding with subsequent hardware-related procedures (serial networks and disk devices cabling), check that the appropriate software is installed on the nodes, as explained below.

Preliminary Remarks

To run a cluster configuration, the needed software on the cluster nodes consists of:

- the HACMP for AIX software, and
- the BullCluster software package

These two complementary packages are provided on the same media (CD–ROM).

Read the Software Release Bulletin

Read the *Software Release Bulletin* (SRB) that comes with the software. It includes environment requirements and restrictions as well as late-breaking news.

The SRB also includes the procedure to install the software. However, do not apply it now since the needed software is normally pre-loaded on the systems you just received (see below).

Check that Both HACMP and BullCluster are Pre-loaded

Both HACMP and BullCluster are normally pre-loaded (i.e. pre-installed) on any node you have ordered to build your cluster solution. However, it is preferable to check that software pre-loading has effectively be done at the factory.

Carry out this check **on each node** of your configuration, by entering this command:

```
lslpp –L '*cluster*'
```

In the **lslpp** output, you must see:

- filesets that begin with “cluster”: they constitute the HACMP software

  AND

- filesets that begin with “bullcluster”: they constitute the BullCluster software

Note: The **cluster.clvm** fileset is the HACMP part that deals with concurrent access. This fileset, which is provided only on dedicated media, is only needed for concurrent access environments. It is useless if you are installing a highly-available cluster that does not implement concurrent access; in that case, your media does not include the **cluster.clvm** fileset.

Chapter 6. Completing the Hardware Setup

This chapter explains these steps:

- Shutdown and Power Off the Nodes
- Interconnect Nodes Via Serial Networks (RS232)
- Preparing the SCSI Adapters for Later Cabling, on page 6-2
- Set Up the Shared Disk Devices, on page 6-6.

Follow the instructions step by step, then proceed with the next chapter.

Shutdown and Power Off the Nodes

Before proceeding with further hardware setup (serial networks, then shared disks):

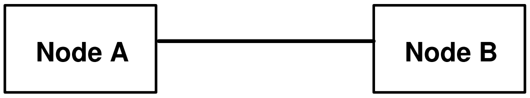
- Shutdown and power off all of the servers you are setting up as cluster nodes.
- Power off any attached device.
- Unplug any power cord (servers and devices). This is a strongly recommended safety precaution.

Interconnect Nodes Via Serial Networks (RS232)

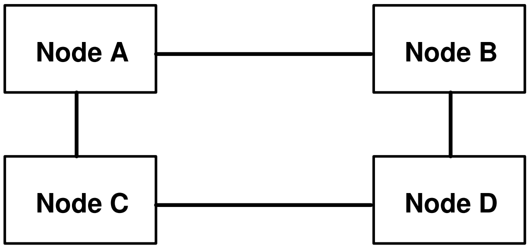
Prior Knowledge

In a cluster, serial networks are RS232 links that connect nodes together. These non-TCP/IP links transport “keep-alive” traffic generated by HACMP. If the TCP/IP subsystem should fail, the serial links will continue to carry node-to-node keep-alive traffic.

Each node in the cluster must have a serial network connection with its neighboring nodes. For clusters that include only two nodes, you have to install a single serial link between the two nodes.



Serial Networks for a 2-node cluster: one RS232 line interconnects the two nodes.



Serial Networks for a 4-node cluster: four RS232 lines interconnect the nodes to form a logical ring.

Notes:

- For details on the role of serial networks, refer to “Planning Networks”, on page 3-4.
- Consider that serial networks are always RS232 (target mode SCSI is not supported).

Physically Install the Needed RS232 Serial Lines

Physically install the RS232 serial lines between the involved nodes.

This operation consists in connecting each end of your RS232 cable to a serial port on each node. Note that if you plan to use the RSF software for remote maintenance via a modem, the S2 port should be reserved for RSF operation (in this case, do not use the S2 port to implement the serial network). For details on serial port concerns, refer to page 3-3.

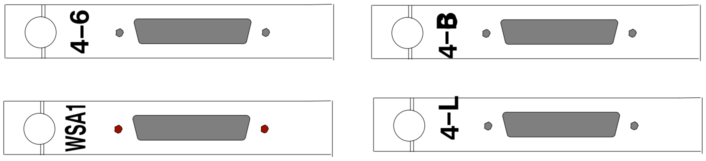
Preparing the SCSI Adapters for Later Cabling

When is this step needed?

- If you are installing a new Powercluster or HA Solution, you do not need to perform this step, since it has been performed at the factory. In this case, directly skip to “Setting Up the Shared Disk Devices”, on page 6-6.
- This step is required only if you are upgrading existing servers to include them in a cluster.

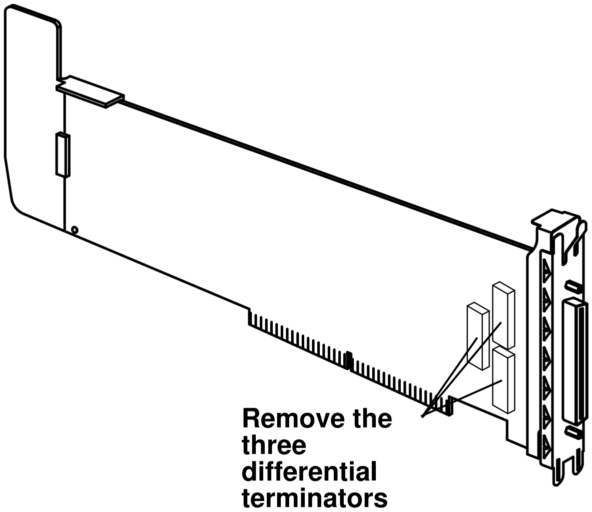
Preparing Type 4-6, 4-B, 4-L or WSA-1 SCSI Differential Adapters

Look for SCSI-2 Differential Fast/Wide adapters in the adapter slots in the back of the servers. These adapters are labeled with the “4-6”, “4-B”, “4-L” or “WSA1” code, as shown in the figure below.



On the adapter card, locate the three terminators to remove, as shown in the figure below. If you do not see terminators, that means they have already been removed at factory, thus you may reinstall the card immediately.

Gently extract the three terminators. Save them for possible future use.

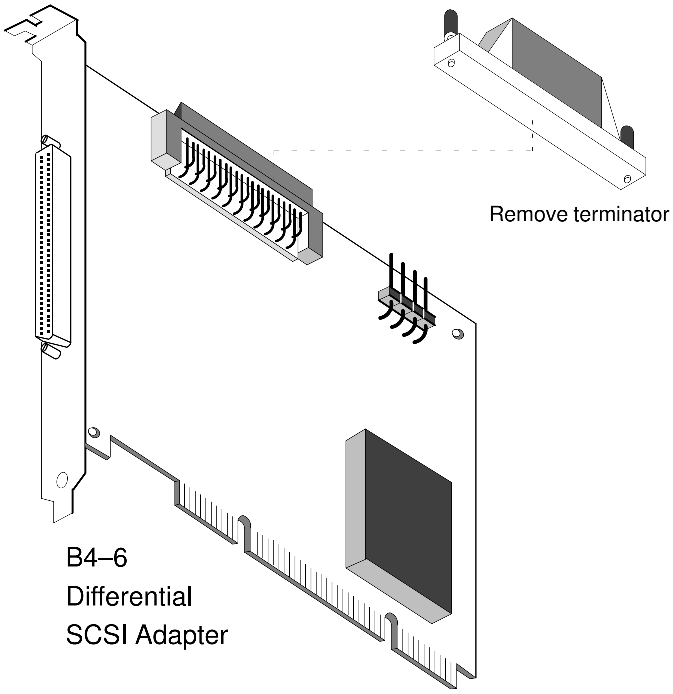

Remove the three differential terminators

Once this operation is completed for each adapter, you may cable the shared disk devices, as explained in the following sections.

Preparing Type B4-6 SCSI Differential Adapters

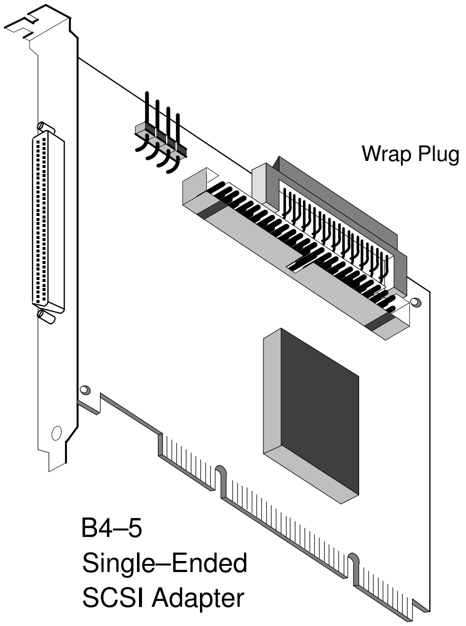
Escala E and T Series and EPC400 servers can be upgraded with additional SCSI DE adapters in order to configure them as cluster nodes.

Note: B4-6 adapters are Ultra SCSI Differential adapters used in Escala E200 and T400 cluster nodes.



Preparing Type B4-5 SCSI Single-ended Adapters

Escala E and T Series servers can be upgraded with additional SE SCSI adapters in order to configure them as cluster nodes. **Note:** The setup of these adapters involves the use of wrap-plugs.



For hints related to these wrap-plugs, refer to “Planning HA Native Shared Storage for Escala HA Solutions”, starting on page 3-7.

Preparing Type B4-7 Fibre Channel Adapters

These adapters are used with DAS 3500 Disk Array Storage Systems.

Connection to the physical layer (FC–0) is accomplished through the industry standard GLM (Gigabaud Link Module) which allows speeds of 1063 Mbps. The GLM offers a DB9 copper connection.

MIA (Module Interface Adapter) are used to provide Fiber Optic dual SC connectors.

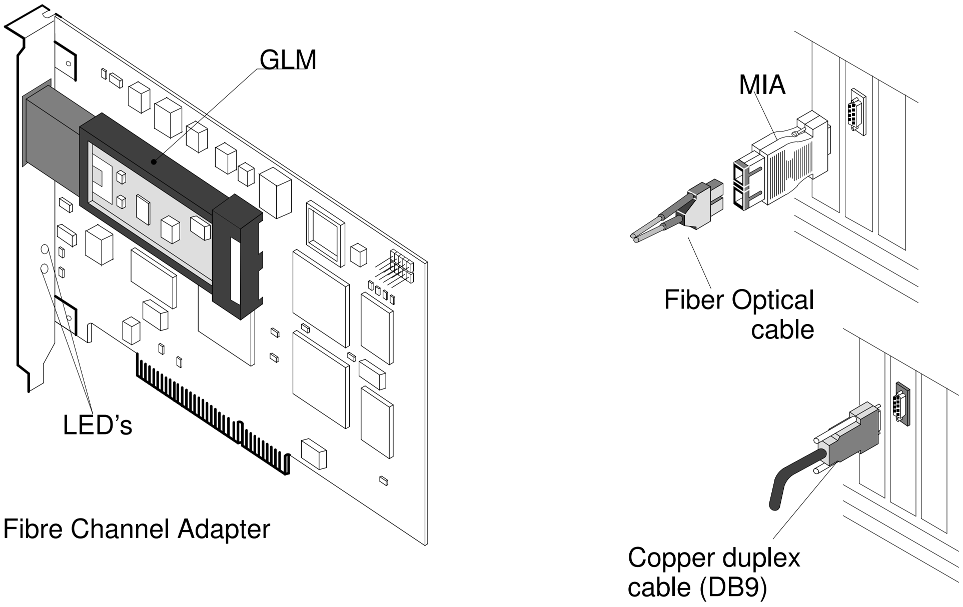


Figure 1. PCI Fibre Channel Adapter (type B4-7)

Preparing Type B4-8 Fibre Channel Adapters

These adapters are used with Fibre DAS Storage Subsystems (DAS5700, DAS5720 and DAS5300).

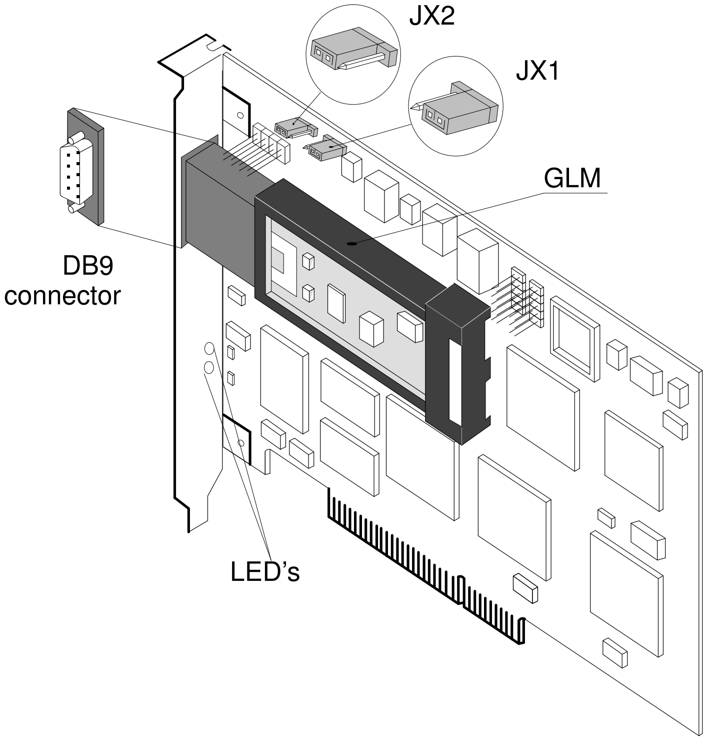


Figure 2. PCI Enhanced Fibre Channel Adapter (type B4-8)

Preparing Type B4-A and B4–B Fibre Channel Adapters

These adapters are used with Fibre DAS Storage Subsystems (DAS5700, DAS5720 and DAS5300).

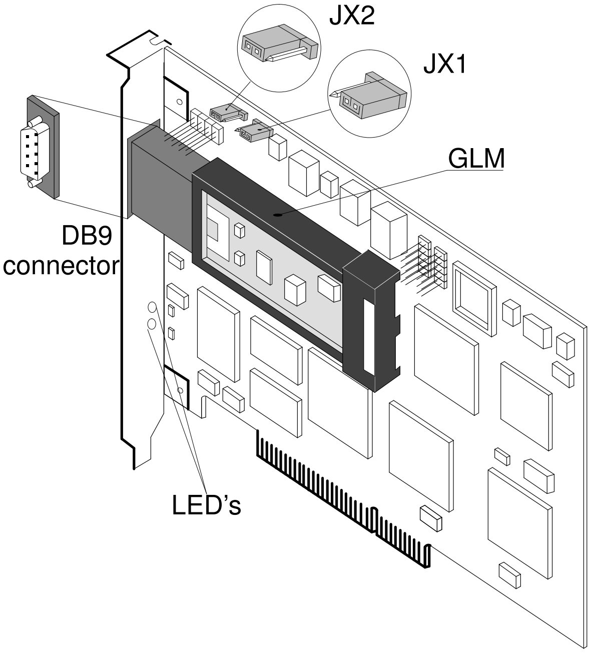


Figure 3. PCI 64–bits Copper Fibre Channel Adapter (type B4-A)

The jumpers JX1 and JX2 must be set as shown in figure 3.

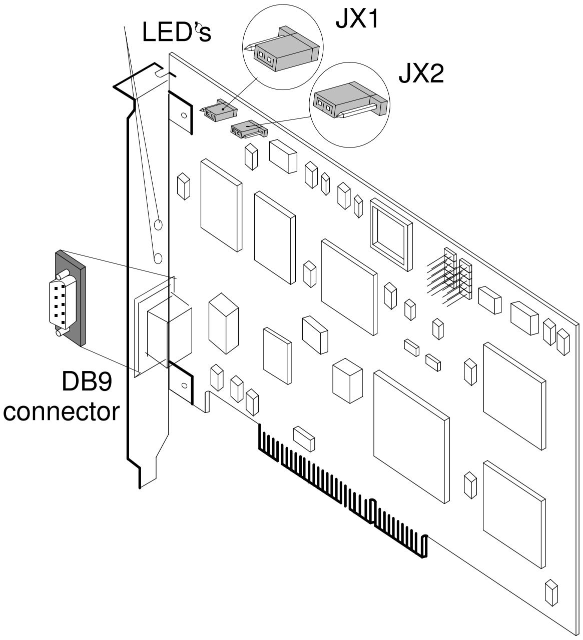


Figure 4. PCI 64–bits Optical Fibre Channel Adapter (type B4-B)

The jumpers JX1 and JX2 must be set as shown in figure 4.

Preparing the Fibre Channel Hub

The Fibre Channel Hub must be connected to the nodes using Fibre Channel Adapters and also to the DAS units.

Note: There is no power switch on the Fibre Channel Hub. The hub must be the first operating component, before powering up the nodes with the FC-AL adapters.

Set Up the Shared Disk Devices

This task consists of cabling and setting up the shared disk devices.

- If you are installing SSA disk subsystems, refer to “Planning Shared SSA Disk subsystems” on page 3-22, and to the documentation of your SSA equipment. Please follow the procedure described in section “Installing Shared IBM 7133 SSA Disk Subsystems” included in the chapter “Installing Shared Disk Devices” of the the *HACMP Installation Guide*. Then, continue with the chapter “Carrying Out Pre-Configuration Tasks”.
- If you are installing SCSI-based or Fibre Channel shared devices, refer to the guidelines below. Other useful and applicable information can be found in the documentation that comes with your disk devices, and in chapter “Installing Shared Disk Devices” of the *HACMP Installation Guide*.

Cabling the Disk Devices and Setting their SCSI ID

Warning: Before you begin, make sure that the systems and any device are powered off. Unplug all the power cords from the wall outlets as well as from the system and device connectors. Observe any safety precautions given in the hardware documentation.

1. Connect all the SCSI cables required for the planned configuration. You may want to look at the chapter “Supplemental Planning Information”, starting on page 3-1, that includes information and figures related to disk cables. For details, also refer to the documentation that comes with your disk devices.
2. Set appropriately the SCSI IDs of your disk devices. Each device on a SCSI bus must have a unique ID, ranging from 0 to 7. Be sure to choose IDs that are not already assigned to the SCSI adapters on the nodes. (For a discussion on SCSI IDs concerns, refer to page 4-7.)
3. For DAS Fibre Channel, assign a loop ID for each SP. See DAS documentation.
4. Complete the hardware installation as directed by the documentation that comes with the disks you are installing (power cords...). If you are installing DAS subsystems, connect the DAS SPs to the node, using RS232 cables.
5. For FC-AL loops (option), power on the hubs.
6. Power on the disk devices (if applicable), then power on and boot the cluster nodes.

   At boot time (or whenever the **cfgmgr** command is used), AIX configures all the devices that are connected to the system. When it configures a disk, it assigns a logical name to the disk of the form `hdisk`*`X`*, where *`X`* is an integer that uniquely identifies the disk.
7. Do not create shared volume groups now. This task is discussed in a subsequent chapter, “Completing the Installation”.
8. Verify the installation:
   a. If you are setting up DAS subsystems, refer to section “Configuring DAS Subsystems” below.
   b. If you are setting up other SCSI disk devices, use the AIX command

      ```
      lsdev -Cc -H
      ```

   to check that the disks are in the “Available” state.

   More details on this procedure can be found in section “Installing Shared IBM SCSI-2 Differential Disks” included in the chapter “Installing Shared Disk Devices” of the *HACMP Installation Guide*. Then, continue with the chapter “Carrying Out Pre-Configuration Tasks”.

Configuring DAS Subsystems

This section summarizes the tasks you must perform to configure your DAS subsystems. We assume here that you have completed the hardware installation (SCSI and RS232 cables, SCSI ID settings...).

Warning: The indications given should be regarded as guidelines only. For detailed instructions, you must refer to the documentation that comes with the DAS subsystem (*Configuring and Managing a DAS*, and *Installing and Maintaining a DAS*).

The configuration tasks involve several issues that are discussed below.

tty Setup

On each node, configure serial ports that connect to the SPs of the DAS subsystems. Use the **smit tty** AIX command, as explained in the chapter “Installing your software” of the *Configuring and Managing a DAS* book.

Physical Disk Units (LUNs) Setup

Now, you must set up physical disk units (LUNs). This is a requisite before you can create volume groups (whether or not you plan to use the Bullcluster Quick Initial Configuration Tool).

Follow the instructions of chapter “Setting physical disk units and storage-system caching” in the *Configuring and Managing a DAS and Using ATF* book. Before that, read the hints below:

Accessing the DAS Manager for the DAS 2900, 3x00 and 3500

To access the DAS manager, use the `smit disk` command, then choose the **DiskArray Storage-System Manager** option.

Accessing the DAS Manager for the Fibre DAS

To access the DAS manager, choose the **Navisphere** option.

Setting Up the Physical Disk Units (RAID Levels, LUNs)

You must create LUNs, i.e. to bind disk modules as LUNs. A LUN (Logical Unit) is one or more disk modules bound into a single entity (note that the terms “LUN” and “physical disk unit”, as used in the DAS documentation, have the same meaning).

To set up the physical disk units (LUNs), go to the **Bind Physical Units** menu. Read carefully the recommendations given in the *Configuring and Managing a DAS and Using ATF* book, keeping in mind the following:

- **Do not create more than 8 LUNs per DAS subsystem**.
- DAS subsystems support RAID levels 0, 0/1, 1, 3, and 5. It is up to the system administrators, to determine which RAID level best suits their needs. Actually, typical cluster configurations use RAID levels 3 or 5, which are the most suitable for high availability purposes.
- As instructed in the *Configuring and Managing a DAS and Using ATF* (“Installing the ATF Software”), if you use ATF, make sure you disable the **auto-assign** option for each LUN.
- The default value for the **Maximum rebuild time** bind option is 4 hours. However, if your site requires fast response time and you want to minimize degradation to normal I/O activity, extend the rebuilding process over a longer period of time, such as 24 hours.
- Once you have set up the desired LUNs, refresh the AIX ODM database as explained below.

Shared Volume Groups

Do not create shared volume groups now:

- If you use the *Bullcluster quick initial configuration* tool to implement an environment dedicated to specific applications, shared volume groups will be created automatically.
- Otherwise, you will have to create shared volume groups manually. This task is discussed in a subsequent chapter, "Completing the Installation".

Refreshing the AIX ODM Database

Once you have created and set up the desired LUNs, we recommend that you perform the following procedure in order to refresh the AIX ODM database:

1. List the disks by entering the command:

   ```
   lsdev -C -c disk
   ```

2. From the displayed list, note the disks that belong to the DAS subsystems. From the AIX viewpoint, any physical disk unit (LUN) you have created is seen as one individual disk (`hdisk`<*n*>).
3. For each DAS disk in the list, run the command:

   ```
   rmdev -d -l hdiskn
   ```

   (In this command, replace *n* with the appropriate disk number.) This removes the device definition from the AIX ODM database.
4. Run the cfgmgr command to instruct AIX to re-configure the devices.

   ```
   cfgmgr
   ```

5. Once this procedure is performed, you are sure that the ODM database is consistent.

It is recommended to perform the steps above each time you change DAS disk bindings, in order to make sure that the ODM database remains consistent.

EMC Storage Additional Setup

An EMC cabinet is capable of supporting a large amount of disks and, in the event of a disk failure, takeover of disks is likely to take a long time. For this reason the choice is made, by default, not to automatically start the takeover process.

smit menus can be used to configure the behavior of a cluster using an EMC in a split bus configuration (takeover or not on an adapter or cable failure).

```
smit hacmp
   Bull Cluster Easy Configurator
      EMC Split-bus Behaviour Configuration
```

Where You Go From Here

The next setup task consists in configuring the cluster topology for HACMP. Proceed with the next chapter to carry out pre-configuration tasks.

Chapter 7. Configuring the Cluster Topology

This chapter explains how to configure an HACMP cluster by defining the HACMP cluster topology. It includes the following sections:

- Determine Which Configuration Method to Use
- Before Using the “BullCluster Quick Initial Configuration” Tool, on page 7-3.
- Preliminary Step: Updating */etc/hosts and /.rhosts* Files, on page 7-3.
- Setting Up a Configuration File, on page 7-4.
- Applying your Configuration File, on page 7-10.
- What Happens During the Quick Initial Configuration Process, on page 7-10.
- Undoing a Quick Initial Configuration, on page 7-12.
- Verifying a Configuration File Manually Edited, on page 7-13.
- Using the BullCluster Snapshot Utility, on page 7-13.
- Where You Go From Here.

Determine Which Configuration Method to Use

The Two Possible Configuration Methods

Two methods are available for defining the HACMP cluster topology: the standard HACMP-based configuration method, and the BullCluster quick configuration method.

Note: For background information about the “topology” concept in HACMP, refer to your HACMP documentation.

The Standard HACMP-Based Configuration Method

This configuration method consists of using various AIX SMIT menus and the standard HACMP utilities and SMIT menus, as explained in the *HACMP Installation Guide*.

This is a general-purpose method, which is suitable for any cluster configuration, whether the configuration is complex or simple.

The Bullcluster Quick Initial Configuration Method

This configuration method consists of using the *Bullcluster Quick Initial Configuration* tool which is included in the BullCluster software package. This tool is implemented as an extension to the SMIT menus of HACMP (accessible through the **smit hacmp** command).

The aim of the *Bullcluster Quick Initial Configuration* tool is to facilitate the configuration of the network interfaces and of the HACMP cluster topology (you need not to go back and forth between the different AIX and HACMP SMIT menus).

However, the *Bullcluster Quick Initial Configuration* tool is not suitable for complex cluster configurations. Consequently, if you intend to use the *Bullcluster Quick Initial Configuration* tool, you must first determine if it is able to handle your cluster configuration: see the criteria in the table below.

Criteria for Determining the Appropriate Method

Refer to the table below and determine whether the *Bullcluster Quick Initial Configuration* tool is appropriate for your cluster configuration.

You can use the Bullcluster quick initial configuration method only if your cluster configuration matches all the conditions below:	**You can NOT use the Bullcluster quick initial configuration method if your cluster configuration matches any condition listed below:**
The cluster includes **2, 3 or 4 nodes**.	The cluster includes **more than 4 nodes.**
The cluster nodes are connected to a **single public network** (**FDDI** or **ethernet**).	The cluster nodes are connected to **several** Ethernet or FDDI **public networks** (as in a dual network configuration), or to one or several **ATM** public networks.

Notes:

- If you plan to implement a private network using Fibre Channel, contact your service representative.
- The notion of “public network” and “private network” is explained on page 3-4.
- Keep in mind that you can always use the standard HACMP-based configuration method, even if the desired configuration matches the requirements of the quick method.

Where You Go From Here

- If the *Bullcluster Quick Initial Configuration* tool is suitable for the cluster configuration you plan, continue with the next section, “Before Using the BullCluster Quick Initial Configuration Tool”.
- Otherwise:
  - Use the standard HACMP-based configuration method, as explained in the *HACMP Installation Guide*, to configure network interfaces and the cluster topology.
  - Once you have configured network interfaces and the cluster topology, refer to chapter “Completing the Installation” (starting on page 8-1) before creating shared volume groups, and defining application servers and resource groups.
  - Note that even once the cluster topology is configured, you can take advantage of menus and facilities that are provided by the BullCluster software package. You can use them to set up cluster resources (resource groups, application servers...) and shared LVM components (volume groups...), and to prepare the cluster for running specific applications.

  Moreover, if your cluster is equipped with shared DAS RAID subsystems or SSA disk subsystems, using these menus and facilities is particularly recommended. They allow you to deal with shared volume groups in a simplified way, that is less error-prone than the manual method.

Before Using the “BullCluster Quick Initial Configuration” Tool

Prior Knowledge

Before you begin, note the following:

Network Adapter Interfaces Naming Conventions

The quick initial configuration tool assumes specific naming conventions for the network adapter interfaces. The `_boot`, and `_stby` strings are always used as the standard suffixes for the boot, and standby adapter labels.

This point was discussed in “Naming Conventions”, on page 4-5.

Other Requisites

Before running the quick initial configuration tool, you must update /**etc**/**hosts** and /**.rhosts** files. This is explained below.

Terminology: Local Node / Remote Node

You will run the quick initial configuration tool from one of the cluster nodes. By convention, in the instructions below:

- the **local node** term refers to the node on which you will run the tool to apply a configuration file,
- the **remote node** term refers to the other node(s)

Reviewing the Quick Initial Configuration Procedure

Important: Before you actually perform the initial configuration, we recommend that you review the steps involved that are described in this chapter. Pay attention to the following sections:

- “Setting Up a Configuration File”, on page 7-4
- “What Happens During the Quick Initial Configuration Process”, on page 7-10

Once you have reviewed these sections, proceed with the instructions below to carry out the configuration.

Preliminary Step: Updating /etc/hosts and /.rhosts Files

When applying configuration settings, the quick initial configuration tool must be able to access the remote nodes in order to run appropriate configuration commands on this node.

Consequently, for the configuration to work properly, you have first to update the /**etc**/**hosts** and /**.rhosts** files on both local and remote nodes, as explained below.

Edit /etc/hosts on the Local Node

Edit the /**etc**/**hosts** file of the local node.

- Add entries for the boot interfaces of the remote nodes. If you plan to not implement IP address takeover, the notion of boot interface makes no sense, thus you add entries for service interfaces of the remote nodes. Note that issues related to IP address takeover are discussed in “Configure TCP/IP For Boot/Service Interfaces, on page 4-4.
- Also, make sure that the /**etc**/**hosts** file has the following entry:

```
127.0.0.1    loopback localhost
```

For example, assuming you will run the quick initial configuration tool on node `foo`, and the unique remote node is `bar`, thus you must edit /**etc**/**hosts** on the `foo` node. Once the appropriate entry is added, and assuming you want to implement IP address takeover, the file resembles the following:

```
127.0.0.1    loopback localhost
# existing entries:

# added entry for bar's boot interface
192.9.200.11    bar_boot
```

Edit /.rhosts on Each Cluster Node

The quick initial configuration tool uses remote shell commands. To ensure that it will work as expected, edit the /**.rhosts** file *on each node in the cluster*. Add entries for the boot interface (or for the service interface, if you do not implement IP address takeover) for all the nodes, with access right granted to **root**.

For example, if you are installing a cluster including the nodes foo and bar, and assuming you want to implement IP address takeover, then the /**.rhosts** file on both nodes must include the two following entries:

```
foo_boot  root
bar_boot  root
```

With a cluster identical to the previous example, but without IP address takeover implementation, the entries would be:

```
foo  root
bar  root
```

For security reasons, you can add these entries only as necessary (i.e. before configuring the cluster through the quick configuration tool or the HACMP menus), and delete them when they are no longer needed (i.e. once the configuration is complete).

Setting Up a Configuration File

Accessing the Bullcluster Quick Initial Configuration Tool

The quick configuration tool is implemented as an extension to the SMIT menus of HACMP.

Note for PowerConsole Users

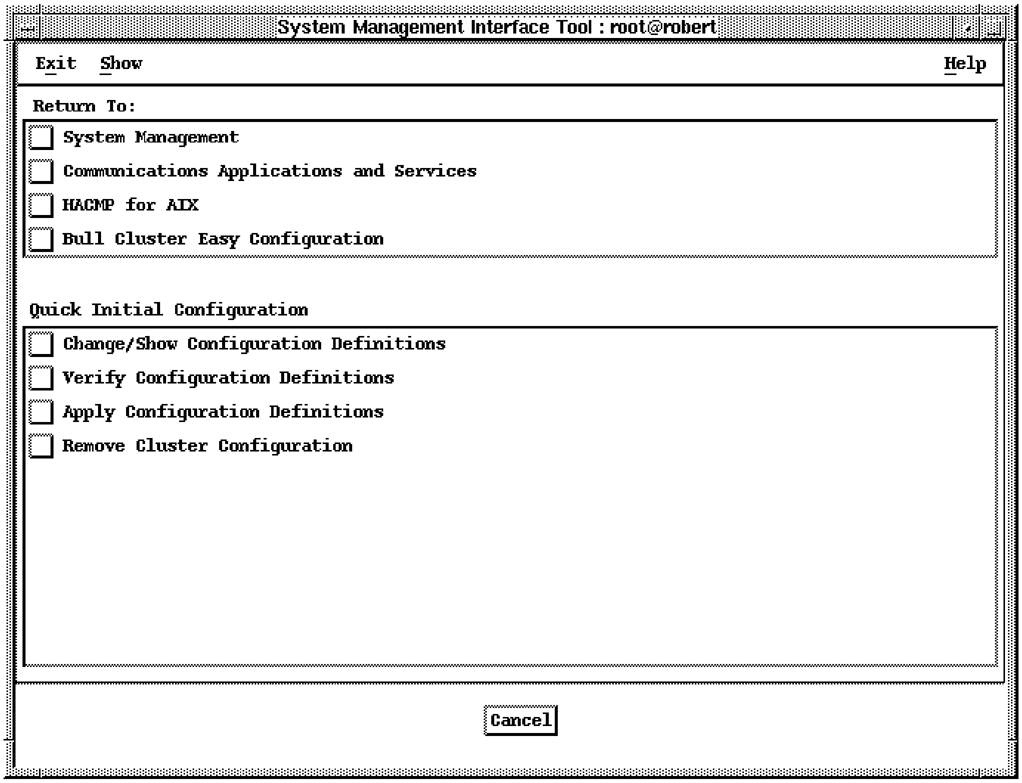
If you are equipped with a PowerConsole, you can take advantage of the ClusterAssistant application, which includes an icon to run the Bullcluster Quick Initial Configuration Tool.

General Procedure

Otherwise (if you are not equipped with a PowerConsole), proceed as follows:

1. Log in to a cluster node as root and enter `smit hacmp` to bring up the HACMP main menu. This is the standard HACMP menu, except it includes the **Bull Cluster Easy Configuration** option.

2. Choose the **Bull Cluster Easy Configuration** option, then **Quick Initial Configuration**. The main menu of the quick initial configuration tool is displayed, as shown below.


System Management Interface Tool : root@robert

Exit Show Help

Return To:

□ System Management
□ Communications Applications and Services
□ HACMP for AIX
□ Bull Cluster Easy Configuration

Quick Initial Configuration

□ Change/Show Configuration Definitions
□ Verify Configuration Definitions
□ Apply Configuration Definitions
□ Remove Cluster Configuration

Cancel

Note: You can also use the SMIT **quickconf** fast path to directly access the quick configuration tool.

Understanding the Quick Initial Configuration Procedure

The quick initial configuration procedure entails two main tasks:

- The first task consists of setting up a configuration file using the functions of the **Change/Show Cluster Definitions** menu. In this menu, you specify information related to the basic cluster configuration, i.e. essentially the network configuration (IP interfaces, addresses) and the cluster topology (number of nodes, cluster and node names).
- The second task consists of applying the configuration file, using the **Apply Configuration Definitions** option. Once the configuration file has been applied, network adapters are setup appropriately, and the HACMP cluster topology is defined.

These tasks are explained below.

Setting Up a Configuration File

Setting up a configuration file consists of specifying the fields of the **Change/Show Cluster Definition** menu. Proceed as explained below:

Bring up the **Change/Show Cluster Definition** menu:

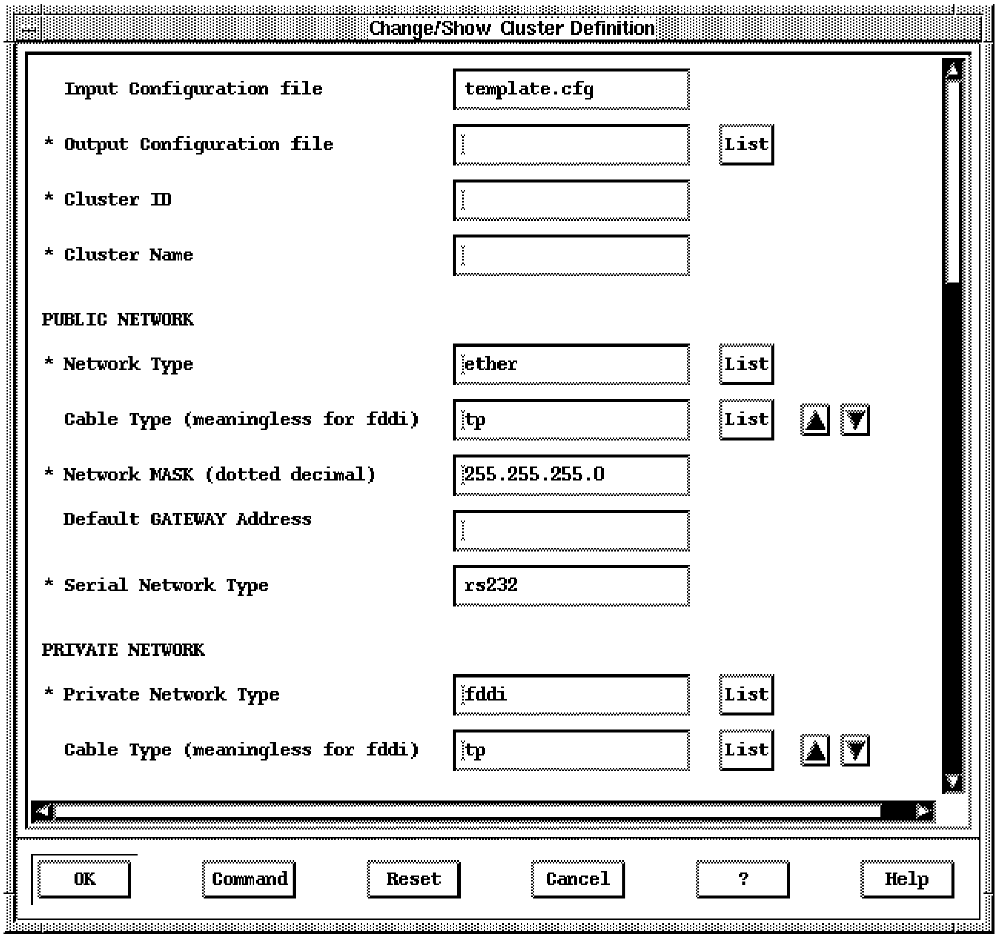
1. From the **Quick Initial Configuration** menu, select the **Change/Show Configuration Definitions** option.
2. You are prompted to enter the configuration file you want to set up. Since you have not yet created a configuration file, choose the default file by entering `template.cfg` (you can also retrieve its name by pressing F4 or clicking on the List button).

3. A new screen appears, that prompts you for node names. In the **Node Names** field, enter the names of the nodes in the cluster.
   - The names cannot exceed 31 characters. They can include alpha and numeric characters and underscores.
   - Separate the names with a space character.
   - If you are about to display or change a configuration file that you have previously set up and saved (i.e. a configuration file other than the default **template.cfg**), press F4 to retrieve the node names defined in this configuration file.

4. Once you have entered the node names, the **Change/Show Cluster Definition** menu is displayed. This is a large menu (such as the menu illustrated on page 7-6), made of several sections:

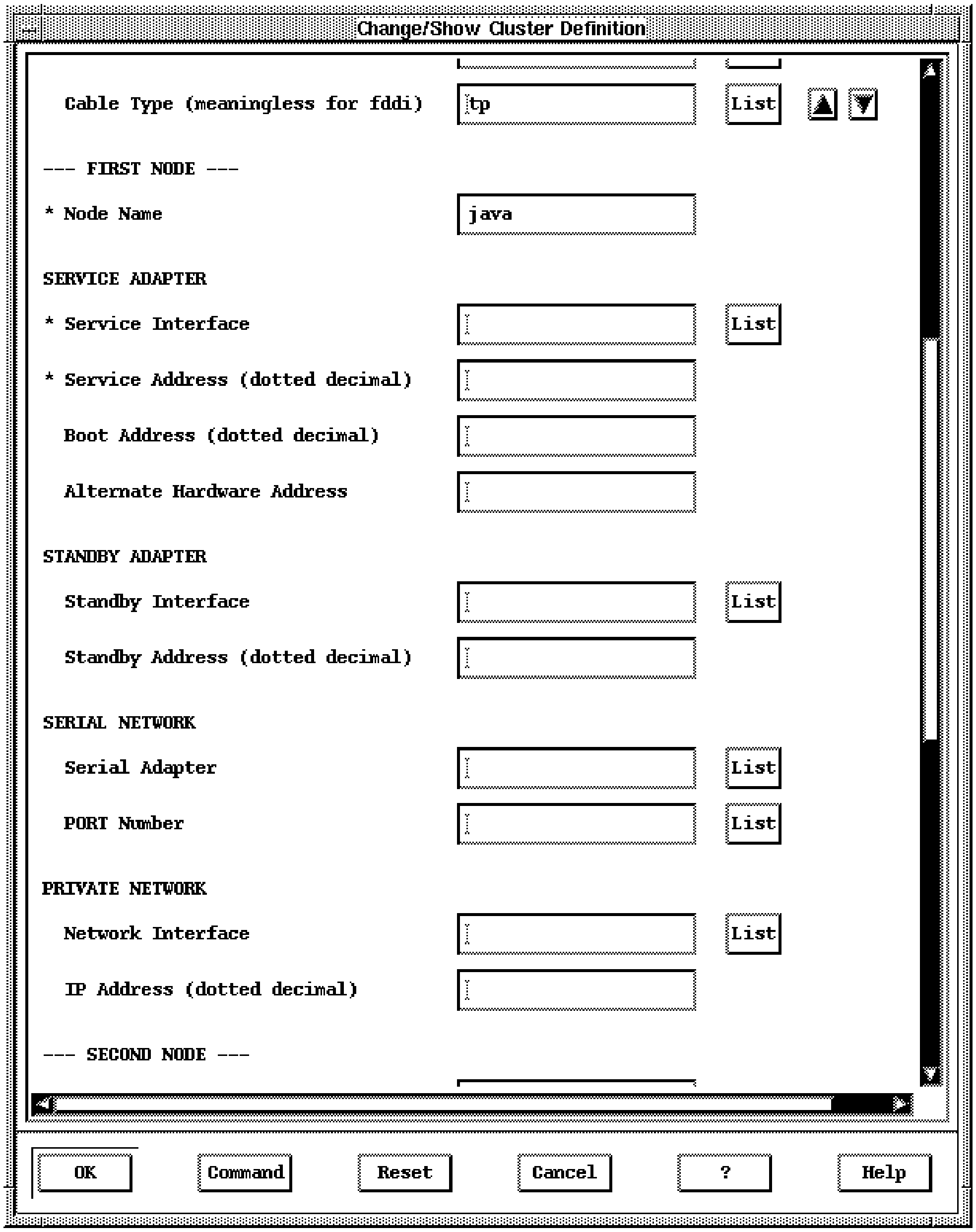
   - At the top, there is a section whose fields apply to the whole cluster.

   - Below this section are displayed as many sections as there are nodes, where the field values are specific to each node.



Input Configuration file	template.cfg
* Output Configuration file	
* Cluster ID	
* Cluster Name	
PUBLIC NETWORK	
* Network Type	ether
Cable Type (meaningless for fddi)	tp
* Network MASK (dotted decimal)	255.255.255.0
Default GATEWAY Address	
* Serial Network Type	rs232
PRIVATE NETWORK	
* Private Network Type	fddi
Cable Type (meaningless for fddi)	tp

Menu Part 1/2: The top part of the "Change/Show Cluster Definition" menu concerns global settings for the whole cluster.



Cable Type (meaningless for fddi) tp

--- FIRST NODE ---

* Node Name java

SERVICE ADAPTER

* Service Interface

* Service Address (dotted decimal)

Boot Address (dotted decimal)

Alternate Hardware Address

STANDBY ADAPTER

Standby Interface

Standby Address (dotted decimal)

SERIAL NETWORK

Serial Adapter

PORT Number

PRIVATE NETWORK

Network Interface

IP Address (dotted decimal)

--- SECOND NODE ---

Menu Part 2/2: The remaining fields of the "Change/Show Cluster Definition" menu concern node-specific settings.

Specify the name of the configuration file where the parameters are to be saved:

1. Note that the **Input Configuration file** information indicates, as a reminder, the current configuration file. As explained above, the default is **template.cfg**.
2. In the **Output Configuration file** field, enter a name for the configuration file you are about to create. The menu parameter values will be saved to this file. Choose a name as meaningful as possible, and append to it the "`.cfg`" suffix, as in the example below:

   ```
   Output Configuration file                foobar.cfg
   ```

   It is not recommended to specify the `template.cfg` name since this would overwrite the default empty template.

   Note: If you are modifying a configuration you have previously created, specify in the **Output Configuration file** field the name of the file you want to modify. In that case, input and output files are the same.

Specify the global settings (cluster ID and name, network characteristics):

1. In the **Cluster ID** field, enter a positive integer unique to your site.
2. In the **Cluster Name** field, enter a text string that identifies the cluster, for example `foobar`. The cluster name can include alpha and numeric characters and underscores. Use no more than 31 characters.
3. Set the parameters related to the **PUBLIC NETWORK**:

   Network Type
   Select either `ether` (if this is an ethernet network) or `fddi`, as appropriate.

   Cable Type (meaningless for fddi)
   If the network type is FDDI, leave this field blank (it makes no sense). If it is ethernet, select `tp`, `dix` or `bnc`, depending on the type of the ethernet. These values correspond respectively to twisted pair ethernet, thick ethernet and thin ethernet.

   Network MASK (dotted decimal)
   Specify the appropriate network mask in the dotted decimal notation. For example, `255.255.255.0` is a typical value. However, this value is site dependent. For details on network masks, refer to the section "Defining a Network Mask", of chapter 3 "Planning TCP/IP Networks", of the *HACMP Planning Guide*.

   Default GATEWAY Address
   This field is optional**.** It makes sense only if your network is equipped with a gateway. Enter the gateway address in dotted decimal form (for example, `192.9.201.1`). For more information about gateways, refer to your AIX documentation.

4. The **Serial Network Type** information is not an editable field. The preset value `rs232` is shown as a reminder. Indeed, the quick initial configuration tool applies only to cluster configurations that use RS232 for their serial network.
5. Set the parameters related to the **PRIVATE NETWORK**. Fill in the fields **Network Type** and, if you are using ethernet for the private network, **Cable Type**.

These are the last fields applicable to the whole cluster. The remaining fields are node-specific.

Specify the node-specific settings:

The remaining fields relate to settings specific to each node. Note that there are as many sections as there are nodes. All sections are identical. At the beginning of each section, a header such as "`--- FIRST NODE --`–" is displayed to show where the section begins, and a **Node Name** field (non editable) shows the name of the concerned node.

The order in which the nodes are listed in the menu correspond to the order in which you have specified the node names when initially bringing up the **Change/Show Cluster Definition** menu.

Specify the field values for each node, the steps that follow must be repeated for each node section of the menu.

1. Specify information related to the **SERVICE ADAPTER** of the node, by entering field values as follows:

   Service Interface
   Enter a network interface name, such as `en0`, for the service adapter of the node.

   Service Address (dotted decimal)
   Enter the IP address for the service adapter. Use the dotted decimal notation (for example, `192.9.201.100`).

Boot Address (dotted decimal)

Enter the IP boot address. Use the dotted decimal notation (for example, `192.9.201.120`). If you do not want to implement IP address takeover, the notion of boot address makes no sense, so leave this field blank. If you specify a boot address, it must be on the same subnet as the service address.

Alternate Hardware Address

This field is optional**.** It makes sense only if you want to implement hardware address swapping. If so, enter an alternate hardware address for the adapter (this must be a 12-digit value, for example `02608c2fb512`). The address must be unique within the physical network, so that it does not conflict with the hardware address of another adapter in the network. For hints on selecting an alternate hardware address, refer to chapter “Planning TCP/IP Networks” of the *HACMP Planning Guide*.

2. Specify information related to the **STANDBY ADAPTER** of the node, by entering field values as explained below.

   Standby Interface

   Enter a network interface name, such as `en1`, for the standby adapter of the node.

   Standby Address (dotted decimal)

   Enter the IP address for the standby adapter. Use the dotted decimal notation (for example, `192.9.202.100`). The standby adapter must be on a different subnet from the service adapter. For details on adapter IP addresses, refer to the *HACMP Planning Guide*.

3. Specify information related to the **SERIAL NETWORK**, by entering field values as explained below. Note that these fields are not available if the cluster you are configuring includes more than two nodes. In that case, you must manually configure the serial network after having performed the quick initial configuration. Also note that the unique serial network supported is based on RS232 serial links.

   Serial Adapter

   Specify the serial adapter (for example `sa1`) that is used for the serial network attachment. Press F4 to display the list of available serial adapters.

   PORT Number

   Specify the corresponding port number (for example `s2`). Press F4 to display the list of available serial ports.

4. Specify information related to the **PRIVATE NETWORK**, by entering field values as explained below. If you do not implement a private network in your cluster, leave these two fields blank.

   Network Interface

   Enter a network interface name, such as `en1`, for the private network adapter of the node.

   IP Address (dotted decimal)

   Enter the corresponding IP address. Use the dotted decimal notation (for example, `192.9.200.3`). Since this is a private network, it must be on a different subnet from any other subnet used for the public network.

5. Do not forget that these steps must be repeated for each node section in the menu.

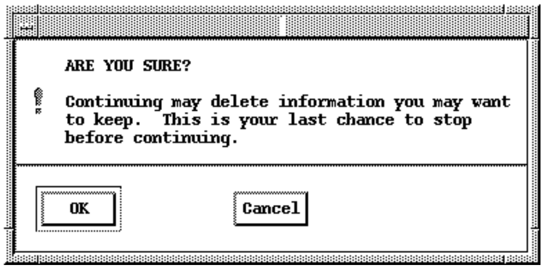
6. Once all the fields are specified appropriately, validate the screen. A message is displayed, and the parameters are saved to the **Output Configuration file** you have specified at the top of the menu.

7. Press F3 several times to return to the **Quick Initial Configuration** menu.

Applying your Configuration File

Once you have set up your configuration file as explained above, proceed as follows to apply it:

1. From the local node, go to the **Quick Initial Configuration** menu, and choose the **Apply Configuration Definitions** option. You are prompted to specify the name of the **Configuration file to apply**.
2. Press F4 to display the list of defined configuration files. From this list, select the configuration file you want to apply. Once the **Configuration to apply** is specified, press Enter.
3. You are prompted to specify whether or not you want detailed messages to be displayed during the quick initial configuration process. Select `Yes` if you want verbose messages or `No` if you prefer less verbose messages.
4. When you validate the screen, a confirmation message is displayed, as shown below.



ARE YOU SURE?

Continuing may delete information you may want to keep. This is your last chance to stop before continuing.

5. Press Enter to run the quick initial configuration process.
6. Wait for the completion of the configuration process, then read the displayed messages to check that the configuration was successful.

Your cluster is now configured. Before proceeding with the next chapter, which describes post-configuration tasks, you may want to review the section below to know what actually happens when applying a configuration file.

Note: If for some reason, you want to undo the quick initial configuration process, refer to "Undoing a Quick Initial Configuration", on page 7-12.

What Happens During the Quick Initial Configuration Process

This section summarizes what the quick initial configuration tool does when you process a configuration file with the **Apply Configuration Definitions** function.

The actions carried out by the quick initial configuration tool can be divided into two categories: configuration actions and verification actions.

Configuration Actions

The quick initial configuration tool sets up the network adapters and configure the HACMP cluster topology.

Network Adapters, IP Address Takeover, hostnames, /etc/hosts and /.rhosts

The quick initial configuration tool:

- defines the IP label, IP address and network mask for the service and standby adapters of both nodes;
- implements IP address takeover if you have specified boot addresses in the configuration file, and configures each service adapter for which IP address takeover might occur to boot from the boot adapter address and not from its service adapter address.

The IP addresses and network masks are set according to the values you have saved to the configuration file that is applied by the quick configuration tool.

Concerning the hostname:

- If you have chosen to **not** enable IP Address Takeover (i.e. if you have not specified boot addresses), the quick initial configuration process does not change the hostname (for example, the initial name `foo` remains `foo`).
- If you have chosen to enable IP Address Takeover, it changes the hostname. For example, the hostname `foo_boot`, is changed in `foo`. This scheme is usually desirable (you rarely want a hostname that matches the boot interface).

Naming Conventions

According to the naming rules used by the quick configuration tool, the `_boot`, and `_stby` strings are always used as the standard suffixes for the names of the boot and standby interfaces, while no suffix is used for the service interface.

For example, assume you are implementing a cluster with two nodes having the names `foo` and `bar`:

- The `foo` node is configured as having the following three interfaces:

`foo_boot`	*(boot)*
`foo`	*(service*)
`foo_stby`	*(standby)*

- Similarly, the `bar` node is configured as having the following three interfaces:

`bar_boot`	*(boot)*
`bar`	*(service)*
`bar_stby`	*(standby)*

This assumes that each server has been properly configured, with a hostname suffixed with the `_boot` string. For details, on these conventions, refer to "Configure TCP/IP for Boot/Service Interfaces", starting on page 4-4.

/etc/hosts and /.rhosts

The quick initial configuration process updates on both nodes the **/etc/hosts** and **/.rhosts** file, by adding entries for the service and standby addresses.

Hardware Address Swapping

If you have specified values in the **Hardware Address** optional fields of the **Change/Show Cluster Definition** screen, the quick initial configuration process implements the hardware swapping facility. The specified hardware addresses (one for each node) are used as alternate hardware addresses.

For details on hardware address swapping, refer to chapter "Planning TCP/IP Networks" of the *HACMP Planning Guide*.

RS232 Serial Network

On each node, the quick initial configuration tool defines and configures the tty device that is connected to the RS232 line used to implement the serial network. The quick configuration tool relies on the **Serial Adapter** and **PORT Number** fields of the **Change/Show Cluster Definition** screen.

Note that if the tty device is already configured, the quick initial configuration process stops. In this case, you have to delete this tty device, using the **rmdev** command or the **Remove tty** option of the **tty** SMITmenu, before you again run the quick configuration tool.

For details on RS232 serial line concerns, refer to the AIX documentation and to section "Configuring Networks" of the *HACMP Installation Guide*.

HACMP Cluster Topology Configuration

The quick initial configuration process configures the cluster topology for HACMP and creates accordingly all the needed HACMP objects: cluster ID and cluster name, node names, adapters (label, type, network attribute...), and so on.

Once the configuration is complete, the quick initial configuration process synchronizes the HACMP cluster topology definition on all nodes.

Verification Actions

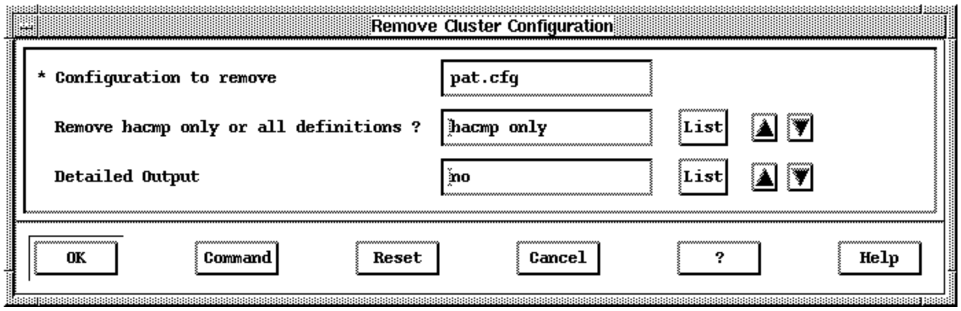
Before running the configuration phase, the quick configuration tool performs preliminary checks related to the concerned adapters (ethernet, SCSI and RS232).

Once the configuration phase is complete, the quick configuration tool runs a verification phase to verify that the cluster environment is actually properly configured.

Undoing a Quick Initial Configuration

If after applying the quick initial configuration process, you discover that it must be undone, proceed as follows:

1. From the local node, go to the **Quick Initial Configuration** menu, and choose the **Remove Cluster Configuration** option.
2. In the displayed screen, specify the name of the configuration file you have applied (be sure to specify the configuration file you have just applied!).
3. The following screen is then displayed:



```
* Configuration to remove                        pat.cfg
  Remove hacmp only or all definitions ?         hacmp only
  Detailed Output                                no
```

4. Enter field values:

Remove hacmp only or all definitions ?

• Enter **hacmp only** if you only want to remove HACMP definitions (i.e. the HACMP-related objects of the ODM database).
• Enter **all** if you want to remove HACMP definitions as well as the volume groups that have been created by the quick initial configuration. Removing volume groups is a destructive action, so be cautious about using the **all** option.

Detailed Output

Select `yes` if you want verbose messages or `no` if you prefer less verbose messages.

5. Press Enter to apply the command with your settings.

Verifying a Configuration File Edited Manually

The **Quick Initial Configuration** menu includes the **Verify Configuration Definitions** option, which can be used to verify that a given configuration file is correct.

This function is useless if you have used the quick initial configuration SMIT menus to edit the configuration file (the normal method). The function is intended for the field personnel, when they have created or modified a configuration file with a text editor.

Using the BullCluster Snapshot Utility

The BullCluster Snapshot Utility allows the user to save, and restore the cluster topology and the BullCluster topology ODM classes.

SMIT menus allowing access to the new Snapshot menus:

```
–>Communications Applications and Services
  –>HACMP for AIX
    –>Bull CLuster Easy Configuration
      –>Bull Cluster Snapshots
          Add a Bull Cluster Snapshot
          Change/Show a Bull Cluster Snapshot
          Remove a Bull Cluster Snapshot
          Apply a Bull Cluster Snapshot
```

Where You Go From Here

If your cluster is equipped with a PowerConsole

- Once the HACMP cluster topology has been defined, you may want to register now the cluster with ClusterAssistant (use the “Register Cluster” icon located in the “Configuration” folder of the ClusterAssistant application group). Once registered, the cluster can be managed from the PowerConsole using the various ClusterAssistant utilities.

Next Step

- Continue with the next chapter to complete the cluster setup. If your cluster is equipped with a PowerConsole, also refer to the *PowerConsole & ClusterAssistant Setup Guide*.

Chapter 8. Completing the Cluster Setup

This chapter explains the tasks to perform in order to complete the installation:

- Note for PowerConsole Users.
- Setting Up HACMP Application Servers and Resources.
- Setting Up Shared LVM Components (Volume Groups...), on page 8-2.
- Synchronizing Time in Your Cluster, on page 8-3.
- Implementing IP Address Aliasing, on page 8-3.
- Enabling Cluster Event Notification by E-Mail, on page 8-3.
- Dealing With Shared Volume Groups on DAS Subsystems, on page 8-5.
- Where You Go From Here.

Note for PowerConsole Users

If your cluster is equipped with a PowerConsole (an optional console solution for PowerCluster, note that you can take advantage of the ClusterAssistant software to perform the setup tasks discussed in this section. Indeed, ClusterAssistant provides handy icons to quickly bring up the different setup menus involved in these tasks.

Setting Up HACMP Application Servers and Resources

Once you have configured the HACMP cluster topology, you must set up the desired HACMP application servers and resource groups using BullCluster and/or HACMP facilities. Whatever method you choose, refer to the HACMP documentation for background information on concepts related to application servers and resource groups.

Using BullCluster Facilities

- The BullCluster software package provides the **Cluster Resources Modification** menu. You can take advantage of this menu, which offers easy-to-use functions to set up resource groups and specific application servers. For example, you can create a resource group using the **Add a Resource Group** option, then create and add the desired resources to this resource group. For further details, refer to the chapter “Managing Cluster Resources and Shared LVM Components”, starting on page 9-1.
- The BullCluster software package also includes sample start and stop scripts useful for implementing various applications in your cluster as highly available applications. For a quick presentation of these sample scripts, refer to “planning Specific Applications”, on page 3-1. For detailed information, refer to the *Powercluster & HA Solutions: Using the Sample HA Scripts* document, which is mainly intended for Bull technical personnel.

Using HACMP Facilties

- Instead of using the BullCluster **Cluster Resources Modification** menu to set up resource groups and specific application servers, you can also use the standard HACMP menus and facilities. The *HACMP Installation Guide* provides detailed instructions on setting up application servers and resource groups using the standard HACMP menus.

Setting Up Shared LVM Components (Volume Groups...)

If, using the quick initial configuration tool, you have applied a configuration file customized for a specific application (X25, Informix, Oracle, OPS, Tuxedo or HVX), shared volume groups and filesystems are already defined and ready for use.

If not, you must create shared volume groups and filesystems as desired, using either the quick method or the manual method (see below).

Quick Method

The BullCluster software package provides the **Cluster Resources Modification** menu. This menu offers easy-to-use functions to create and set up shared LVM components (volume groups, logical volumes, filesystems). The quick method consists in using the following options of the **Cluster Resources Modification** menu to create the desired shared LVM components and declare them as resources of a given resource group:

- **Add a Volume Group to a Resource Group**
- **Add a Logical Volume to a Resource Group**
- **Add a File System to a Resource Group**

For further details, refer to chapter “Managing Cluster Resources and Shared LVM Components”, starting on page 9-1.

Note: If you are using DAS subsystems, before creating volume groups, you must have defined the desired physical disk units (LUNs) and refreshed the AIX ODM database, as explained in “Physical Disk Units (LUNs) Setup” on page 6-7.

Manual Method

The manual procedure to set up volume groups differs depending on the type of shared disk devices that are installed:

- If you are using non-RAID SCSI disk devices or SSA disk subsystems:

  No special consideration applies: simply refer to the AIX and HACMP documentation. In particular, you can find useful information in chapter “Defining Shared LVM Components” of the *HACMP Installation Guide*.

- If you are using DAS subsystems:

  Special considerations apply, that are explained in the section “Dealing With Shared Volume Groups on DAS Subsystems”, on page 8-5.

For further details on manually setting up shared volume groups and filesystems, refer to the AIX and HACMP documentation.

Synchronizing Time in your Cluster

It is recommended that you implement the **timed** daemon in your cluster so that clocks are maintained synchronized on all the nodes. For details, refer to “Synchronizing Time in your cluster with timed”, on page 10-25.

Implementing IP Address Aliasing

If IP address takeover is enabled in your configuration, you may want to implement the IP address aliasing feature (optional). This feature provides the means to alias the service and boot IP addresses used by a service adapter. For details, refer to page 10-5.

Enabling Cluster Event Notification by E-Mail

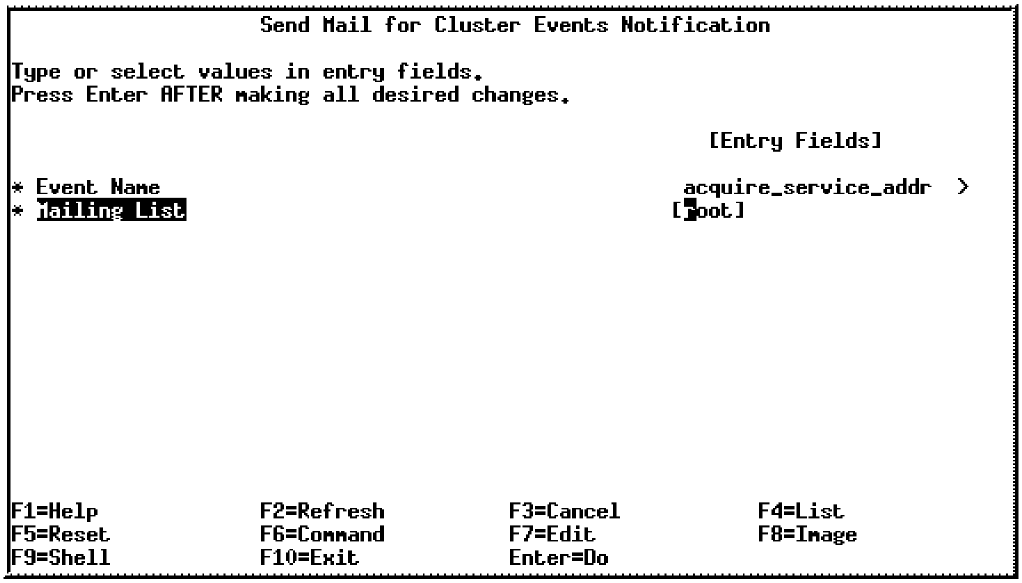
You may want to customize cluster event processing so that when a cluster event occurs, an e-mail message is send to the administrator.

To implement this feature, you may use the standard **Change/Show Cluster Events** HACMP menu and specify the appropriate **Notify Command**. However, this menu applies only to one event at a time. To configure many events at once for notification by e-mail, use the **Send Mail for Cluster Events Notification** menu. This BullCluster-specific menu is an extension to the standard HACMP menus.

Procedure

Use the **Send Mail for Cluster Events Notification** menu as follows:

1. From the main SMIT menu of HACMP, choose **Bull Cluster Easy Configuration**, then **Send Mail for Cluster Events Notification**.
2. The event list is then displayed.
3. From the list, select one or several events using the F7 key. Press Enter. A menu similar to the following is displayed:



```
                     Send Mail for Cluster Events Notification

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                                        [Entry Fields]

* Event Name                                          acquire_service_addr  >
* Mailing List                                       [root]











F1=Help              F2=Refresh           F3=Cancel            F4=List
F5=Reset             F6=Command           F7=Edit              F8=Image
F9=Shell             F10=Exit             Enter=Do
```

4. Note that **Event Name**, which is not an editable field, contains the event names you have just selected.

5. In the **Mailing List** field, specify the e-mail addresses to which notification messages must be sent. You can specify several e-mail addresses separated with a blank, as in the example below:

```
root cladmin@foosite.com jack@barsite.com
```

6. Press Enter. When the confirmation message appears, press Enter again.

What Happens When Using the “Send Mail for Cluster Events Notification” Menu

The procedure above modifies the notify command associated with each event you have selected.

- Thus, if you go to the **Change/Show Cluster Events** HACMP menu (accessible through the **Manage Node Environment** menu), you can see that for these events, the **Notify Command** field has been modified. This field contains the following command:

```
/usr/sbin/bullcluster/admin/mailtoroot
```

The **Notify Command** comes with the BullCluster software package and is run both before and after the event command execution. Thus mailtoroot sends an e-mail message both before and after the event command execution.

Note: For details on cluster events processing, refer to chapter “Maintaining Cluster Events Processing” in the *HACMP Administration Guide*.

The procedure described above also modifies the **/usr/sbin/bullcluster/admin/mailtoroot.cfg** file, by adding to it any mail address you have specified in the **Mailing List** field of the **Send Mail for Cluster Events Notification**. The `mailtoroot` script relies on mailtoroot.cfg, i.e. it sends e-mail messages to all the addresses specified in this file.

- When you use the **Send Mail for Cluster Events Notification** facility, if you change the **Mailing List** field, the `mailtoroot.cfg` list is updated accordingly. Consequently, this affects all the events that are currently configured to use `mailtoroot` as the notify command.

Note: The `mailtoroot` script and the `mailtoroot.cfg` files are part of the BullCluster software package, and are installed in the **/usr/sbin/bullcluster/admin/utils** directory. You may want to take a look at them.

Dealing With Shared Volume Groups on DAS Subsystems

This section explains the tasks you must perform when manually creating and setting up shared volume groups on DAS subsystems.

Warning: The manual method described below is rather complex. Using the quick method, presented on page 8-2, is recommended. The quick method allows you to deal with shared volume groups in a simplified way, that is less error-prone than the manual method.

Preliminary Remarks

- We assume here you have defined the desired physical disk units (LUNs) and you have refreshed the AIX ODM database, as explained in "Physical Disk Units (LUNs) Setup" on page 6-7.
- The tasks differ depending on whether you use the split-bus or the dual-initiator/dual-adapter DAS configuration.
- The split-bus and dual-initiator/dual-adapter configurations are discussed in "Planning Shared DAS Subsystems", on page 3-11.

Case of a Dual-Initiator/Dual-Adapter Configuration

To manually create and set up a shared volume group:

- Simply follow the standard manual procedure described in the section "Maintaining Shared Volume Groups", of the *HACMP Administration Guide*. This setup procedure is outlined below.
- Note that for a dual-initiator/dual-adapter configuration to work properly, the ATF (Application-Transparent Failover) software is required. For further information, refer to the SRB (*Software Release Bulletin*) that comes with your cluster software.

Procedure Outlines

The procedure described in the HACMP documentation involves the steps below. For a detailed description, refer to chapter 4 "Maintaining Shared Volume Groups", of the *HACMP Administration Guide*.

1. If needed, complete prerequisite tasks. The shared physical volumes (**hdisks**) should be installed, configured and available (you can use the `lsdev -Cc disk` command to verify their status). In addition, HACMP must be stopped on all nodes.
2. On one node, create the desired shared volume group using the **smit mkvg** fastpath. Keep in mind that you must set the field **Activate volume group AUTOMATICALLY at system restart?** to **No**, and the field **ACTIVATE volume group after it is created?** to **Yes**.
3. Once the volume group is created, create the desired logical volumes and filesystems.
4. Unmount the newly created filesystem(s).
5. Vary off the newly created volume group, by using the `varyoffvg` command.
6. Import volume group information onto the second node: being logged in to the second node, use the `importvg` command.
7. Set up the imported volume group to remain dormant at startup: use the **smit mkvg** fastpath and set the field **Activate volume group AUTOMATICALLY at system restart?** to **No**.
8. Vary off the volume group.

Note that any volume group must be left in the state "varied off". They will be varied on by HACMP at cluster services startup.

Case of a Split-Bus Configuration

With the split-bus configuration, the shared volume group manual setup is a bit more complicated and involves the use of the **trespass** command which comes with the DAS software.

Below are outlined the steps involved when setting up shared volume groups for a split-bus configuration. For detailed instructions, you must refer to:

- The *Configuring and Managing a DAS* book; in particular, see chapter “Making physical disk units available to AIX”.
- The AIX and HACMP documentation; in particular, see chapter 4 “Maintaining Shared Volume Groups”, of the *HACMP Administration Guide*.

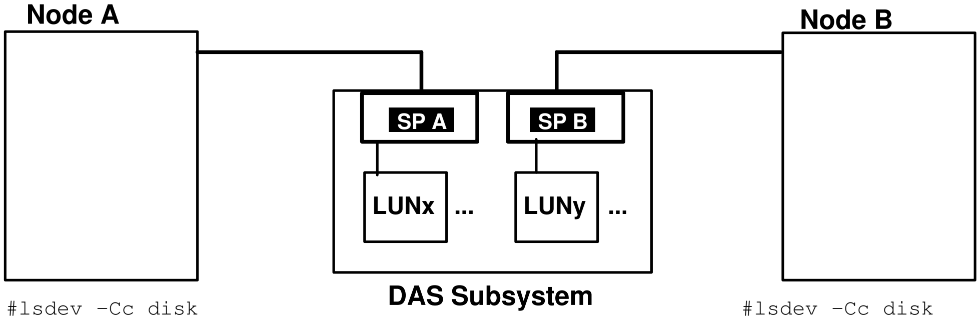
Procedure Outlines

Prerequisites and Assumptions

1. Make sure that HACMP is stopped on both cluster nodes, and shut down any applications running on the DAS disks.
2. For the following instructions, we assume that:
   - On the DAS, you have bound disk modules into LUNs (with the desired RAID levels). In the example below, we assume you have set up two LUNs, *LUNx* and *LUNy*.
   - We assume the systems are configured so each hdisk has the same name on both nodes. For example, we assume that *LUNx* is known as *hdiskx* on both nodes, and that *LUNy* is known as *hdisky* on both nodes.

     However, it is important to note that in practice, depending on your setup, a LUN can be assigned different names by the two nodes. For example, a LUN can be known as *hdisk4* on one node, and as *hdisk5* on the other node.
3. Note that initially, the disks that are seen as *available* from one node are seen only as *defined* (and thus not available) from the other node. You can use the `lsdev -Cc disk` command to verify the status of the **hdisks**.

   In our example, illustrated below, we see that *hdiskx* is *available* when seen from Node A but only *defined* when seen from Node B. (and vice–versa for *hdisky*).



```
#lsdev -Cc disk
...
hdiskx Available...
hdisky Defined...
...
```

```
#lsdev -Cc disk
...
hdiskx Defined...
hdisky Available...
...
```

Setting Up Volume Group and Filesystems on Node A

4. From the Node A, create a shared volume group including a disk available from Node A: *hdiskx* in our example. Use the **smit mkvg** fastpath. Keep in mind that you must set the field **Activate volume group AUTOMATICALLY at system restart?** to **No**, and the field **ACTIVATE volume group after it is created?** to **Yes**.
5. Once the volume group is created, create the desired logical volumes and filesystems.
6. Unmount the newly created filesystem(s), then vary off the newly created volume group, by using the `varyoffvg` command.

Setting Up Volume Groups and Filesystems on Node B

7. From the node B, create a shared volume group including a disk available from Node B: *hdisky* in our example. Use the **smit mkvg** fastpath. Keep in mind that you must set the field **Activate volume group AUTOMATICALLY at system restart?** to **No**, and the field **ACTIVATE volume group after it is created?** to **Yes**.

8. Once the volume group is created, create the desired logical volumes and filesystems.

9. Unmount the newly created filesystem(s), then vary off the newly created volume group by using the `varyoffvg` command.

Transfering to SP A the Ownership of the LUN Owned by SP B

10. Use the **trespass** command to transfer to *SP A* the ownership of the *LUNy*. (the trespass command is described in the DAS documentation). On the Node A, enter the command:

    ```
    /usr/sbin/cluster/events/utils/trespass –d –c hdisk<n> –l <m>
    ```

    This command tells the SP that currently owns `hdisk<n>` to take control of LUN `<m>`. In our example, the appropriate command is:

    ```
    /usr/sbin/cluster/events/utils/trespass –d –c hdiskx –l y
    ```

    (With this command, you tell to the SP A to take control of *LUNy*)

    Note that, instead of using the **trespass** command, you can use the **Change SP Ownership of a Physical Unit** of the **GridMgr** DAS menu.

Making Available To Node A the Newly Acquired hdisk

11. Make the *hdisky* disk available from Node A. On Node A, enter:

    ```
    mkdev –l hdisky
    ```

    Note that no application should attempt to access the DAS during the execution of the `mkdev` command.

    From Node A, *hdisky* is now seen as *available* (you can use `lsdev –Cc disk` to list hdisk states).

Setting Up the Volume Groups for the Newly Acquired hdisk

12. On node A, use the `importvg` command to import the volume group that includes the newly acquired disk (i.e. *hdisky*).

13. Set up the volume group to remain dormant at startup: use the **smit chvg** fastpath and set the field **Activate volume group AUTOMATICALLY at system restart?** to **No**.

14. Vary off the volume group.

Unconfigure the hdisk

15. On Node A, unconfigure *hdisky* by entering the following command:

    ```
    rmdev –l hdisky
    ```

    From the Node A, *hdisky* is now seen as only *defined* (i.e., it is no longer *available*).

Transferring again to SP B the Ownership of the Involved LUN

16. Use the **trespass** command **on Node B** to transfer to *SP B* the ownership of the *LUNy*. In our example, the appropriate command to enter on node B is:

    ```
    /usr/sbin/cluster/events/utils/trespass –d –c hdisky –l y
    ```

    With this command SP B retakes control of *LUNy*. From the LUN ownership standpoint, the configuration returns to its initial state (as depicted in the figure at the beginning of the procedure).

Note: The steps that follow are the same as steps 10 to 16, except they are performed from Node B (instead of Node A) and concern *hdiskx* (instead of *hdisky*).

Steps to Perform on Node B

17. Use the **trespass** command to transfer to *SP B* the ownership of the *LUNx*. (the trespass command is described in the DAS documentation). On the node B, enter the command:

    ```
    /usr/sbin/cluster/events/utils/trespass –d –c hdisky –l x
    ```

    (With this command, you tell to the SP B to take control of *LUNx*.

18. Make the *hdiskx* disk available from Node B. On Node B, enter:

    ```
    mkdev –l hdiskx
    ```

    From the Node B, *hdiskx* is now seen as *available* (you can use use `lsdev –Cc disk` to list hdisk states).

19. On Node B, use the `importvg` command to import the volume group that includes the newly acquired disk (i.e. *hdiskx*).

20. Set up the volume group to remain dormant at startup: use the **smit chvg** fastpath and set the field **Activate volume group AUTOMATICALLY at system restart?** to **No**.

21. Vary off the volume group.

22. On Node B, un-configure *hdiskx* by entering the following command:

    ```
    rmdev –l hdiskx
    ```

    From the Node B, *hdiskx* is now seen as only *defined* (i.e., it is no longer *available*).

Last Step to Perform on Node A

23. Use the **trespass** command **on Node A** to transfer to *SP A* the ownership of the *LUNx*. In our example, the appropriate command to enter on node A is:

    ```
    /usr/sbin/cluster/events/utils/trespass –d –c hdiskx –l x
    ```

The procedure is complete. (Note that any volume group is left in the state “varied off”. They will be varied on by HACMP at cluster services startup.)

Where You Go From Here

The tasks in this chapter complete the cluster setup procedure.

- At this point, you have a cluster that is ready to operate. Refer to the HACMP documentation for details on customizing the HACMP configuration and starting cluster services.
- You may want to take a look at the “System Management Tools and Tips” chapter, starting on page 10-1, for information related to cluster management.

Chapter 9. Managing Cluster Resources and Shared LVM Components

This chapter explains how to manage cluster resources and shared LVM components using the **Cluster Resources Modification** menu that comes with the BullCluster software package. It includes the following sections:

- Overview and Menu Summary.
- Adding, Removing, or Showing a Resource Group, on page 9-3.
- Adding or Removing a Volume Group, on page 9-7.
- Adding or Removing a Logical Volume, on page 9-10.
- Adding or Removing a File System, on page 9-12.
- Adding or Removing a Service IP Label, on page 9-15.
- Adding or Removing an Application Server, on page 9-17.
- Propagating Volume Group Definition to All Nodes, on page 9-18.

Overview and Menu Summary

Managing cluster resources and shared LVM components involves complex procedures, based on using multiple HACMP and AIX menus and commands. These procedures are described in the *HACMP Administration Guide*.

To make these administrative tasks simpler, the BullCluster software package provides the **Cluster Resources Modification** menu. This menu offers easy-to-use functions to manage cluster resources and shared LVM components (volume groups, logical volumes, filesystems). Each management task is achieved through a specific, unique, submenu, so that you do not have to go back and forth between the different HACMP and AIX menus.

Using the **Cluster Resources Modification** menu not only saves time, but also is safer, since it ensures that all the required configuration/de-configuration operations are performed and checked as appropriate.

Accessing the Cluster Resources Modification Menu

To access the **Cluster Resources Modification** menu:

- From the main HACMP SMIT menu (accessed with **smit hacmp**), select **Bull Cluster Easy Configuration**, then **Cluster Resources Modification**. The screen illustrated below is displayed.

Note for PowerConsole Users

If you are equipped with a PowerConsole, you can take advantage of the ClusterAssistant application. Indeed, the “Configuration” folder of the ClusterAssistant application group includes the “Resources” folder, whose icons allows you to easily access the various features of the **Cluster Resources Modification** menu.

Menu Summary

The options of the **Cluster Resources Modification** menu are summarized below.

Add a Resource Group (see page 9-3)

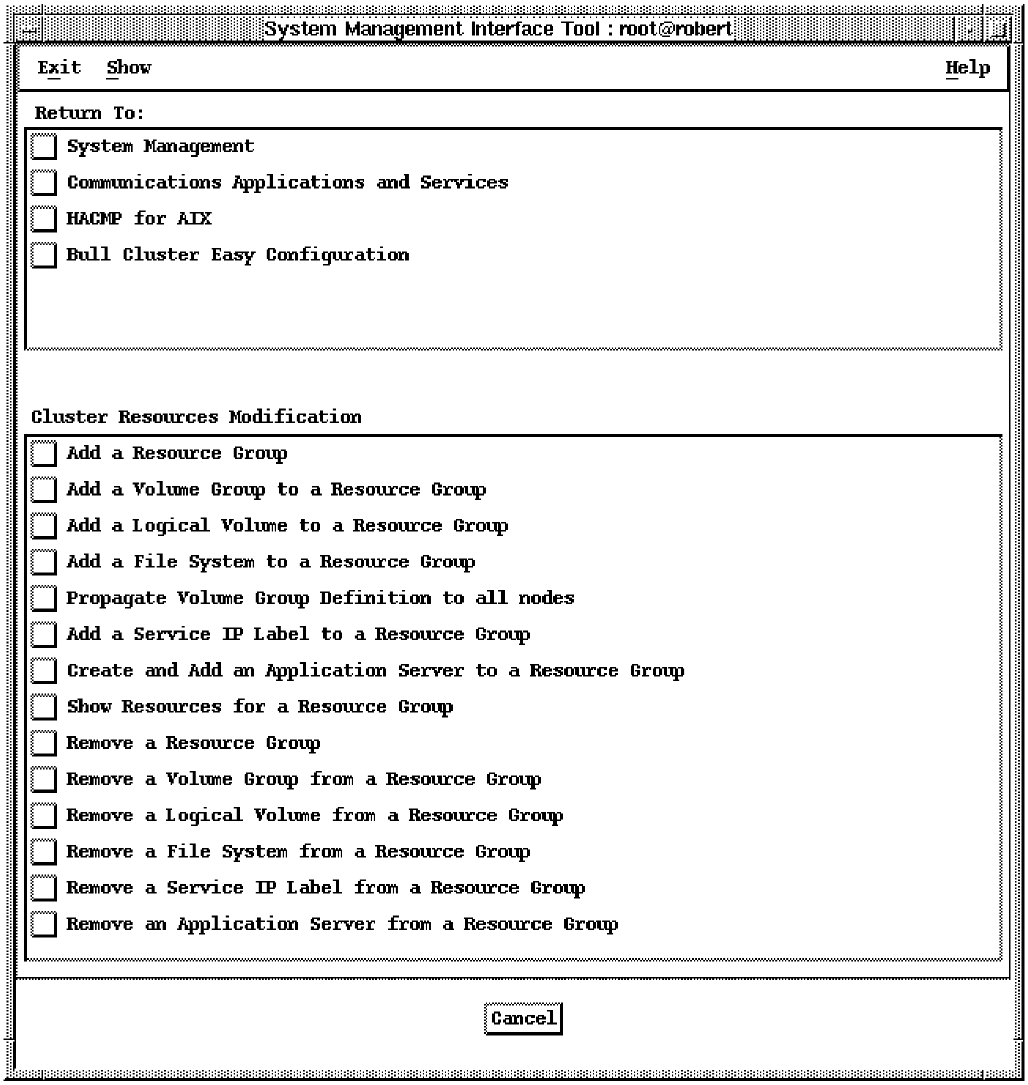
Creates a new resource group.

Add a Volume Group to a Resource Group (see page 9-7)

Creates a new shared volume group (or uses an already existing shared volume group), and declares it as a resource of a given resource group. Appropriate operations required to set up the volume group (such as vary on/off, export, import) are automatically performed.

Add a Logical Volume to a Resource Group (see page 9-10)
Creates a new logical volume and declares the volume group to which it belongs as a resource of a given resource group. The characteristics of the logical volume are configurable (mirroring, size...), and appropriate operations required to set up the corresponding volume group (such as vary on/off, export, import) are automatically performed.

Add a File System to a Resource Group (see page 9-12)
Creates a new filesystem along with a new logical volume, and declares the filesystem as a resource of a given resource group. The characteristics of the filesystem are configurable (mirroring, size...), and appropriate operations required to set up the corresponding volume group (such as vary on/off, export, import) are automatically performed.



Return To:

□ System Management
□ Communications Applications and Services
□ HACMP for AIX
□ Bull Cluster Easy Configuration

Cluster Resources Modification

□ Add a Resource Group
□ Add a Volume Group to a Resource Group
□ Add a Logical Volume to a Resource Group
□ Add a File System to a Resource Group
□ Propagate Volume Group Definition to all nodes
□ Add a Service IP Label to a Resource Group
□ Create and Add an Application Server to a Resource Group
□ Show Resources for a Resource Group
□ Remove a Resource Group
□ Remove a Volume Group from a Resource Group
□ Remove a Logical Volume from a Resource Group
□ Remove a File System from a Resource Group
□ Remove a Service IP Label from a Resource Group
□ Remove an Application Server from a Resource Group

Propagate Volume Group Definition to all nodes (see page 9-18)
Propagates volume group information across all the cluster nodes.

Add a Service IP Label to a Resource Group (see page 9-15)
Declares a given Service IP Label as a resource of a given resource group (this enables IP address takeover).

Create and Add an Application Server to a Resource Group (see page 9-17)
Sets up application servers and resources dedicated to specific applications.

Show Resources for a Resource Group (see page 9-6)
Shows the resources currently associated with a given resource group.

Remove a Resource Group (see page 9-4)
Removes the definition of a given resource group and synchronizes information across all cluster nodes.

Remove a Volume Group from a Resource Group (see page 9-9)
Physically removes a volume group and updates the resource group definitions accordingly.

Remove a Logical Volume from a Resource Group (see page 9-12)
Physically removes a logical volume and updates the resource group definitions accordingly.

Remove a File System from a Resource Group (see page 9-14)
Physically removes a filesystem and updates the resource group definitions accordingly.

Remove a Service IP Label from a Resource Group (see page 9-16)
Removes a service IP Label from the list of resources associated with a given resource group (this disables IP address takeover).

Remove an Application Server from a Resource Group (see page 9-17)
Removes an application server from the list of resources associated with a given resource group.

Adding, Removing, or Showing a Resource Group

This section discusses the three following options, available from the **Cluster Resources Modification** menu:

- **Add a Resource Group**
- **Remove a Resource Group**
- **Show Resources for a Resource Group**

Note: These menus are similar in functionality to what is offered in the **Define Resource Groups** menu of HACMP. In addition, they automatically perform information synchronization across cluster nodes.

For information related to resource group configuration, refer to the section “Configuring Resources” in the chapter “Configuring Cluster Resources” of the *HACMP Installation Guide*.

Add a Resource Group

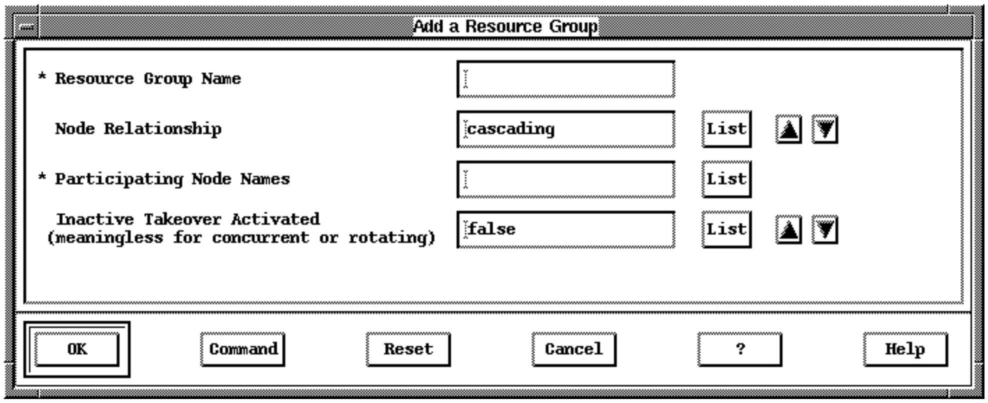
The **Add a Resource Group** menu performs the following:

- Creation of a new resource group (configured in accordance with your settings).
- Synchronization of the information across the cluster nodes.

Utilization

1. Using the **Add a Resource Group** requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).

2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Add a Resource Group**. The screen shown below is displayed.



Add a Resource Group

* Resource Group Name

Node Relationship: cascading

* Participating Node Names

Inactive Takeover Activated (meaningless for concurrent or rotating): false

3. Enter field values:

Resource Group Name
Enter the desired name. Use no more than 31 characters. You can use alphabetic or numeric characters and underscores.

Node Relationship
Toggle the entry field between **cascading**, **concurrent**, and **rotating**.

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.

Inactive Takeover Activated (meaningless for concurrent or rotating)
Set this parameter to control the initial acquisition of a resource by a node when the node/resource relationship is **cascading**. This parameter does not apply to **concurrent** or **rotating** resource groups.

If the field value is **true**, then the first node in the resource chain to join the cluster acquires the resource. Subsequently the resource will cascade to nodes in the chain with higher priority as they join the cluster.

If the field value is **false**, only the node with the highest priority for that resource will initially acquire the resource. Nodes lower in the resource chain only acquire the resource during a fallover.

4. Press Enter to validate the screen and to create the resource group.

5. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

Remove a Resource Group

The **Remove a Resource Group** option performs the following:

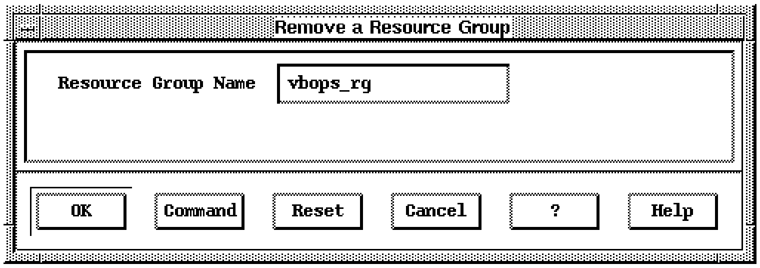
- Removal of the specified resource group from the cluster configuration. The resource group definition is removed from the ODM HACMP database, without physically impacting the resources (such as volume groups, filesystems...) that were associated with the resource group.
- Synchronization of the information across the cluster nodes.

Utilization

1. Using the **Remove a Resource Group** requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).

2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Remove a Resource Group**.

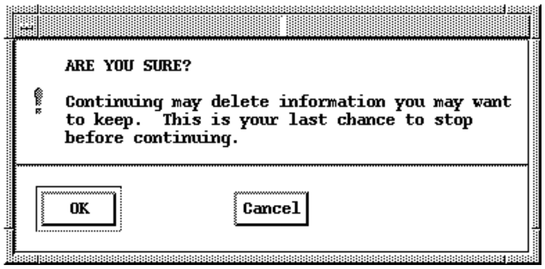
3. From the displayed list, select the resource group you want to remove. The Remove a Resource Group screen is displayed, as shown below.



Resource Group Name vbops_rg

4. The **Resource Group Name** field is not editable. Make sure it shows the resource group you want to remove.

5. When you validate the screen, a confirmation message is displayed, as shown below.



ARE YOU SURE?

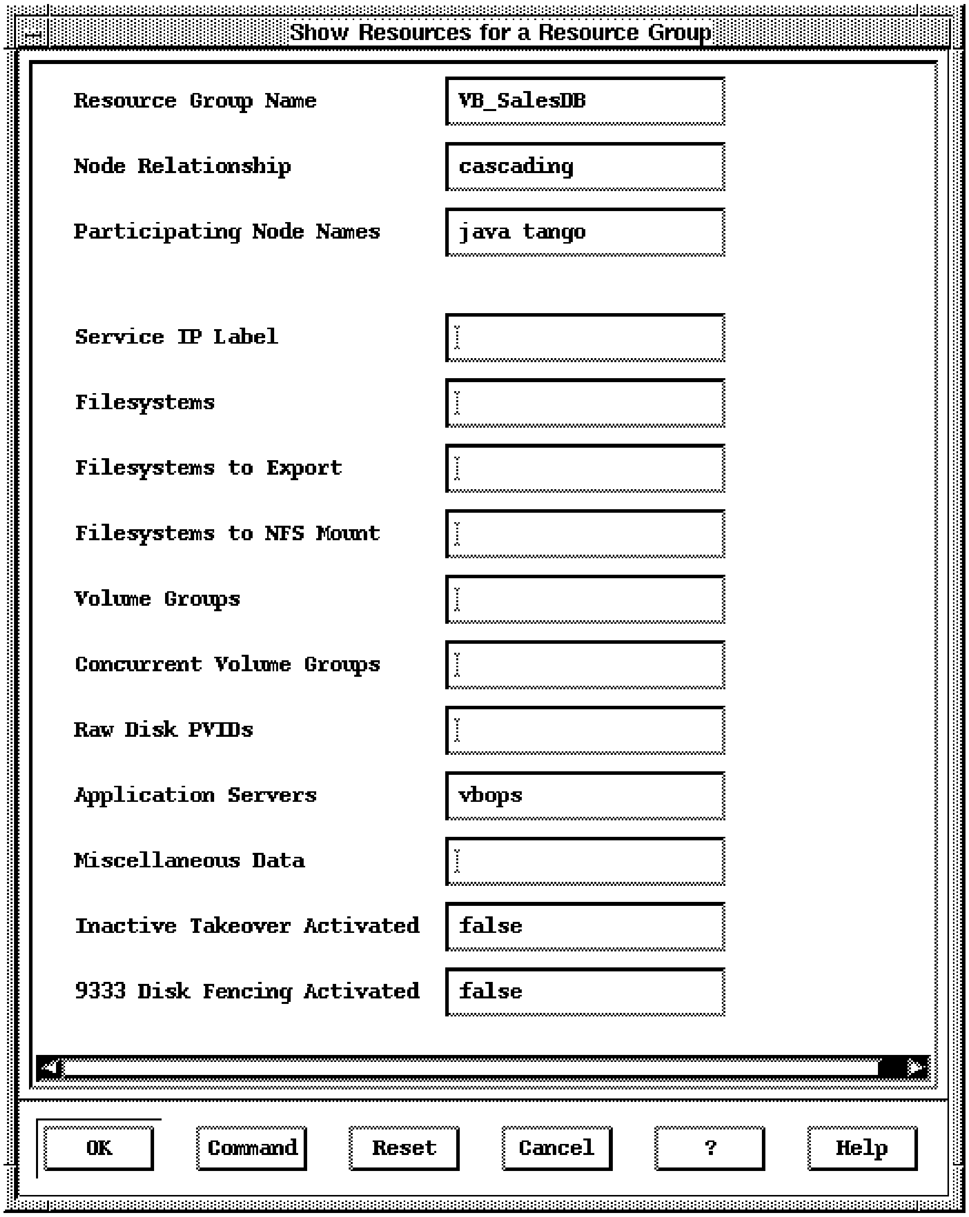
Continuing may delete information you may want to keep. This is your last chance to stop before continuing.

6. Press Enter to remove the resource group.

7. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

Show Resources for a Resource Group

The **Show Resources for a Resource Group** option displays the resources that are currently associated with an existing resource group.



Resource Group Name	VB_SalesDB
Node Relationship	cascading
Participating Node Names	java tango
Service IP Label	
Filesystems	
Filesystems to Export	
Filesystems to NFS Mount	
Volume Groups	
Concurrent Volume Groups	
Raw Disk PVIDs	
Application Servers	vbops
Miscellaneous Data	
Inactive Takeover Activated	false
9333 Disk Fencing Activated	false

The fields in this screen are not editable. For details on the meaning of the different fields, refer to section “Managing Resources” in the chapter “Changing Run-Time Parameters and Resource Groups” of the *HACMP Administration Guide*.

Adding or Removing a Volume Group

This section discusses the following two options, available from the **Cluster Resources Modification** menu:

- **Add a Volume Group to a Resource Group**
- **Remove a Volume Group from a Resource Group**

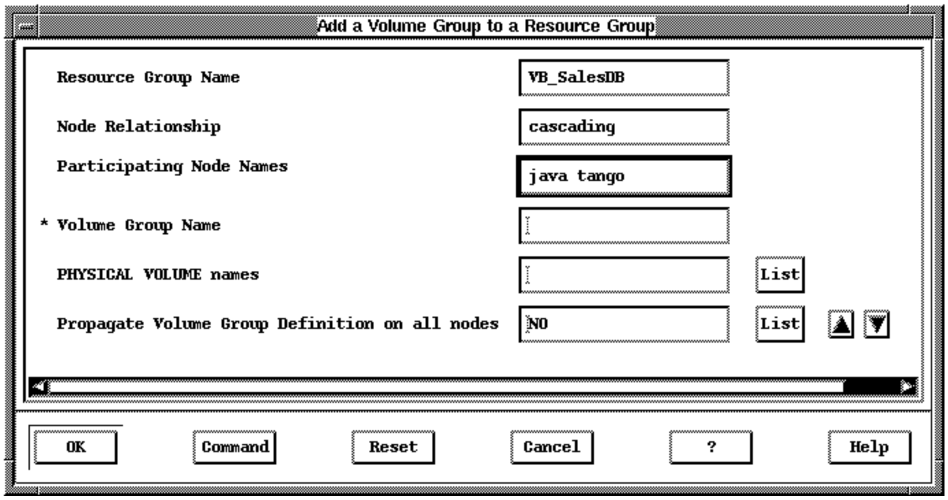
Add a Volume Group to a Resource Group

The **Add a Volume Group to a Resource Group** menu performs the following:

- If the volume group you specify does not exist, it creates it. From any node, the shared volume group will always be seen as having the same major number. An appropriate physical partition size is chosen.
- In any case (whether the shared volume group is new or already existed):
  - The shared volume group is varied off. This is a desired feature, because varying on shared volume groups is under the responsibility of HACMP (this is done when cluster services are started).
  - The shared volume group is made known to all the nodes. This is a very valuable feature, since the manual procedure is rather complex. (The manual procedure involves commands such as **varyon**, **varyoff**, **exportvg**, **importvg** and **trespass**, see "Dealing with Shared Volume Groups on DAS Subsystems", on page 8-5).
- The volume group is declared as a resource of the specified resource group.
- The cluster configuration is synchronized across the cluster nodes.

Utilization

1. Using the **Add a Volume Group to a Resource Group** requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).
2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Add a Volume Group to a Resource Group**.
3. From the displayed list, choose the resource group to which you want to add a volume group. The **Add a Volume Group to a Resource Group** screen is displayed, as illustrated below.



Add a Volume Group to a Resource Group

Resource Group Name	VB_SalesDB	
Node Relationship	cascading	
Participating Node Names	java tango	
* Volume Group Name		
PHYSICAL VOLUME names		List
Propagate Volume Group Definition on all nodes	NO	List

4. Note that **Resource Group Name**, **Node Relationship** and **Participating Node Names** are not editable fields. They are just provided as a reminder of the characteristics of the resource group to which you are about to add a volume group.

5. Enter field values:

Volume Group Name

Enter either the name of an existing volume group or the desired name for the new volume group to create.

PHYSICAL VOLUME names

If you are about to create a new volume group, specify here the names of the shared disks that must comprise the volume group. The list may include as many disk names as needed. For example, if you specify "`hdisk4`", the new volume group will include a single shared disk, `hdisk4`. If you specify "hdisk4 hdisk5", it will include two shared disks, `hdisk4` and `hdisk5`.

Leave the field blank if you are setting up a volume group that already exists.

Note about AIX Mirroring

AIX mirroring is a concern if the shared disks are ordinary (non-RAID) disks; note the following:

– Mirroring is possible only if the volume group includes more than one disk. So, if you intend to implement logical volume mirroring, you must specify at least two disks in the **PHYSICAL VOLUME names** field.

– If you decide to create a volume group that includes two disks in order to implement mirroring, you should specify, if possible, two disks that are installed on two distinct SCSI buses. In that case, a disk on a given SCSI bus is mirrored on the disk of the other SCSI bus. This configuration is preferable to increase data availability. On the contrary, if the two disks are on the same SCSI bus, the mirroring feature takes place on this SCSI bus. In this latter case, data availability is not optimal in case of bus failure.

– If you decide to create a volume group that includes more than two disks, it is not guaranteed that the mirroring feature will take place across the two SCSI buses. This restriction applies even when the specified disks are not all on the same SCSI bus.

– Similarly, if the cluster is equipped with SSA disk subsystems using two SSA loops, mirroring should be implemented across the two loops (disks on one loop should be mirrored to disks on the other loop.).

–For details on logical volumes and mirroring, refer to the chapter "Logical Volumes" in the *AIX System Management Guide*.

Propagate Volume Group Definition on all nodes

Enter `YES` so that the cluster configuration changes are immediately synchronized across the nodes.

Note that, since synchronization is time-consuming, you may prefer to specify `NO` if you are about to set up numerous shared LVM components (either volume groups, logical volumes or filesystems). In that case, once you have set up all the LVM components, do not forget to synchronize the cluster configuration using the **Propagate Volume Group Definition to all nodes** menu (see page 9-18). If in doubt, enter `YES`.

6. Press Enter to validate the screen, and read the messages that are displayed.

7. Once the command is executed, you can set up other shared LVM components using the menus for adding a volume group, a logical volume, or a file system. Then, if needed, do not forget to synchronize the cluster configuration using the **Propagate Volume Group Definition to all nodes** menu (see page 9-18).

8. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

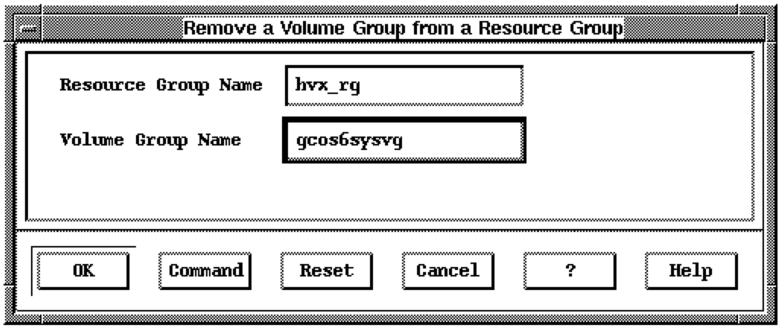
Remove a Volume Group from a Resource Group

The **Remove a Volume Group from a Resource Group** menu performs the following:

- Physically destroys the shared volume group you specify.
- Removes the definition of the destroyed volume group from every cluster node.
- Updates and synchronizes the cluster resources definitions across the cluster nodes.

Utilization

1. Using the **Remove a Volume Group from a Resource Group** requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).
2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Remove a Volume Group from a Resource Group**.
3. From the displayed list, select a resource group in which the volume group you want to destroy is declared as a resource.
4. If several volume groups are declared as members of the selected resource group, a volume group list is displayed. Select the volume group you want to destroy. The **Remove a Volume Group from a Resource Group** screen is displayed, as illustrated below.

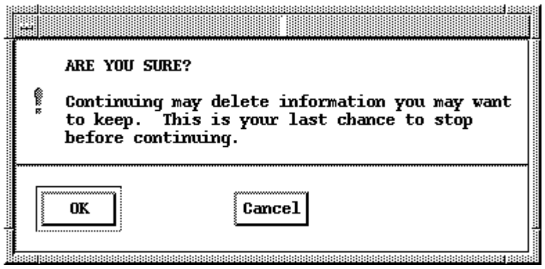


Remove a Volume Group from a Resource Group

Resource Group Name: hvx_rg

Volume Group Name: gcos6sysvg

5. The fields in this screen are not editable. Make sure that the **Volume Group Name** field shows the volume group you want to destroy.
6. When you validate the screen, a confirmation message is displayed, as shown below.



ARE YOU SURE?

Continuing may delete information you may want to keep. This is your last chance to stop before continuing.

7. Press Enter to destroy the volume group.
8. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

Adding or Removing a Logical Volume

This section discusses the following two options, available from the **Cluster Resources Modification** menu:

- **Add a Logical Volume to a Resource Group**
- **Remove a Logical Volume from a Resource Group**

Add a Logical Volume to a Resource Group

Preliminary Remarks

The **Add a Logical Volume to a Resource Group** menu is of interest if you are implementing a concurrent access cluster configuration. Indeed, the concurrent mode makes use of raw logical volumes to store data on shared disks. The AIX JFS (Journaled File System) is not supported, and you cannot use filesystems.

Note: For related information, refer to “Maintaining Shared LVM Components in a Concurrent Access Environment”, in the *HACMP Administration Guide*.

On the other hand, if you are implementing a non-concurrent configuration and you make use of filesystems (the usual case), the **Add a Logical Volume to a Resource Group** is of little interest: use the **Add a File System to a Resource Group** menu (explained on page 9-12).

Actions Performed

The **Add a Logical Volume to a Resource Group** menu performs the following:

- It creates a new logical volume (in an already existing shared volume group) with the characteristics you specify in the menu.
- It makes the logical volume known to all the nodes. This is a very valuable feature, since the manual procedure is rather complex. (The manual procedure involves commands such as **varyon**, **varyoff**, **exportvg**, **importvg** and **trespass**, see “Dealing with Shared Volume Groups on DAS Subsystems”, on page 8-5).
- It ensures that the volume group in which the new logical volume is created is declared as a resource of the specified resource group.
- It updates and synchronizes the cluster resource definitions across the cluster nodes.

Utilization

1. Using the **Add a Logical Volume to a Resource Group** requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).
2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Add a Logical Volume to a Resource Group**.
3. From the displayed list, choose a resource group to which belongs the volume group in which you want to create a new logical volume. The **Add a Logical Volume to a Resource Group** screen is displayed.
4. Note that **Resource Group Name**, **Node Relationship**, **Participating Node Names**, **Volume Group Name** and **Volume Group Disk(s)** are not editable fields. They are provided as a reminder of the characteristics of the involved resource group and volume group.
5. Enter field values:

   Logical Volume Name

   Enter a unique name (up to 15 characters) for the logical volume you want to create. Leave this field blank to accept the default name generated when creating the logical volume. (Note that within a cluster, the name of any shared logical volume must be unique. Moreover, those unique names must be the same on all nodes sharing the logical volumes. Checks are automatically performed when you apply this menu.)

Logical Volume Disks

When a logical volume is created, the default allocation policy is to use a minimum number of disks per logical volume copy, and to place the physical partitions belonging to a copy as contiguously as possible.
– To accept the default allocation policy, leave this field blank.
– If you prefer to create a logical volume whose physical partitions are spread across several disks, specify the desired disks. Leave a space between disk names (example: “`hdisk5 hdisk6`”).

AIX Mirroring

Specify in this field `yes` or `no` depending on whether or not you want to implement mirroring.
– If the shared disks are ordinary (non-RAID) disks: you probably want to specify `yes` to enable mirroring and thus, to increase data availability. This setting means you want to maintain two copies of the shared JFS log and of the logical volumes.
– If the shared disks are RAID disk devices: specify `no`. Indeed, since RAID disk devices provide their own data redundancy management, mirroring is useless.

Concerning the mirroring issue, which is a concern if the shared disks are ordinary (non-RAID) disks, see also the “Note about AIX Mirroring”, on page 9-8

Logical Volume Size (in PPs)

Specify the number of physical partitions to allocate to the logical volume.

Propagate Volume Group Definition on all nodes

Enter `YES` so that the cluster configuration changes are immediately synchronized across the nodes.

Note that, since synchronization is time-consuming, you may prefer to specify `NO` if you are about to set up numerous shared LVM components (either volume groups, logical volumes or filesystems). In that case, once you have set up all the LVM components, do not forget to synchronize the cluster configuration using the **Propagate Volume Group Definition to all nodes** menu (see page 9-18). If in doubt, enter `YES`.

6. Press Enter to validate the screen, and read the messages that are displayed.
7. Once the command is executed, you can set up other shared LVM components using the menus for adding a volume group, a logical volume, or a file system. Then, if needed, do not forget to synchronize the cluster configuration using the **Propagate Volume Group Definition to all nodes** menu (see page 9-18).
8. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

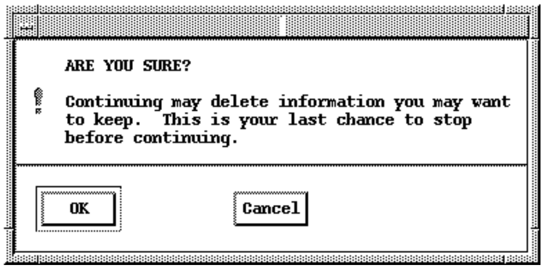
Remove a Logical Volume from a Resource Group

The **Remove a Logical Volume from a Resource Group** menu performs the following:

- Physically destroys the shared logical volume you specify.
- Removes the definition of the destroyed logical volume from every cluster node.
- Updates and synchronizes the cluster resources definitions across the cluster nodes.

Utilization

1. Using the **Remove a Volume Group from a Resource Group** requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).
2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Remove a Logical Volume from a Resource Group**.
3. From the displayed list, select a resource group to which belongs the volume group that includes the logical volume you want to destroy.
4. A list of logical volumes is displayed. Select the logical volume you want to destroy. The **Remove a Logical Volume Group from a Resource Group** screen is displayed.
5. The fields in this screen are not editable. Make sure that the **Logical Volume Name** field shows the logical volume you want to destroy.
6. When you validate the screen, a confirmation message is displayed, as shown below.



ARE YOU SURE?

Continuing may delete information you may want to keep. This is your last chance to stop before continuing.

7. Press Enter to destroy the logical volume.
8. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

Adding or Removing a Filesystem

This section discusses the following two options, available from the **Cluster Resources Modification** menu:

- **Add a File System to a Resource Group**
- **Remove a File System from a Resource Group**

Add a File System to a Resource Group

The **Add a File System to a Resource Group** menu performs the following:

- It creates a new filesystem along with a new logical volume (in an already existing shared volume group), with the characteristics you specify in the menu.
- It makes the filesystem known to all the nodes. This is a very valuable feature, since the manual procedure is rather complex. (The manual procedure involves commands such as **varyon**, **varyoff**, **exportvg**, **importvg** and **trespass**, see “Dealing with Shared Volume Groups on DAS Subsystems”, on page 8-5).
- It declares the new shared filesystem as a resource of the specified resource group.
- It updates and synchronizes the cluster resources definitions across the cluster nodes.

Utilization

1. Using the **Add a File System to a Resource Group** menu requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).
2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Add a File System to a Resource Group**.
3. From the displayed list, choose a resource group to which belongs the volume group in which you want to create a new filesystem.
4. A new screen is displayed that prompts you to specify a volume group. Select the volume group in which you want to create a new filesystem. The **Add a File System to a Resource Group** screen is displayed.
5. Note that **Resource Group Name**, **Node Relationship**, **Participating Node Names** , **Volume Group Name** and **Volume Group Disk(s)** are not editable fields. They are provided as a reminder of the characteristics of the involved resource group and volume group.
6. Enter field values:

Logical Volume Name

Enter a unique name (up to 15 characters) for the logical volume that will be created to hold the new filesystem. Leave this field blank to accept the default name generated when creating the logical volume.

Logical Volume Disks

When a logical volume is created, the default allocation policy is to use a minimum number of disks per logical volume copy, and to place the physical partitions belonging to a copy as contiguously as possible.
– To accept the default allocation policy, leave this field blank.
– If you prefer to create a logical volume whose physical partitions are spread across several disks, specify the desired disks. Leave a space between disk names (example: “`hdisk5 hdisk6`”).

AIX Mirroring

Specify in this field `yes` or `no` depending on whether or not you want to implement mirroring.
– If the shared disks are ordinary (non-RAID) disks: you probably want to specify `yes` to enable mirroring and thus, to increase data availability. This setting means you want to maintain two copies of the shared JFS log and of the logical volumes.
– If the shared disks are RAID disk devices: specify `no`. Indeed, since RAID disk devices provide their own data redundancy management, mirroring is useless.

Concerning the mirroring issue, which is a concern if the shared disks are ordinary (non-RAID) disks, see also the “Note about AIX Mirroring”, on page 9-8

Log Size (in PPs)

Specify the desired size for the JFS log, expressed in number of physical partitions.

Logical Volume Size (in PPs)

Specify the number of physical partitions to allocate to the logical volume.

File System Mount Point

Specify the desired mount point (i.e. the directory where the filesystem will be made available).

Propagate Volume Group Definition on all nodes

Enter `YES` so that the cluster configuration changes are immediately synchronized across the nodes.

Note that, since synchronization is time-consuming, you may prefer to specify `NO` if you are about to set up numerous shared LVM components (either volume groups, logical volumes or filesystems). In that case, once you have set up all the LVM components, do not forget to synchronize the cluster configuration using the **Propagate Volume Group Definition to all nodes** menu (see page 9-18). If in doubt, enter `YES`.

7. Press Enter to validate the screen, and read the messages that are displayed.
8. Once the command is executed, you can set up other shared LVM components using the menus for adding a volume group, a logical volume, or a file system. Then, if needed, do not forget to synchronize the cluster configuration using the **Propagate Volume Group Definition to all nodes** menu (see page 9-18).
9. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

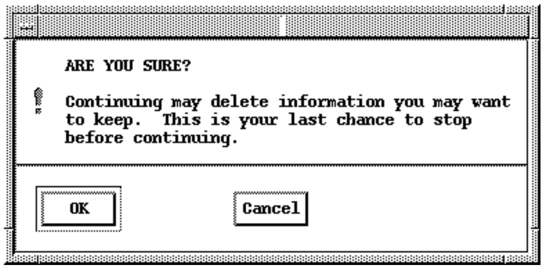
Remove a File System from a Resource Group

The **Remove a File System from a Resource Group** menu performs the following:

- Physically destroys the filesystem you specify.
- Removes the definition of the destroyed filesystem from every cluster node.
- Updates and synchronizes the cluster resources definitions across the cluster nodes.

Utilization

1. Using the **Remove a File System from a Resource Group** menu requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).
2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Remove a File System from a Resource Group**.
3. From the displayed list, select a resource group where the filesystem you want to destroy is declared as a resource.
4. A list of filesystems is displayed. Select the filesystem you want to destroy. The **Remove a File System from a Resource Group** screen is displayed.
5. The fields in this screen are not editable. Make sure that the **File System Name** field shows the filesystem you want to destroy.
6. When you validate the screen, a confirmation message is displayed, as shown below.



ARE YOU SURE?

Continuing may delete information you may want to keep. This is your last chance to stop before continuing.

7. Press Enter to destroy the filesystem.
8. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

Adding or Removing a Service IP Label

This section discusses the following two options, available from the **Cluster Resources Modification** menu:

- **Add a Service IP Label to a Resource Group**
- **Remove a Service IP Label from a Resource Group**

Note: For details on the concepts related to IP labels and IP address takeover, refer to your HACMP documentation.

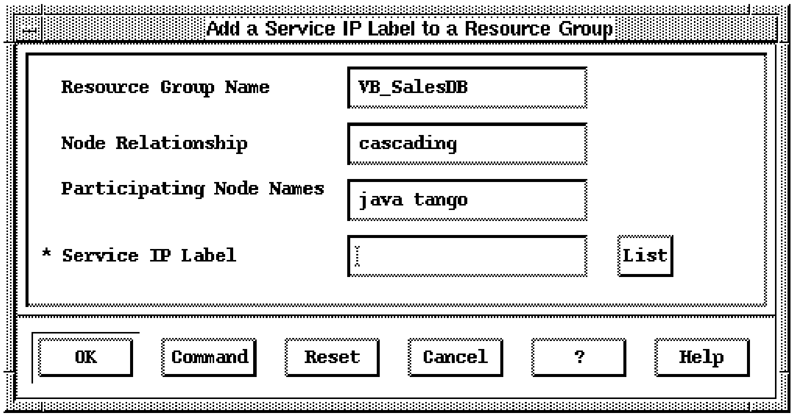
Add a Service IP Label to a Resource Group

The **Add a Service IP Label to a Resource Group** menu enables you to declare a service IP label as a resource of an existing resource group. In other words, it enables you to implement IP address takeover. The **Add a Service IP Label to a Resource Group** menu performs the following:

- It declares the specified service IP label as a resource of the selected resource group.
- It updates and synchronizes the cluster resources definitions across the cluster nodes.

Utilization

1. Using the **Add a Service IP Label to a Resource Group** menu requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).
2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Add a Service IP Label to a Resource Group**.
3. From the displayed list, choose the resource group to which you want to add a service IP label. The **Add a Service IP Label to a Resource Group** screen is displayed, as illustrated below:



Add a Service IP Label to a Resource Group

Resource Group Name VB_SalesDB

Node Relationship cascading

Participating Node Names java tango

* Service IP Label

4. Note that **Resource Group Name**, **Node Relationship** and **Participating Node Names** are not editable fields. They are just provided as a reminder of the characteristics of the selected resource group.
5. Specify the **Service IP Label** to be taken over when this resource group is taken over. Press F4 to see a list of valid IP labels.
6. Press Enter to add the selected service IP label to the resource group.
7. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

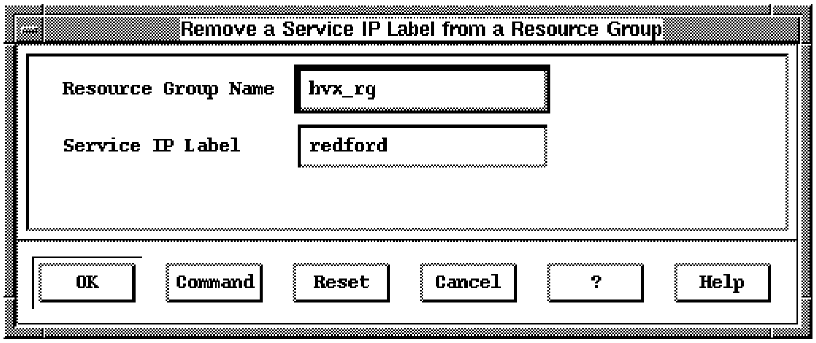
Remove a Service IP Label from a Resource Group

The **Remove a Service IP Label to a Resource Group** menu performs the following:

- It removes a service IP label from the list of resources associated with the specified resource group. (in other words, it disables IP address takeover).
- It updates and synchronizes the cluster resources definitions across the cluster nodes.

Utilization

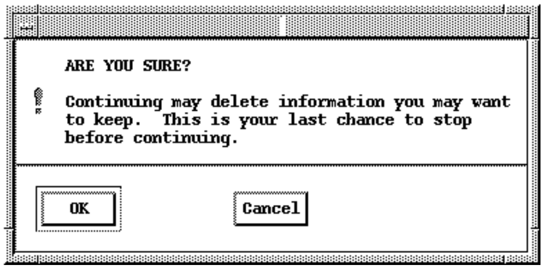
1. Using the **Remove a Service IP Label from a Resource Group** menu requires cluster services to be stopped. Consequently, ensure that cluster services are stopped (if needed, stop them using the **Stop Cluster Services** HACMP menu).
2. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Remove a Service IP Label from a Resource Group**.
3. From the two lists successively displayed, choose the resource group from which you want to remove the associate the service IP label, then select the appropriate service IP label. The **Remove a Service IP Label from a Resource Group** menu is displayed, as illustrated below:



Resource Group Name: hvx_rg

Service IP Label: redford

4. The fields in this screen are not editable. Make sure that the **Service IP Label** field shows the service IP label you want to remove from the specified resource group.
5. When you validate the screen, a confirmation message is displayed, as shown below.


ARE YOU SURE?

Continuing may delete information you may want to keep. This is your last chance to stop before continuing.

6. Press Enter to remove the service IP label from the resource group.
7. If desired, restart the cluster services using the **Start Cluster Services** HACMP menu.

Adding or Removing an Application Server

This section discusses the following two options, available from the **Cluster Resources Modification** menu:

- **Create and Add an Application Server to a Resource Group**
- **Remove an Application server from a Resource Group**

Create and Add an Application Server to a Resource Group

The BullCluster Software package includes sample start and stop scripts to help you to implement specific applications in your cluster. The **Create and Add an Application Server to a Resource Group** menu enables you to set up application servers dedicated to these specific applications. The menu creates an application server for HACMP, configuring it with the sample start and stop scripts that correspond to the chosen application.

The sample start and stop scripts as well as the **Create and Add an Application Server to a Resource Group** menu are mainly intended for Bull technical personnel.

- For an overview of the provided sample scripts, refer to “Planning Specific Applications”, on page 3-1.
- For details, refer to the *Powercluster & HA Solutions: Using the Sample HA Scripts* document.

Remove an Application Server from a Resource Group

If you want to remove an application server you have created using the **Add an Application Server to a Resource Group** menu, then use the **Remove an Application Server from a Resource Group**.

This ensures that all needed de-configuration operations are done properly.

Note: If you have created X25 or TUXEDO application servers with the **Create and Add an Application Server to a Resource Group** menu, using the **Remove an Application Server from a Resource Group** is mandatory to ensure that the specific objects defined in the ODM database are de-configured.

Propagating Volume Group Definition to All Nodes

The **Cluster Resources Modification** menu includes several options that allow you to create shared LVM components:

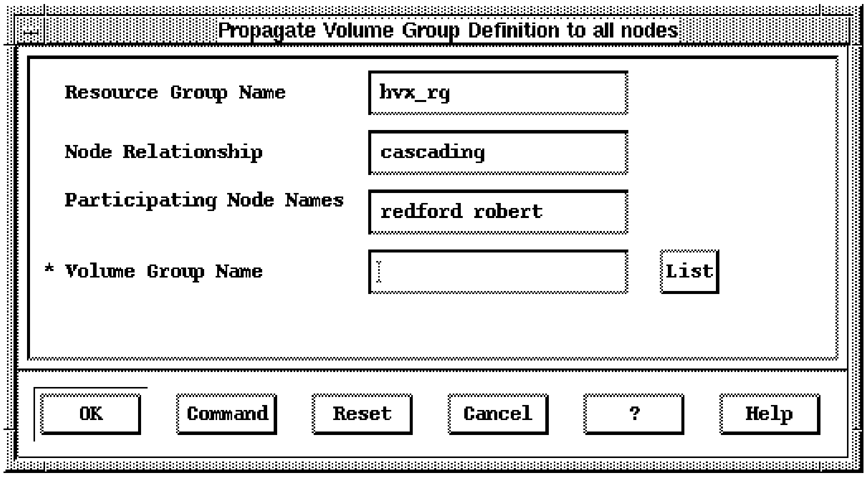
- **Add a Volume Group to a Resource Group**
- **Add a Logical Volume to a Resource Group**
- **Add a File System to a Resource Group**

Running the **Propagate Volume Group Definition on all nodes** option from the **Cluster Resources Modification** general menu is necessary only if:

> you have used one of the three menus mentioned above
> AND
> in these menus, you have specified NO in the **Propagate Volume Group Definition to all nodes** field.

In that case, you must make the concerned volume group(s) known to the different cluster nodes. Proceed as follows:

1. From the **Bull Cluster Easy Configuration** menu, choose **Cluster Resources Modification**, then **Propagate Volume Group Definition on all nodes**.
2. A list of resource groups is displayed. From this list, select a resource group to which you have added a LVM component using one of the three menu options mentioned above. The **Propagate Volume Group Definition on all nodes** menu is displayed, as illustrated below.



Propagate Volume Group Definition to all nodes

Resource Group Name	hvx_rg
Node Relationship	cascading
Participating Node Names	redford robert
* Volume Group Name	

3. Note that **Resource Group Name**, **Node Relationship** and **Participating Node Names** are not editable fields. They are provided as a reminder of the characteristics of the selected resource group.
4. In the **Volume Group Name** field, specify the volume group to be made known to all cluster nodes.
5. Press Enter to perform the operation.

Chapter 10. System Management Tools and Tips

The first two sections discuss two BullCluster utilities related to cluster management:

- Using the BullCluster Diagnostic Tool
- Using the Clusterview Utility, on page 10-3.

The two following sections discuss two BullCluster utilities related to cluster setup:

- Implementing IP Address Aliasing, on page 10-5.
- Implementing Rolling Applications, on page 10-7.
- Implementing Native HA, on page 10-16.
- Implementing the Extended HA Shared Storage, on page 10-22.
- Customizing Cluster Events Processing, on page 10-24.

Note: BullCluster includes other utilities useful for cluster setup and management.

The remaining sections (starting with page 10-25) give hints on using various tools useful for managing a cluster, but that are not included in the BullCluster software package:

- Synchronizing Time in Your Cluster with timed, on page 10-25.
- Hints for Dealing with iFOR/LS Nodelocked License Keys, on page 10-26.
- Hints for Using DSMIT, on page 10-26.
- Managing DAS Subsystems with ArrayGUIde, on page 10-26.
- Managing DAS Subsystems with Navisphere, on page 10-27.

Using the BullCluster Diagnostic Tool (bclerrdiag)

The BullCluster software package includes a diagnostic tool that allows you to generate diagnostic information. A cluster administrator may want to use this tool on a regular basis (for example daily) to check the health of the cluster.

The BullCluster diagnostic tool is provided as a UNIX command, **bclerrdiag**. Below is the man page for this command.

Note for PowerConsole Users

If you are equipped with a PowerConsole, you can take advantage of the ClusterAssistant application, which includes a windows-based utility ("Cluster Diagnostics" icon) that allows you to use the BullCluster diagnostic tool in an easy way.

Purpose

The **bclerrdiag** BullCluster diagnostic tool reports diagnostic information about cluster operation.

Description

bclerrdiag analyzes the state and consistency of cluster resources. In addition (and optionally, in accordance with the flags you specify), it performs further subsystem-specific diagnostics by scanning log files on the involved cluster nodes.

Cluster Resources Consistency Analysis

bclerrdiag analyzes the state and consistency of cluster resources. It looks for the definitions of HACMP resource groups, and checks the state of the different resources involved in these groups. After analysis, it reports any potential problem or inconsistency. For example, if none of the cluster nodes is assuming a resource that should be highly available (for example, a volume group or an IP address), this abnormal condition is reported.

Subsystem-Specific Diagnostics

In accordance with the flags you specify, the **bclerrdiag** tool is able to generate diagnostics related to these three subsystems: AIX, HACMP and DAS (Disk Array Subsystems):

- **AIX Subsystem**:
  The tool looks for hardware errors logged in the AIX error log file (**errlog**) of each involved node. Any logged hardware error is reported.
- **HACMP Subsystem**:
  The tool scans error messages logged by HACMP in the /**var**/**adm**/**cluster.log** file and reports messages that contain one of the strings `error`, `fail` and `network_down`.
- **DAS Subsystem**:
  The tool queries the event log (**DASlog**) of the DAS (Disk Array Subsystems) and reports relevant error messages, i.e. messages identified by `0x900`–series and `0xa00`–series codes.

Diagnostic Information Consolidation

bclerrdiag gathers the information extracted from the involved log files of the different nodes. Then it consolidates the collected data, taking into account timestamps and summarizing the information for repeated errors.

Syntax

/usr/sbin/bullcluster/monitoring/diagtool/**bclerrdiag** [–b *subsystems*]
[–n *nodes*]
[–p *period*]

The **bclerrdiag** command can be executed on any node, as long as this node is member of the cluster whose operation is to be checked. **bclerrdiag** reports diagnostic information to its standard output as well as to specific log files (see “Files” below).

Flags

[–b *subsystems*]
Lists the subsystems to which diagnostics must be applied. The *subsystems* parameter is a comma-separated list of subsystem names. Valid subsystem names are `AIX`, `HACMP` and `DAS`. By default (if the **–b** flag is not specified), diagnostics are performed against all these three subsystems. Note in addition that, whether or not the **–b** flag is specified, **bclerrdiag** always analyzes the state and consistency of cluster resources.

[–n *nodes*] Lists the nodes to be taken into account to perform the diagnostics. The *nodes* parameter is a comma-separated list of node names. By default (if the **–n** flag is not specified), **bclerrdiag** takes into account all the nodes that are members of the cluster.

[–p *period*] **bclerrdiag** scans the subsystem logs only for messages that occurred in the period specified by the *period* parameter. The default is a 1-day period (and thus, error messages older than one day are ignored). The *period* must be specified as an integer number of hours.

Files

/usr/sbin/bullcluster/monitoring/diagtool/bclerrdiag is the executable file.

The /**var**/**bullcluster**/**logs** directory is used to keep the various log files generated by **bclerrdiag** (these log files are overwritten each time **bclerrdiag** runs). The **cluster_global_report** log file contains the information that was displayed the last time the command was executed. **cluster_detailed_report** contains more detailed information, and can be of interest for troubleshooting purposes. Both files are stored on the node where the

bclerrdiag command was executed. Note that **bclerrdiag** stores other intermediate log files on the different concerned nodes.

Using the Clusterview Utility

The BullCluster software package includes **clusterview**, a utility that reports the status of clusters. This utility is provided as a UNIX command. It is comparable to **clstat**, which comes with the HACMP software.

Below is the **clusterview** man page.

Purpose

The clusterview utility is provided for reporting the status of HACMP clusters.

Syntax

/usr/sbin/cluster/**clusterview**

Description

The **clusterview** utility is a Clinfo client program that uses the Clinfo API and information in the **/usr/sbin/cluster/etc/clhosts** file to display node and interface information about all detected clusters in ASCII display mode. The **clinfo** program must be running to use the clusterview command.

clusterview reports whether the cluster is up, down or unstable. It also reports whether a node is up, down, leaving or joining, and the number of nodes in the cluster.

For each node, **clusterview** displays the IP label and address of each network interface attached to the node and whether this interface is up or down.

To run the utility, enter `/usr/sbin/cluster/clusterview`. You can also invoke **clusterview** through SMIT, by using the **smit clusterview** fastpath, or selecting the **Show Cluster State** option from the **Bull Cluster Easy Configuration** menu.

See a sample output below.

Example of Output

clusterview displays information similar to the following:

```
clinfo daemon is running.
Number of clusters active: 2

---------------------- CLUSTER blues --------------------------------

ID             NAME           STATE          SUBSTATE       PRIMARY NODE
4              blues          UP             STABLE         otis

Cluster blues (4) has 2 nodes:

NODE           STATE          CLUSTER ID     NB INTERFACES
otis           UP             4              2
        INTERFACE      STATE          ADDRESS
        otis_srv       UP             130.183.1.95
        otis_tty1      UP             0.0.0.0

NODE           STATE          CLUSTER ID     NB INTERFACES
redding        UP             4              2
        INTERFACE      STATE          ADDRESS
        redding_srv    UP             130.183.1.85
        redding_tty1   UP             0.0.0.0

---------------------- CLUSTER cluster_tests ----------------------

ID             NAME           STATE          SUBSTATE       PRIMARY NODE
27             cluster_tests  UP             STABLE         mars_srv

Cluster cluster_tests (27) has 2 nodes:

NODE           STATE          CLUSTER ID     NB INTERFACES
mars_srv       UP             27             2
        INTERFACE      STATE          ADDRESS
        mars_srv       UP             130.183.5.2
        mars_tty0      UP             0.0.0.0

NODE           STATE          CLUSTER ID     NB INTERFACES
mercure_srv    UP             27             2
        INTERFACE      STATE          ADDRESS
        mercure_srv    UP             130.183.5.1
        mercure_tty1   UP             0.0.0.0
```

The information displayed shows cluster name, cluster ID, cluster state and substate. In this example, the 2 clusters are up and have 2 nodes, which are also up. In addition, each node has 2 network adapters.

Files

/usr/sbin/cluster/clusterview is the executable file.
/usr/sbin/cluster/etc/clhosts is the file listing available host names.

Related Information

See the **clinfo** command in the HACMP documentation.

Implementing IP Address Aliasing

The BullCluster software package includes a menu option that enables IP address aliasing.

Purpose

IP address aliasing is an optional feature that may be of interest only if IP address takeover (IPAT) is implemented in the cluster.

When IPAT is enabled, a service adapter uses two different IP addresses, depending on the current state:

- at boot time, and until cluster services are started, the service adapter is assigned the boot IP address
- when cluster services are started on the node, it changes over to the service IP address

As an example, assume that the node `foo` has a service adapter associated with the `foo` service IP address and with the `foo_boot` boot IP address. Depending on the current cluster state, applications (or users) that want to connect to this service adapter must sometimes use the `foo` address, and at other times the `foo_boot` address. Moreover, an application that relies on the `foo` service address may actually connect to another cluster node that has taken over this service address.

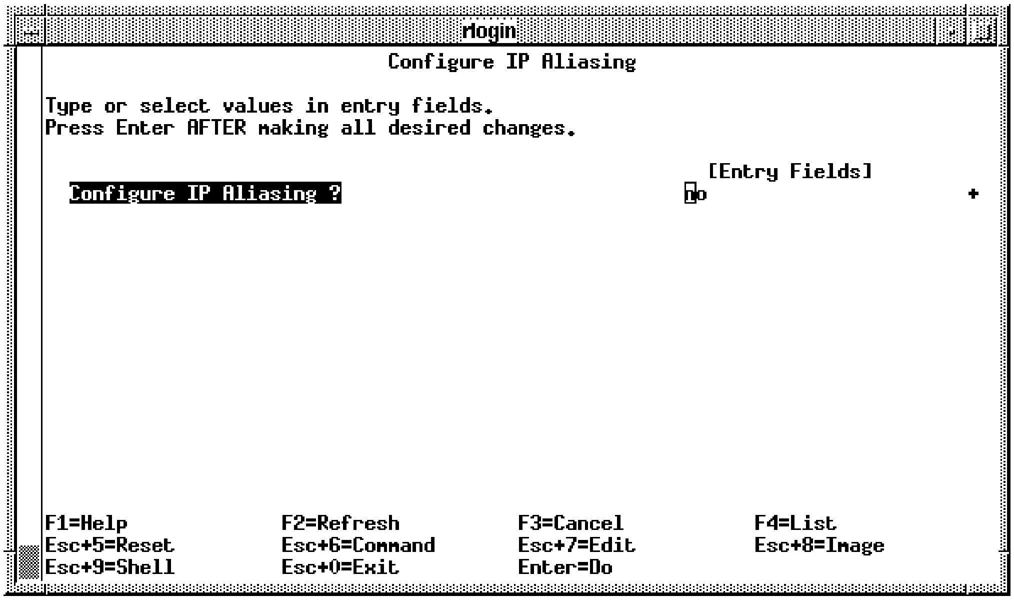
From the user's or application's standpoint, these changing addresses may be seen as unpractical. For example, to remote login to the `foo` node through its service adapter, you must sometimes enter "`rlogin foo`", sometimes "`rlogin foo_boot`", but you do not know in advance which address is valid.

IP address aliasing circumvents this problem, by making the service and boot addresses aliases.

Implementation Procedure

To implement the IP aliasing feature (optional):

1. **If cluster services are running, stop cluster services on all the nodes.**
2. From the main HACMP SMIT menu, successively select **Bull Cluster Easy Configuration** and **IP Aliasing Configuration**. The following screen is displayed:



```
                              Configure IP Aliasing

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                                         [Entry Fields]
  Configure IP Aliasing ?                               no                 +

F1=Help             F2=Refresh          F3=Cancel           F4=List
Esc+5=Reset         Esc+6=Command       Esc+7=Edit          Esc+8=Image
Esc+9=Shell         Esc+0=Exit          Enter=Do
```

3. The current **Configure IP Aliasing?** field shows the current setting (it shows "`no`" if IP aliasing is not currently enabled, otherwise it shows "`yes`"). Toggle the value as desired, then press Enter to validate your choice.
4. As soon as you validate the screen, IP aliasing is enabled (or disabled) on the current node, as well as on the other cluster nodes (those nodes whose service adapter is associated with both a service and a boot IP address).

Remarks:

For your information, note that the IP aliasing capability described above relies on the `ifconfig` AIX command, which allows to declare aliases for an IP address.

In addition, note that when IP aliasing is enabled:

- In the HACMP log files, the messages related to the `swap_adapter` event appear twice.
- A post event is customized for acquire_service_addr (based on BullCluster Event Customization Management).
- If an IP address takeover occurs, the service IP address of the failing node moves to the standby adapter on the backup node, and the service adapter of the failing node reverts to its boot address.

Implementing Rolling Applications

Purpose

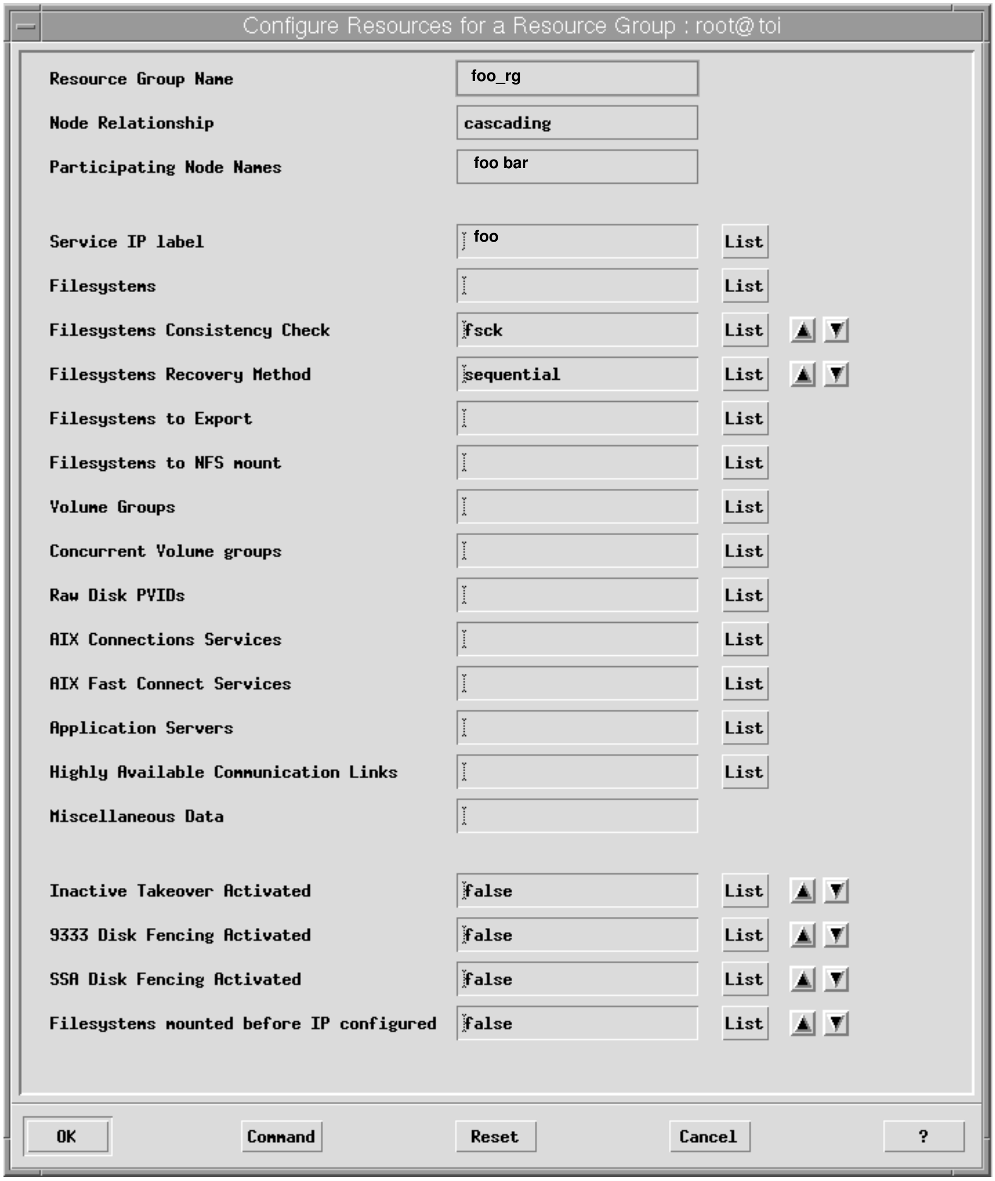
When several applications are running on a node for performance or maintenance operations, you need to transfer an application from a node to another without inducing a takeover.

Rolling Applications allows the transfer of an application (or a group of applications) from a node to another without inducing a takeover.

This function needs a particular configuration of the resources with HACMP:

- A resource group is created with a service IP label configured as a resource, but without having the application server, file system, or volume group configured.

  An example is given below:

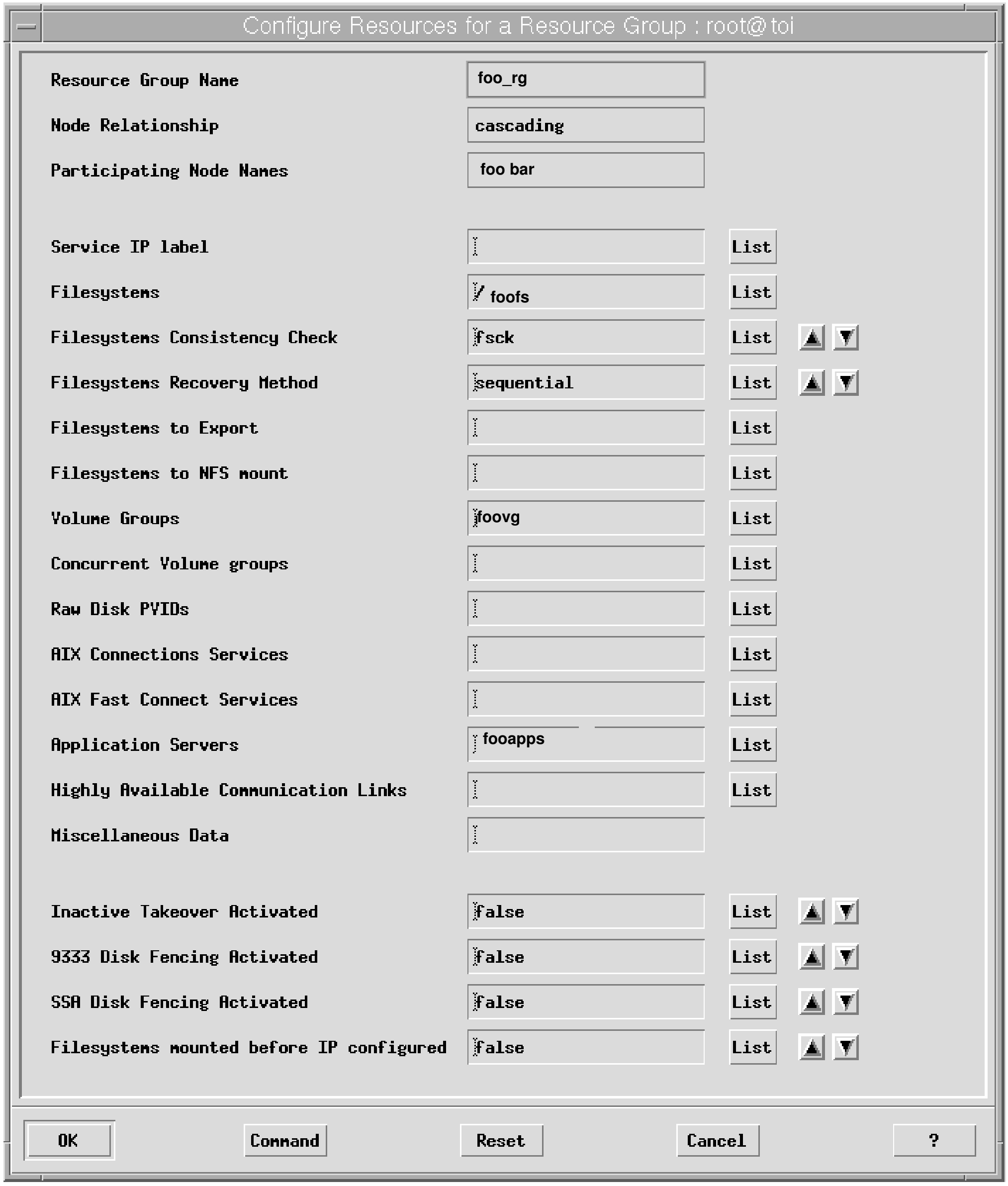


Configure Resources for a Resource Group : root@toi

Resource Group Name	foo_rg
Node Relationship	cascading
Participating Node Names	foo bar
Service IP label	foo
Filesystems	
Filesystems Consistency Check	fsck
Filesystems Recovery Method	sequential
Filesystems to Export	
Filesystems to NFS mount	
Volume Groups	
Concurrent Volume groups	
Raw Disk PVIDs	
AIX Connections Services	
AIX Fast Connect Services	
Application Servers	
Highly Available Communication Links	
Miscellaneous Data	
Inactive Takeover Activated	false
9333 Disk Fencing Activated	false
SSA Disk Fencing Activated	false
Filesystems mounted before IP configured	false

- For each application (or group of applications) which can be migrated, a resource group which contains application server(s) that starts the application(s) and lvm resources (volume groups, file systems) is created but without service IP labels configured.

  An example is given below:



Configure Resources for a Resource Group : root@ toi

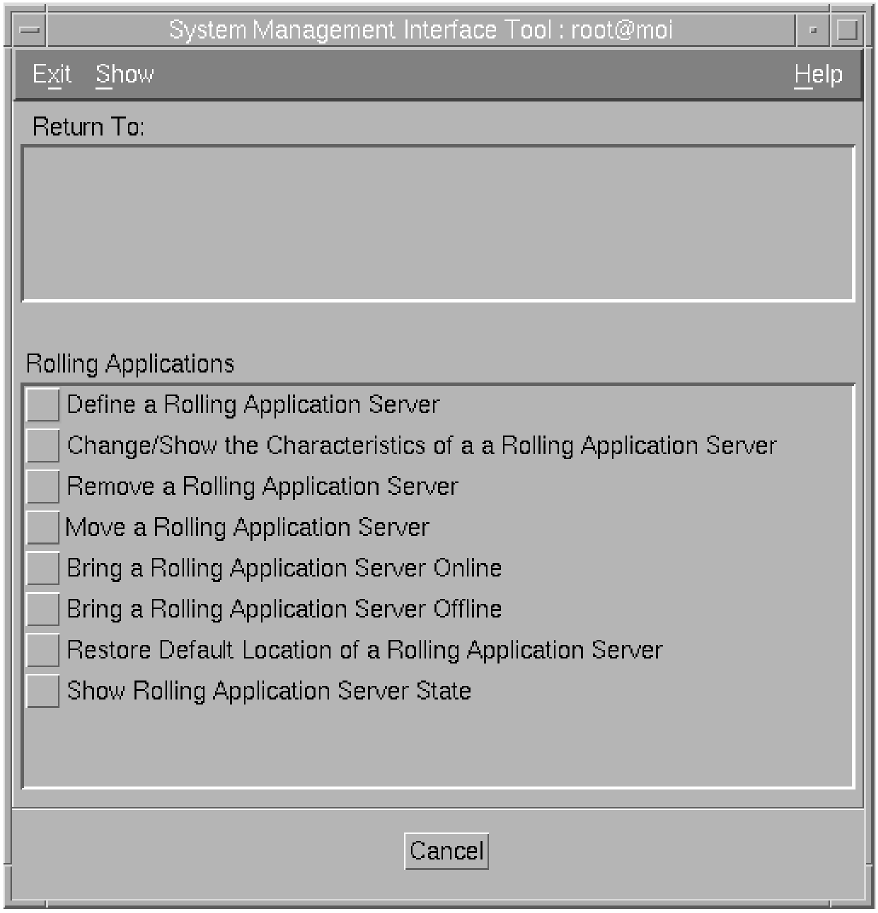
Resource Group Name	foo_rg
Node Relationship	cascading
Participating Node Names	foo bar
Service IP label	
Filesystems	/ foofs
Filesystems Consistency Check	fsck
Filesystems Recovery Method	sequential
Filesystems to Export	
Filesystems to NFS mount	
Volume Groups	foovg
Concurrent Volume groups	
Raw Disk PVIDs	
AIX Connections Services	
AIX Fast Connect Services	
Application Servers	fooapps
Highly Available Communication Links	
Miscellaneous Data	
Inactive Takeover Activated	false
9333 Disk Fencing Activated	false
SSA Disk Fencing Activated	false
Filesystems mounted before IP configured	false

As the purpose of this feature is to migrate an application (or a group of applications) without stopping HACMP, the dynamic configuration mechanism is used (*clRGdare* command).

Implementation Procedure

To implement the Rolling Applications feature (optional):

- From the main HACMP SMIT menu successively select **Bull Cluster Easy Configuration** and **Rolling Applications**. The following screen is displayed:


System Management Interface Tool : root@moi
Exit
Show
Help
Return To:
Rolling Applications
Define a Rolling Application Server
Change/Show the Characteristics of a a Rolling Application Server
Remove a Rolling Application Server
Move a Rolling Application Server
Bring a Rolling Application Server Online
Bring a Rolling Application Server Offline
Restore Default Location of a Rolling Application Server
Show Rolling Application Server State
Cancel

Overview of the Rolling Applications menu

When you use the features of the Rolling Applications menu, you can:

- Configure a Rolling Application Server
  - Define a Rolling Application Server
  - Change/Show the Characteristics of a Rolling Application Server
  - Remove a Rolling Application Server
- Manage a Rolling Application Server
  - Bring a Rolling Application Server Online
  - Bring a Rolling Application Server Offline
  - Move a Rolling Application Server
  - Restore Default Location of a Rolling Application Server

  CAUTION:
  Before managing a Rolling Application Server, be sure that you have configured the specified Rolling Application Server in an HACMP Resource Group.

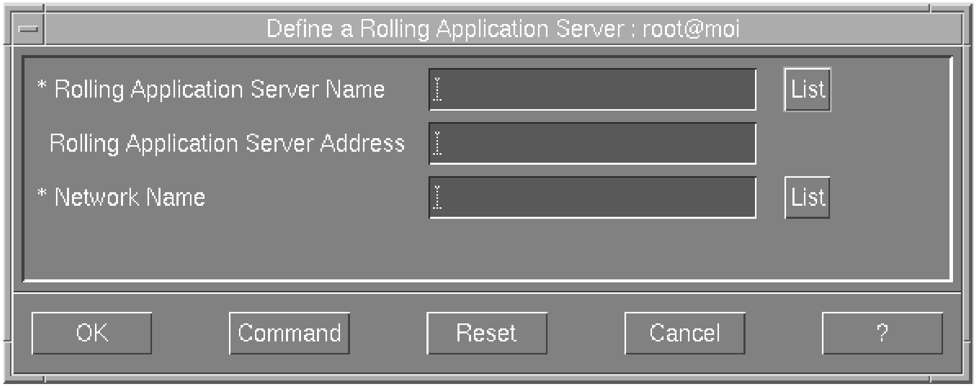
- View Rolling Application Server State
  - Show Rolling Application Server State

Define a Rolling Application Server

This action adds a Rolling Application Server in the configuration.

From any node, bring up the **Bullcluster Rolling Applications** menu, and select the **Define a Rolling Application Server** menu.

The **Define a Rolling Application Server** screen appears.



Define a Rolling Application Server : root@moi

* Rolling Application Server Name

Rolling Application Server Address

* Network Name

1. Select the Application Server Name from the list, or enter the Name for a Rolling Application Server. The List button gives the list of all Application Servers configured in an HACMP Resource Group.
2. Enter an IP address corresponding to the selected Rolling Application Server.
3. Select the Network Name from the list, or enter the Name of a Network.
4. Validate the screen to apply your settings.

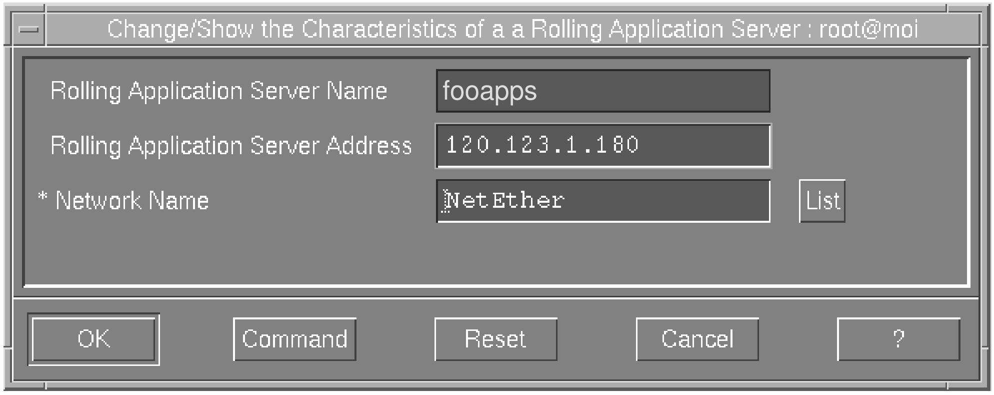
Note: Configuration information for Rolling Applications is automatically synchronized across all the cluster nodes.

Change/Show the Characteristics of a Rolling Application Server

This action allows you to change or show the characteristics of a Rolling Application Server.

From any node, bring up the **Bullcluster Rolling Applications** menu, and select the **Change/Show the Characteristics of a Rolling Application Server** menu. A list of Rolling Application Servers is displayed.

1. Select the Rolling Application Server that you want to show or change. The **Change/Show the Characteristics of a Rolling Application Server** screen appears.



Change/Show the Characteristics of a a Rolling Application Server : root@moi

Rolling Application Server Name: fooapps

Rolling Application Server Address: 120.123.1.180

* Network Name: NetEther

2. In this menu, the Rolling Application Server Name field cannot be edited and is displayed as a reminder only.

   If you want to change the Rolling Application Server Address, enter another address.

If you want to change the Network Name, select another Network Name from the list or enter the Name of a Network.

3. Once you have specified the field desired, validate the screen (click on **OK**) to apply your settings.

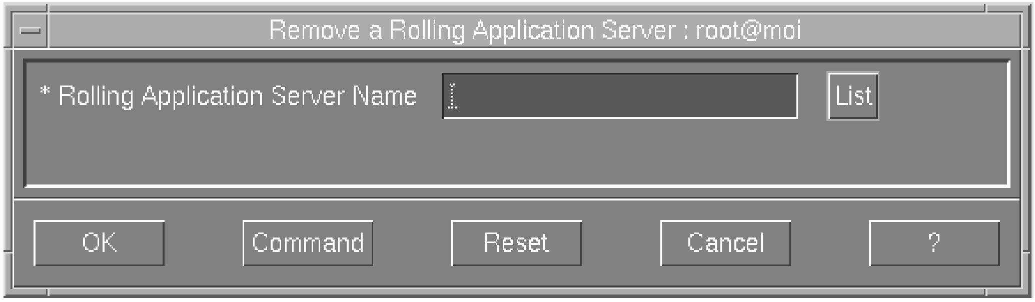
Note: Configuration information for Rolling Applications is automatically synchronized across all the cluster nodes.

Remove a Rolling Application Server

This action removes a Rolling Application Server from the configuration.

From any node, bring up the **Bullcluster Rolling Applications** menu, and select the **Remove a Rolling Application Server** menu.

The **Remove a Rolling Application Server** screen appears.



Remove a Rolling Application Server : root@moi

* Rolling Application Server Name

1. Select the Rolling Application Server Name to be removed from the list. The List button gives the list of all Rolling Application Servers.

2. Validate the screen (click on **OK**) to apply your settings.

Note: Configuration information for Rolling Applications is automatically synchronized across all the cluster nodes.

Bring a Rolling Application Server Online

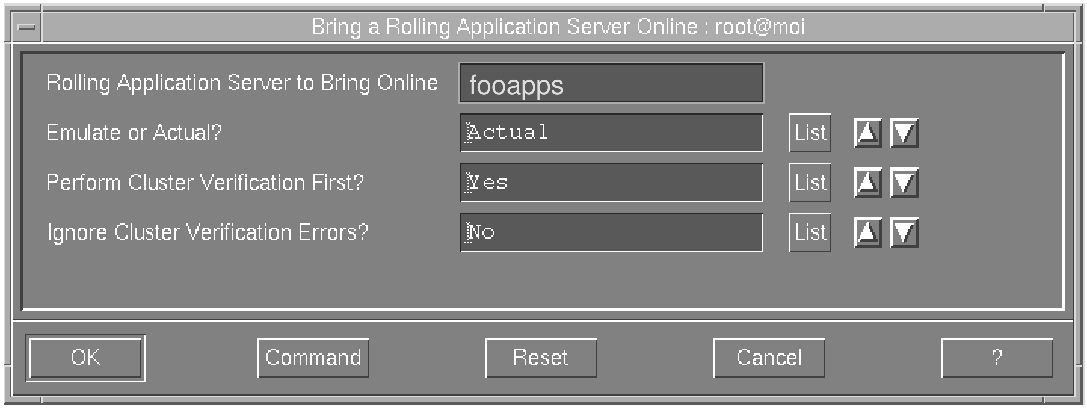
This action activates (starts) the associated Resource Group of the specified Rolling Application Server on the node identified as its highest priority node.

From any node, bring up the **Bullcluster Rolling Applications** menu, and select the **Bring a Rolling Application Server Online** menu.

A list of Rolling Application Servers is displayed.

1. Select from the list, the Rolling Application Server to bring Online.

   The **Bring a Rolling Application Server Online** screen appears.



Bring a Rolling Application Server Online : root@moi

Rolling Application Server to Bring Online fooapps

Emulate or Actual? Actual

Perform Cluster Verification First? Yes

Ignore Cluster Verification Errors? No

2. Choose **Actual** (the default) to bring the Rolling Application Server 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.

3. 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.

4. 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**.

5. Validate the screen to apply your settings.

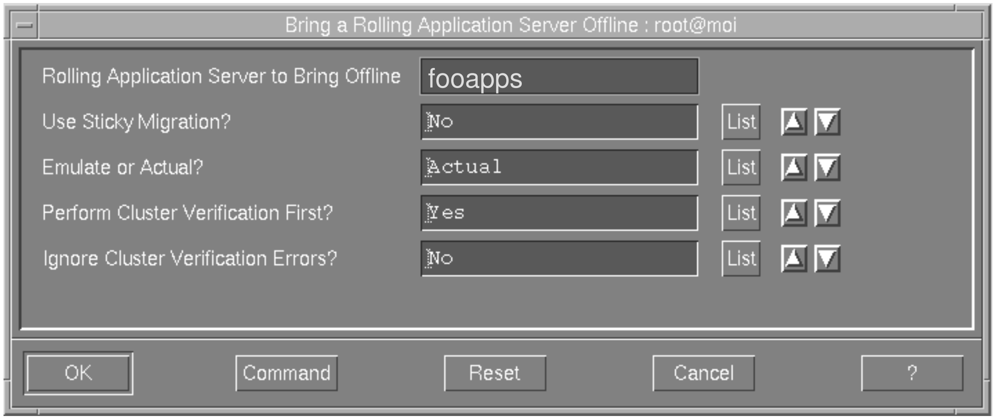
Bring a Rolling Application Server Offline

This action deactivates, or stops the associated Resource Group of the specified Rolling Application Server.

From any node, bring up the **Bullcluster Rolling Applications** menu, and select the **Bring a Rolling Application Server Offline** menu. A list of Rolling Application Servers is displayed.

1. Select from the list, the Rolling Application Server to bring Offline.

   The **Bring a Rolling Application Server Offline** screen appears.



Bring a Rolling Application Server Offline : root@moi

Rolling Application Server to Bring Offline: fooapps

Use Sticky Migration?: No

Emulate or Actual?: Actual

Perform Cluster Verification First?: Yes

Ignore Cluster Verification Errors?: No

2. Specify sticky or normal (nonsticky) migration.

   The default is **No**, causing a nonsticky migration. The associated Resource Group of the Rolling Application Server will be stopped, but its highest–priority location will remain unchanged, and the associated Resource Group may be restarted at the time of the next cluster event.

   If you choose **Yes**, the associated Resource Group of the Rolling Application Server will be stopped, and it attempts to remain offline during a cluster failover or reintegration.

3. Choose **Actual** (the default) to bring the Rolling Application Server Offline.

   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.

4. 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.

5. 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**.

6. Validate the screen to apply your settings.

Move a Rolling Application Server

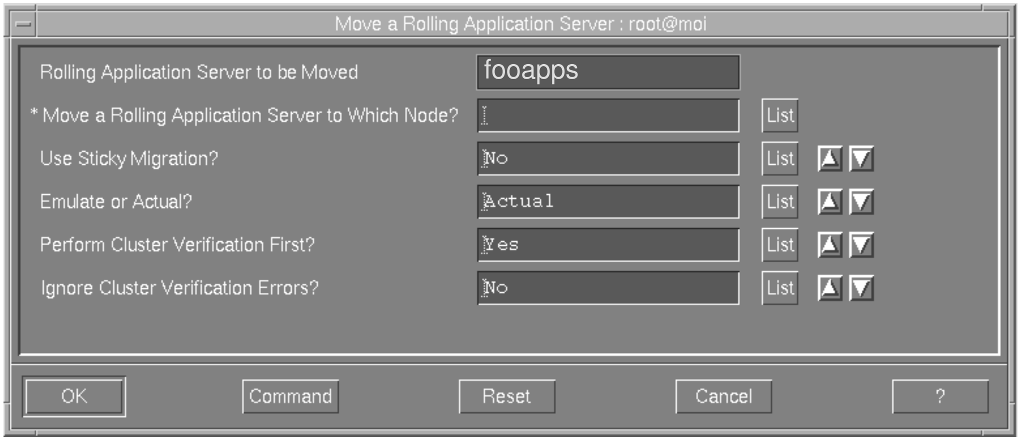
This action moves the associated Resource Group of the specified Rolling Application Server.

From any node, bring up the **Bullcluster Rolling Applications** menu, and select the **Move a Rolling Application Server** menu.

A list of Rolling Application Servers is displayed.

1. Select from the list, the Rolling Application Server to be moved.

   The **Move a Rolling Application Server** screen appears.



Move a Rolling Application Server : root@moi

Rolling Application Server to be Moved	fooapps
* Move a Rolling Application Server to Which Node?	
Use Sticky Migration?	No
Emulate or Actual?	Actual
Perform Cluster Verification First?	Yes
Ignore Cluster Verification Errors?	No

2. Select from the list, the node to which you are moving the Rolling Application Server, or enter the name of the target node.

3. Specify sticky or normal (nonsticky) migration.

   The default is **No**, causing a nonsticky migration. The highest–priority location will remain unchanged, and the associated Resource Group of the Rolling Application Server may be restarted at the time of the next cluster event.

   If you choose **Yes**, the specified target node becomes the highest–priority node for this associated Resource Group of the Rolling Application Server.

   The associated Resource Group of the Rolling Application Server will attempt to remain on that target node, regardless of that node's original priority, during a cluster failover or reintegration.

4. Choose **Actual** (the default) to move the Rolling Application Server.

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.

5. 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.

6. 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**.

7. Validate the screen to apply your settings.

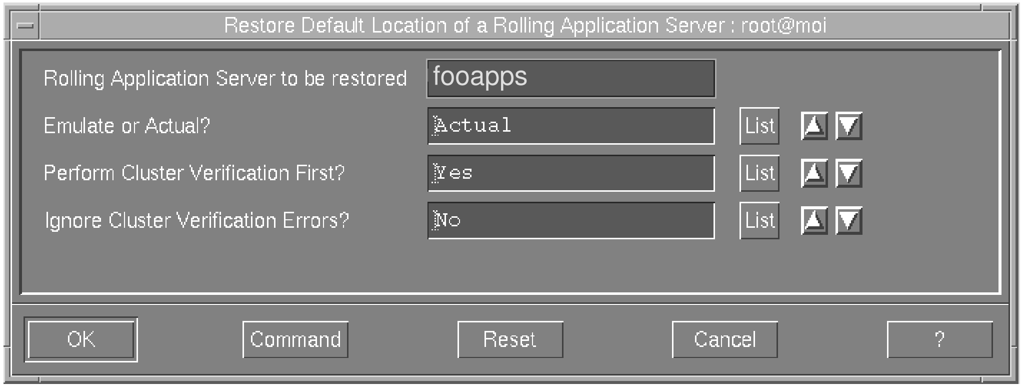
Restore Default Location of a Rolling Application Server

This action moves the associated Resource Group of the Rolling Application Server back to initial designated node, removing any stickiness assigned previously.

From any node, bring up the **Bullcluster Rolling Applications** menu, and select the **Restore Default Location of a Rolling Application Server** menu. A list of Rolling Application Servers is displayed.

1. Select from the list, the Rolling Application Server to be restored.

   The **Restore Default Location of a Rolling Application Server** screen appears**.**



Restore Default Location of a Rolling Application Server : root@moi

Rolling Application Server to be restored	fooapps	
Emulate or Actual?	Actual	List ▲ ▼
Perform Cluster Verification First?	Yes	List ▲ ▼
Ignore Cluster Verification Errors?	No	List ▲ ▼

2. Choose **Actual** (the default) to restore the default location of the specified Rolling Application Server.

   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.

3. 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.

4. 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.**

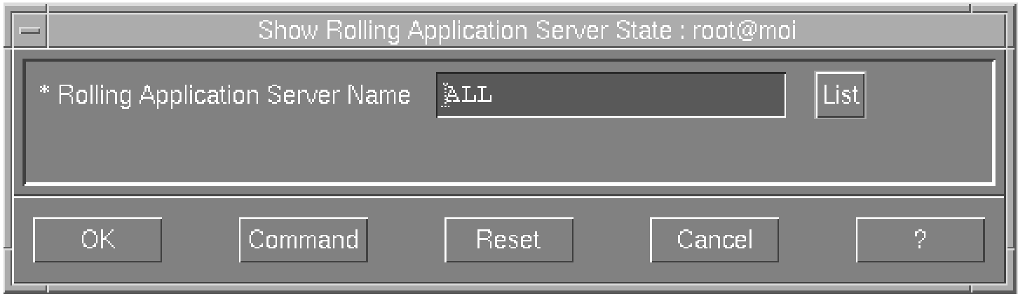
5. Validate the screen to apply your settings.

Show Rolling Application Server State

This action displays the state(s) of the Rolling Application Server(s).

From any node, bring up the **Bullcluster Rolling Applications** menu, and select the **Show Rolling Application Server State** menu.

The **Show Rolling Application Server State** screen appears.



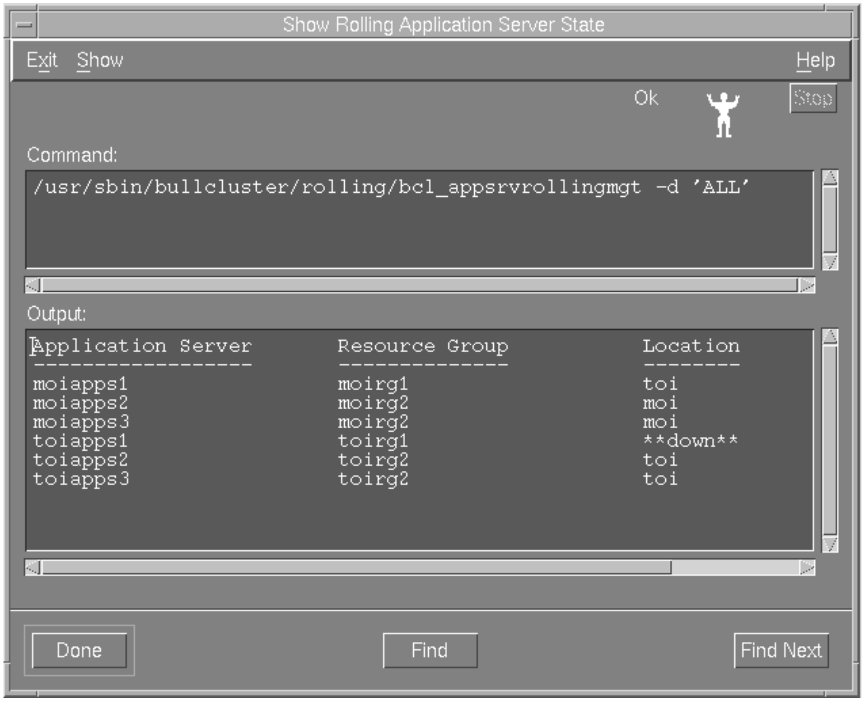
Show Rolling Application Server State : root@moi

* Rolling Application Server Name ALL

1. Select from the list, one Rolling Application Server, or **ALL** to display the state(s) of the selected Rolling Application Server(s).
2. Validate the screen to see the information about the state of the Rolling Application Server in accordance with your settings.

The concerned Rolling Application Server, is displayed along with the associated Resource Group and the current Location (or state: **down**, if no Location exists).

An example of the result of this menu is given below:



Show Rolling Application Server State

Command:

```
/usr/sbin/bullcluster/rolling/bcl_appsrvrollingmgt -d 'ALL'
```

Output:

```
Application Server       Resource Group          Location
------------------       --------------          --------
moiapps1                 moirg1                  toi
moiapps2                 moirg2                  moi
moiapps3                 moirg2                  moi
toiapps1                 toirg1                  **down**
toiapps2                 toirg2                  toi
toiapps3                 toirg2                  toi
```

Implementing Native HA (With or Without Shared Disk Expansion Cabinet)

Purpose

Native HA is an optional feature which must be activated if you have the HA Native Shared Storage configuration, described in **Planning "HA Native" Shared Storage for Escala HA Solutions**, on page 3-7.

When a failure occurs, in the standard HACMP environment for shared storage, automatic takeover is not possible, particularly in the case of a node failure.

Native HA overcomes this difficulty by:

- modifying the activation of shared Volume Groups.
- modifying the de-activation of shared Volume Groups.
- adding a pre-event for mirror synchronization (based on BullCluster Event Customization Management).

Configuration

Use the following check list to configure Native HA:

- On both nodes, all the internal or external shared disks must be AVAILABLE. (External means Expansion Cabinet).

  Each node must be able to see **all** the AVAILABLE internal or external shared disks. If all the shared disks are not counted then there is a problem in the cabling, check your installation with the cabling diagram. For details, see **SCSI Adapters and Bus Cabling**, on page 3-9.

From one node, create the shared VG (s) using internal shared disks and external shared disks for Native HA with the Disk Expansion Cabinet. Follow the instructions in the *Bull System Management Guide Operating System and Devices for AIX:*

- Each VG must contain at least two disks.
- Check the QUORUM and the auto–varyon of each VG: **BOTH** must be set to **NO**.
- Each Logical Volume of the shared VGs must be created entirely on one disk and must be mirrored (2 copies) on another disk belonging to the other node or Expansion Cabinet on a different SCSI bus.
- When creating a Logical Volume set the SCHEDULING POLICY for writing logical partitions copies to **SEQUENTIAL** (the default in smit is PARALLEL).
- Mirror the **jfslog** (default value) logical volume of your VG.

Import the information of the VGs from the other node:

- Be careful about the naming of disk-drives which may be different on this node.

  Take into account the Physical Volume Identifier (PVID) when preparing to import the correct Volume Group.

Note: To identify which hard disk to import, type:

```
lspv
```

- Check again the **QUORUM** and the **auto–varyon** on this node. After running the **importvg** command, set the QUORUM and the auto–varyon to the default value **YES**.

```
# chvg –a n –Qn <your VG>
```

It is important to record on the PLANNING sheet of HACMP the name of disks and their PV IDs on each node. Write also beside each disk the location code on each node. For each disk add a column to write down the name of the system where the disk is really located. In addition, also the following information about the use of the two SCSI–SE adapters by machine:

- PCI Slot of the adapter allowing to access the disks of the other node.
- PCI Slot of the adapter allowing to access the internal disks of this node.

This information will help in hardware problem determination and is useful in the RECOVERY sections described below.

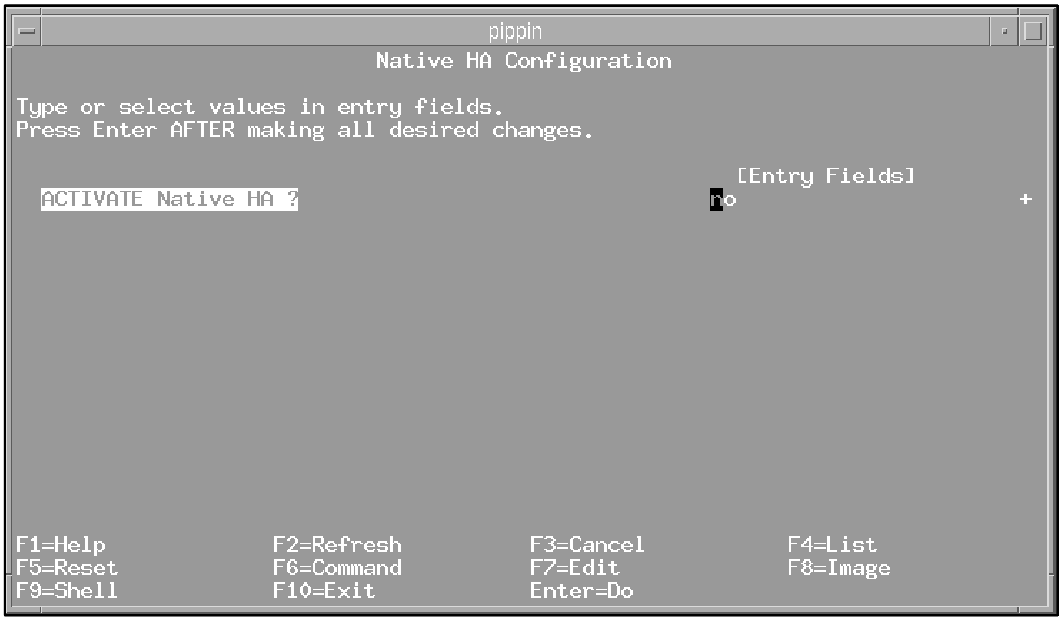
For HACMP cluster configuration, proceed as follows:

- AIX configuration for IP addresses, ttys, etc.
- create the topology : nodes, adapters.
- create the resource groups and their resources : VG, FS IP label, application servers, etc.

Implementation Procedure

To implement the Native HA feature (optional):

1. **If cluster services are running, stop cluster services on all the nodes.**
2. From the main HACMP SMIT menu, successively select **Bull Cluster Easy Configuration**, **Native HA Configuration.** The following screen is displayed:



```
pippin
                         Native HA Configuration

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                                  [Entry Fields]
  ACTIVATE Native HA ?                           no                       +

F1=Help          F2=Refresh        F3=Cancel         F4=List
F5=Reset         F6=Command        F7=Edit           F8=Image
F9=Shell         F10=Exit          Enter=Do
```

3. The current **ACTIVATE Native HA** field shows the current setting (it shows “`no`” if Native HA is not currently enabled, otherwise it shows “`yes`”). Toggle the value as desired, then press Enter to validate your choice.
4. As soon as you validate the screen, Native HA is enabled (or disabled) on the current node, as well as on the other cluster nodes.

Recovery From Hardware Problems

Disk Expansion Unit – Hardware Set Up

When using a Disk Expansion Unit with HA, all external devices which are shared in an HA configuration, and connected to the CPU drawer, must **NOT** be powered off during any hardware set up procedure.

Native HA Shared Disk Replacement

The procedure describes all the steps required to replace an internal or external shared hard disk in a configuration where Native HA is set up and HACMP installed. A disk failure is fully transparent to the applications since the data is still available on the mirror, but this section is concerned with the procedure to apply in order to re-introduce the new disk into the mirroring plan.

A volume group VG1 has 2 mirrored disks (each disk on a different node) where one, hdisk4 for example has to be replaced.

Note: In this example, disks have the same name on both systems. This may not be always the case: the same physical disk may have different names on 2 systems (e.g. hdisk4 on node A and hdisk5 on node B), for this reason check, on the different nodes, the shared disks using the PV IDs.

Initial Checking

1. Issue all the commands from the system where the volume group VG1 is active. No administrative commands should be executed on the other systems until the end of the procedure (except if specially mentioned).
2. Check which disk on which volume group has to be changed:

```
# errpt | grep hdisk or  errpt –a | pg
```

3. Check the cabling before going forward in this procedure. If one cable is unplugged go to the **Recovery from Faulty Cable Connection**, otherwise continue.
4. If no takeover happened, check if the adapters used in the SCSI bus on both nodes are OK. If on any node you have an errlog record about the adapter at the time the record of the disk is obtained, exit this procedure and go to the **Recovery From Adapter Failure** section.
5. Find the VG and check the logical volumes involved on the failed disk:

```
# lspv
hdisk4         000002634f03cc16    vg1
hdisk5         000002634e995408    vg1
...
```

```
# lsvg –M VG1 (or lsvg –p VG1)
hdisk2:1        sharedlv:1:1
hdisk2:2        sharedlv:2:1
hdisk2:3–50
hdisk4:1        sharedlv:1:2    stale
hdisk4:2        sharedlv:1:2    stale
hdisk4:3–50
```

The output ”stale” indicates that hdisk4 has a bad condition.

Note the name of the failed disk drive:

```
Disk drive:     hdisk4
```

6. Find the location code of the drive:

```
# lsdev –Cc disk
hdisk4 Available 04-07-00-8,0 Bull 2.1 GB 16 Bit SCSI Disk Drive.
hdisk5 Available 04-08-00-8,0 Bull 2.1 GB 16 Bit SCSI Disk Drive.
...
```

In this example, the location code is 04–07–00–8,0 which means PCI Adapter slot 7 and SCSI ID 8 as the disk is in the upper data area (disks in the upper shared disk cage are recognized with SCSI IDs: 8,9,a from top to bottom, and in the lower shared disk cage as: 0,1,2)

Recovery Procedure

1. Identify all the logical volumes present on hdisk4:

```
# lspv –l hdisk4
List of logical volumes:
```

2. Reduce the number of copies of all the logical volumes mirrored on this hdisk4 disk, specifying that the copy to be removed is on the hdisk4 failed disk:

```
# rmlvcopy sharedlv 1 hdisk4
```

etc. (for all logical volumes in the volume group)

Do not forget to also remove the copy to the JFSLOG LV, if there is one on this disk.

3. Remove the disk from the volume group:

```
# reducevg VG1 hdisk4
```

You can do this even if your File Systems are mounted and your applications are still running.

4. ON EACH NODE, remove the failed disk from the configuration database:

```
# rmdev –l hdisk4 –d
```

Note: Remember that the name of the disk may not be the same on all nodes. Use the **lspv** command to find the PV ID of the disk on each node.

```
# lspv
```

5. NOW, remove the disk from the cage of the system you had identified previously, and insert the new one.

Note: On some Escala servers the disks are hot-removeable.

6. On each node, configure the new disk:

```
# cfgmgr
```

7. Find the new disk name:

```
# lsdev –Cc disk
hdisk4     Available    Location     Comments
# lspv
hdisk4      PVID
```

Update this information in your PLANNING Sheet for the shared VGs and disks.

In the following steps, it is assumed that the new drive has been assigned the same logical name (hdisk4).

If no PVID is assigned to the disk

on one node, type:

```
chdev –l hdisk4 –a pv=yes
```

on the other node, type:

```
rmdev –l hdisk4
mkdev –l hdisk4
```

8. Add the disk to the volume group:

```
# extendvg VG1 hdisk4
```

Update the PVID of this new disk in your PLANNING Sheet for the shared VGs and disks if different.

9. Add (and synchronize) a mirror copy to all the logical volumes on the disk, do not forget to mirror the JFSLOG:

   ```
   # mklvcopy -k sharedlv 2 hdisk4
   ```
   etc.

On the other node, the new disk will be taken into account on the next fallover.

Recovery From Faulty Cable Connection

You arrive here after detecting an error record on a disk and there was no subsequent takeover. It is noticed that one of the Y cables is unplugged or incorrectly plugged.

Initial Checking

Make sure that the cable is secured correctly and that the internal SCSI flat cables and wrap plugs are correctly plugged-in.

Recovery Procedure

1. After having checked the SCSI cable connections, try to access your disk in read mode on each system (with root privileges).

   ```
   # dd if=/dev/rhdisk4 of=/dev/null bs=128k count=1
   ```

   If an I/O error is seen and a new record is added to errlog, this means that either your adapter or your disk is out of order and must be replaced. So, quit this procedure and follow the usual procedure for replacing the adapter on the node where the command is run. If this disk is an internal disk, go to section NATIVE HA recovery from adapter failure.

2. On the node where the VG1 is active and where the dd command succeeded, run the command `varyonvg` again without stopping the applications:

   ```
   # varyonvg VG1
   ```

   This command will synchronize the staled copies.

Recovery After Node Failure

The node where the VG1 was active crashed (888) or was powered off and the VG1 is now active on the other node (after takeover of HACMP).

Initial Checking

On the surviving node, the errlog contains entries about the SCSI bus failure and about the internal disks of the other system that are no longer accessible.

In the */tmp/hacmp.out* or */var/adm/cluster.log* there is an event node_down_remote because the other node failed.

Recovery Procedure

The applications are automatically restarted on the surviving node despite the fact that the disks of the failed system are now missing in the volume group VG1.

These inaccessible disks may have the name of the VG changed to NONE when the VG1 is taken over on the surviving node.

```
# lspv VG1

hdisk4       PVID        NONE
hdisk5       PVID        VG1
```

Here hdisk4 is an internal disk of the down machine.

This situation is normal and does not prevent proper operation of the application, since VG1 is still accessible via hdisk5.

Procedure After Repair

Check that the machine is up again and all the disks are here.

Start the cluster manager on the repaired machine:

```
# smit clstart
```

This procedure is done automatically; no human intervention is needed.

Note: After the start of HACMP on the failed node, the status NONE for hdisk4 on the surviving node is no more visible. In fact, the new *cl_deactivate_vgs* script (called to release the VG1) delivered with this version, de-activates, exports and re-imports the information of VG1 before releasing it. This is possible as the applications are already stopped and the shared File Systems of VG1 are already unmounted. On this surviving node, all the accessible disks in VG1 no longer have the NONE status.

On the surviving node, if another VG for example VG2 is active before the failure of the other node, you no longer need to synchronize the staled mirror by hand. The customized pre-event script added to the event *node_up_complete* is in charge of synchronizing any staled copies on it, when all the disks are up again.

When the repaired machine had finished joining the cluster, check that all the mirrors of each VG, active on each node, are synchronized and no shared disk has the NONE status:

```
# lsvg -l VG1  (VG2)
# lspv
```

Recovery From Adapter Failure

If the node where the VG1 is active has a bad adapter to be changed then follow the procedure CASE 1.

If the node where the adapter is bad, has no VG active on it then follow the procedure CASE 2.

CASE 1 Recovery Procedure

The applications must be stopped on this node and be restarted on the other node, as one cannot change an adapter without powering-off the system.

1. Disconnect the cable in the SCSI chain at the terminated end.

   Note: Ensure that the disconnected cable remains terminated with the terminator.

2. Stop the HACMP in takeover mode.
3. Wait until the application has restarted and logical volumes are synchronized on the backup node.
4. Power-off the machine.
5. Replace the failed adapter with a good one.

Procedure After Repair

Follow the same procedure as in previous paragraph, **Recovery After Node Failure**.

CASE 2 Recovery Procedure

No application is running on this node (NO resource group is active):

1. Disconnect the cable in the SCSI chain at the terminated end.

   Note: Ensure that the disconnected cable remains terminated with the terminator.

2. Power-off the machine
3. Change the adapter
4. Power-up the machine.

Procedure After Repair

On this system, start **hacmp**:

```
   # smit clstart
```

On the other node, the customized pre-event script added to the event *node_up_complete* will synchronize the staled copies on the surviving machine.

Implementing the Extended HA Shared Storage

Precautions

Each node in a cluster defined over two sites, must belong to the same subnet to respect HA requirements for public network connections (IP takeover capability).

Serial networks used for HACMP heartbeat between cluster nodes must be extended using RS232 micro-modems.

Network configuration aspects must be examined on the customer site by the Bull Service Representative before any installation.

Configuration

Use the following check list to configure Extended HA:

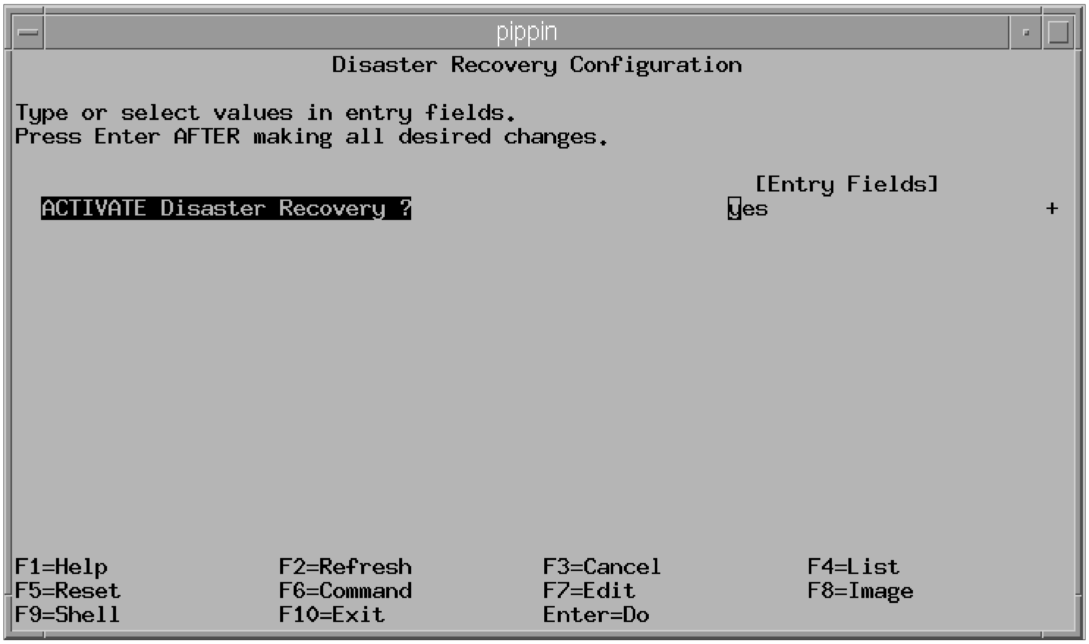
- On both nodes array/atf and al_slot objects must be available. See AIX Emulex adapter/driver documentation.
- On both nodes all the disks must be available.
- Create the shared VGs as usual, using luns from the first DAS. For each LV in a VG add a second copy on a disk (lun) from the second DAS system.
- Choose the two luns (at least two) with the same RAID type (RAID1, RAID3 or RAID5 for High Availability) and with the same disk capacity.
- Choose the copies on different FC–AL loops. For example, if you choose disk1 (lun1) on DAS1 binded on SPA then choose disk2 (lun2) on DAS2 binded on SPB.

Follow the same instructions given in the HA Native configuration.

Implementation Procedure

To implement the Extended HA environment:

1. **if cluster services are running, stop cluster services on all the nodes.**
2. From the main HACMP SMIT menu, successively select **Bull Cluster Easy Configuration, Disaster Recovery Configuration.**



```
                    Disaster Recovery Configuration

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                                         [Entry Fields]
  ACTIVATE Disaster Recovery ?                          yes                +











F1=Help             F2=Refresh          F3=Cancel           F4=List
F5=Reset            F6=Command          F7=Edit             F8=Image
F9=Shell            F10=Exit            Enter=Do
```

3. The current **ACTIVATE Disaster Recovery** field shows the current setting (it shows "no" if Disaster Recovery is not currently enabled, otherwise it shows yes"). Toggle the value as desired, then press Enter to validate your choice.

4. As soon as you validate the screen, Disaster Recovery (Extended HA) is enabled (or disabled) on the current node, as well as on the other cluster node.

Note: You can get to this last menu by calling also smit with the fast path disaster.

Recovery From Hardware Problems

When a hardware failure (node, adapter, SP, general power supply, hub, DAS system) is detected on a Site by ATF or/and the HACMP cluster manager, there is a failover mechanism on the redundant component (adapter, hub, node,AIX mirrors) which allows the continuity of the service (applications); this is the HACMP environment expected behavior.

Depending of the failure type, follow the general/specific instructions described below in order to repair and to return to the initial state before the failure.

Recovery from Faulty Node

HACMP on the surviving node will take over all the resources on the failed node.

After repair of the failed node, power on this node and start HACMP. The resources will be re–activated back on this node.

Recovery from Adapter Failure

When an adapter failure is detected on a node, ATF (Application Transparence Failure) will switch the control of any lun, binded to the SP connected to this adapter through a HUB, accessed at this time by the applications.

Repair Procedure

Procedure as follows:

- Stop HACMP with takeover on this node, power off the node, change the adapter.
- On the other node, restore ATF (atf0 or atf1) if any trespass has occurred. Select smit menu atf and then select **restore LUNs to originated SPs**.
- Power on the first node, check that both FC Emulex adapters have their green light on and that the yellow light is blinking at the end of boot time.
- Start HACMP back on this node.

Recovery after SP Failure

ATF changes the control of accessed luns to the other SP and the applications continue working on both nodes.

Repair Procedure

Procedure as follows:

- Remove the failed SP and change it without stopping the applications.
- After SP booting check the status of the SP: the LED on the HUB port connected to this SP must be off.
- Restore ATF on each node (atf0 or atf1) if atfx (x: 0 or 1) had trespassed the luns on this SP.

Recovery after Site General AC Power Problems

The surviving site (Site 1) will take over all the resources activated on the site having AC power problems (Site 2).

When the AC power problem is repaired, do the following repair procedure.

One must **plan** this procedure as in some cases it may be required to stop all the applications in the cluster.

Repair Procedure

Procedure as follows:

1. Power on the HUBs on this site (Site 2).
2. Power on the DAS system on this site (Site 2); wait until the service light goes on and then off.

   If you find a SP status off–line you have to power off its HUB and then to power it on, to re-initialize the insertion of this SP in the FC–AL loop.
3. Power on the node on this site (Site 2).
4. Check ATF status on both sites as ATF may have trespassed.

   If atfx has trespassed then restore trespassed Luns to their default SPs.
5. Start HACMP on the node of Site 2; all the resources will be taken back on this Site.

   Atfx may trespass on both Sites when the HACMP manager is started on the repaired site; if this is the case then:

   – wait until the event node_up_complete on this Site (Site 2);

   – restore the luns for each atfx on this Site (Site 2);

   – on the other Site (Site 1), restore the luns again to their SPs.

Recovery From DAS Disks Failure

Contact your Bull Service Representative for the procedure of hardware exchange and AIX procedures to make all the VGs available again with all the mirrored disks (new ones).

Customizing Cluster Events Processing

Purpose

In HACMP, any cluster event can be associated with processing scripts to be executed when the associated event occurs (for background information about concepts and HACMP features related to cluster events, refer to the HACMP documentation).

Synchronizing Time in Your Cluster with timed

Timing is very important for many applications that make decisions based on the current time or logged timestamps. Database software, as well as analysis, accounting and auditing tools, are examples of such applications. Incoherent time clocks across the nodes are prone to mislead applications.

Thus, it is strongly recommended to implement the **timed** daemon in your cluster so that clocks are maintained synchronized on all the nodes.

timed in Brief

timed is a standard time server that comes with AIX. It synchronizes one machine's clock with those of other machines on the network that are also running **timed**. By slowing down the clocks of some machines and speeding up the clocks on other machines, **timed** maintains a uniform network time.

timed is a daemon that can run in two modes: master or slave. Through this mechanism, **timed** ensures on its own high availability of time service.

- When the **timed** daemon is started with the `-M` flag (master mode):
  - If the network has a master **timed** already running, the **timed** just started becomes a so called *submaster*. Submaster time servers are here only to back up the master **timed**, should it fail.
  - If the network does not have a master **timed** already running, the **timed** just started becomes the master time server for this network. The master **timed** is responsible for performing transactions with the other **timed** servers in order to synchronize the time. If the master **timed** dies (for example, due to a failure of the host on which it is running), a new master time server is elected from the submaster time servers (i.e. from other **timed** servers that have been enabled with the `-M` flag).
- When the **timed** daemon is started without the `-M` flag, it runs as a slave server (never candidate to become a master time server).

Hints for Implementing timed in Your Cluster

Here are two hints for implementing time service in your cluster:

- Enable **timed** on each node. Consider that at least two **timed** servers (each on a different node) must be enabled in master mode (`-M` flag).
- The most straightforward solution consists of enabling **timed** as a master server (`-M` flag) on all the cluster nodes. With this strategy, you are sure that time service continues even in the situation where only one node is up.

For details on **timed** and on how to enable it automatically at system startup, refer to the AIX documentation.

Note for PowerConsole Users

If you are setting up a Powercluster equipped with a PowerConsole, we recommend you set up timed using the configuration utility that is included in the ClusterAssistant application set. Note the following:

- To access the utility, activate the **AIX Tools** icon (which is available from the "Configuration" folder in the ClusterAssistant application group), then use the "timed Configuration" feature.
- When using this feature, **timed** is automatically set up in the following way:
  - it is set up to run as a master timed on the PowerConsole,
  - it is set up to run as a slave timed on the different nodes of the cluster.

  Note that this setup, although it is not the same as the sample setup given in the hints above, is the recommended one when implementing a cluster equipped with a PowerConsole.

Hints for Dealing with iFOR/LS Nodelocked License Keys

iFOR/LS, through the use of encrypted keys, monitors the software licenses used by stand-alone machines, or by machines within a network.

If you experience problems when dealing with license keys for your specific software products, refer to the iFOR/LS documentation. Note that an appendix of the *AIX iFOR/LS Tips and Techniques* book discusses special considerations on using iFOR/LS within an HACMP environment.

Hints for Using DSMIT

DSMIT (Distributed System Management Interface Tool) is an optional product from the AIX product family. DSMIT is similar to the standard SMIT tool, excepted it is able to execute management commands over several networked systems at the same time.

Please note the following:

- **DSMIT must not be used to run most of the different HACMP SMIT menus.** DSMIT can only be used to start and stop HACMP cluster services. Almost all other HACMP SMIT operations will cause unpredictable results if performed via DSMIT.
- On the other hand, you can take advantage of DSMIT for other system management tasks that do not involve the HACMP menus. (For example, DSMIT is a valuable tool for managing user accounts and installing software packages on multiple nodes).

Here are some additional hints for using DSMIT in your cluster:

- Nodes that belong to the same cluster must be gathered in a “working collective” or in a “domain”.
- The simplest DSMIT implementation consists of configuring a single cluster node as a DSMIT server. However, you may want to make your DSMIT highly available, by configuring several cluster nodes as DSMIT servers. You may even configure all cluster nodes as DSMIT servers, so that you do not have to worry about which node you must use to execute DSMIT commands.

For further information, refer to the DSMIT documentation.

Managing DAS Subsystems with ArrayGUIde

Note: From AIX 4.3.3, you must use Navisphere to manage all DAS subsystems.

The DAS (2900, 3x00 and 3500) disk-array subsystems come with the *ArrayGUIde* and the **dassmgr** utilities. To manage your DAS subsystems, you can use either ArrayGUIde or **dassmgr**.

- ArrayGUIde is a graphical user interface that lets you manage, configure and monitor multiple DAS subsystems connected to multiple servers on the network. You can run ArrayGUIde from any system on the network, not necessarily from a node to which the DAS is connected.
- **dassmgr** communicates with the DAS through a serial connection from a storage processor to an RS-232 port on a node.

You may find that ArrayGUIde is more friendly than **dassmgr** for daily monitoring tasks. For details, refer to the *Configuring and Managing a DAS and Using ATF* and *Using the ArrayGUIde Utility* books.

Note for PowerConsole Users

If you are equipped with a PowerConsole, you can take advantage of the ClusterAssistant application, which includes utilities to easily set up and run ArrayGUIde.

Managing DAS Subsystems with Navisphere

Note: From AIX 4.3.3, all DAS subsystems are managed by Navisphere.

DAS storage subsystems are managed using the Navisphere Manager, an interactive graphical interface on the Powerconsole.

Navisphere Manager provides hierarchical status information about the physicaldisks making up the storage system.

It also provides configuration and management functions to set up storage system memory, find disk modules into logical units and change parameters of the storage system.

Chapter 11. Remote Monitoring a Cluster With ISM

This chapter explains how to remote monitor an **HACMP 4.3** cluster using the **ISM 4.54** product. It includes the following sections:

- Concepts.
- Setting Up the SNMP Environment on the Cluster Nodes, on page 11-3.
- Installing the Needed Files on the ISM Manager Station, on page 11-4.
- Drawing a Basic ISM Picture Representing a Cluster, on page 11-5.
- Hints For Interpreting Some Basic Information, on page 11-9.
- Browsing the HACMP MIB and Adding Objects to the Picture, on page 11-11.
- Understanding the HACMP MIB Objects, on page 11-13.
- Dealing With Traps, on page 11-16.

Note: It must be stressed that ISM is intended to be installed and run on a dedicated AIX workstation of the network, not on a cluster node. A workstation dedicated for running ISM is referred to as an “ISM Manager Station”.

Concepts

What is ISM?

ISM (Integrated System Management) is a Bull product designed to remotely monitor and globally manage distributed systems and networks throughout an Enterprise.

ISM offers a user-friendly and configurable graphical interface. In addition, ISM employs a sophisticated architecture in which all networked resources are defined as objects, making it easy to integrate different components into a single, consistent view.

The figure on next page shows a sample ISM display.

ISM and Cluster Monitoring

A cluster can change over time. For example, a node can join or leave the cluster, or a standby adapter can replace a service adapter. Such changes generate events reported to ISM enabling you to monitor them.

With ISM, you can draw a picture, i.e. a graphical representation of a cluster with its nodes and their network interfaces. After proper configuration, you can display this picture to visually monitor the cluster.

When events occur in the cluster, ISM updates the displayed picture to reflect the changes in the cluster. For example, some green labels may become red to denote certain events. You can then browse the HACMP MIB (*MIB* stands for *Management Information Base*) or display the received traps to obtain further information on the cluster state.

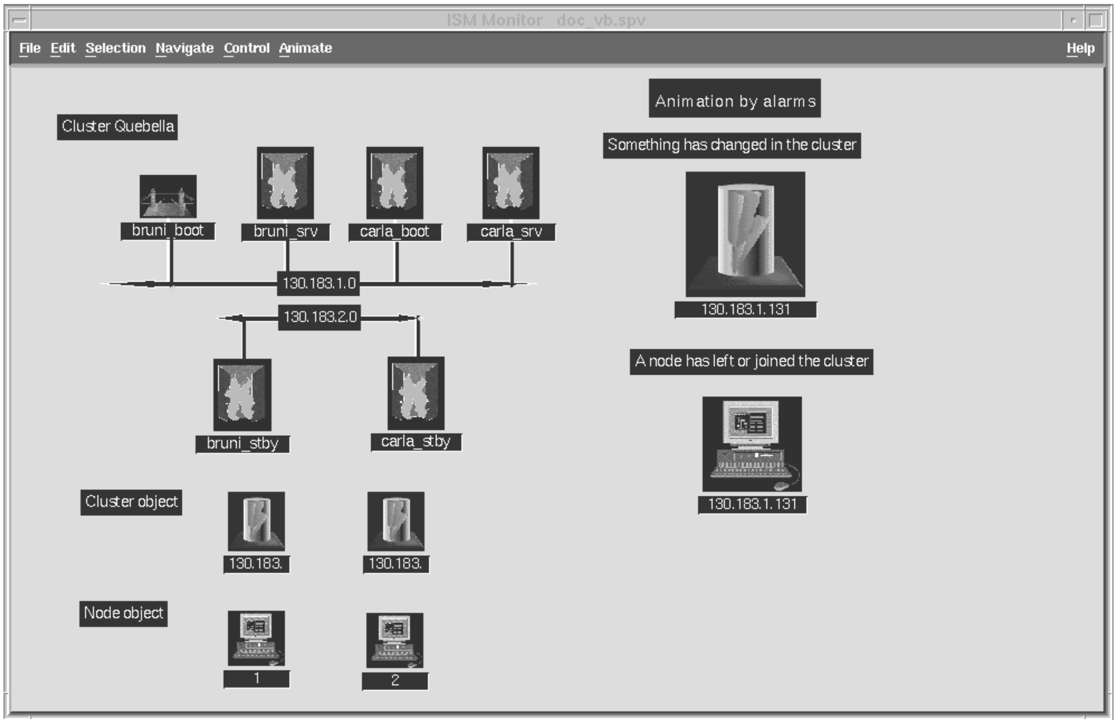
The HACMP for AIX software includes an enterprise-specific MIB dedicated to cluster environments. HACMP provides notification of cluster state changes through the Cluster SMUX peer daemon, **clsmuxpd**, which should run on each cluster node. To monitor a cluster, ISM relies on the HACMP MIB.

Related Information

- The rest of this chapter is for administrators who are responsible for setting up ISM, and who are familiar with ISM. The given recommendations should be regarded as hints only.

For details on ISM configuration, customization and operation, refer to the ISM documentation.

- For further information on concepts and utilities related to cluster monitoring, refer to the HACMP documentation. You may want to look at the “SNMP”, “clsmuxpd” and “clinfo” entries in the *HACMP Master Index and Glossary*. (**clinfo**, which comes with the HACMP for AIX software, is an SNMP-based cluster monitor.)



Sample ISM picture

Required ISM Applications

Three standard ISM application modules are used to set up the cluster monitoring environment: ISM Monitor, ISM IP Discovery, and ISM Alarm.

ISM Monitor

This is the main facility provided with ISM for monitoring managed objects. It allows you to:

- Build and display pictures, i.e. graphical representations showing the managed objects.
- Animate the pictures to reflect state changes of the managed objects. These changes can be shown through color changes in the animated icons.
- Visualize (“browse”) objects and attributes maintained in the different involved MIBs.

ISM IP Discovery

This ISM facility allows you to scan the networks to detect the presence of networks and devices that support the IP protocol. Once the objects are discovered, you can include them in a picture, using ISM Monitor.

ISM Alarm

This ISM facility allows you to monitor alarms, i.e. traps emitted by the SNMP daemons (**snmpd** and **clsmuxpd**). Within an ISM picture, you can animate the icons according to these traps. When alarms (i.e. traps) arrive, you can examine their detailed contents.

Setup Procedure Summary

If you want to monitor your cluster with ISM, you must set up your environment. The setup procedure consists of the following tasks:

- Setting Up the SNMP Environment on the Cluster Nodes
- Installing the Needed Utilities on the ISM Manager Station
- Creating ISM Pictures

This chapter explains these tasks.

Setting Up the SNMP Environment on the Cluster Nodes

This task must be carried out for each cluster node. It consists essentially in performing some checks.

Setting Up the snmpd Configuration Files

During the installation of HACMP, two configuration files, **/etc/snmpd.conf** and **/etc/snmpd.peers**, are updated with appropriate identification strings and passwords. On each cluster node, check and update these files as explained below.

1. Edit the **/etc/snmpd.conf** file.

   Check that the line below has been added to the end of the file (add it if necessary):

   ```
   smux   1.3.6.1.4.1.2.3.1.2.1.5   clsmuxpd_password  # HACMP
   ```

   As needed, add trap lines to specify which machines should receive traps generated by **clsmuxpd**. You must add one line for each ISM manager station you plan to use for cluster monitoring. In the example below, two trap lines were added corresponding to two ISM stations, **ism_foo** and **ism_bar**, whose IP addresses are 130.183.1.1 and 130.183.1.51. Both **ism_foo** and **ism_bar** stations will receive traps.

   ```
   trap    public  130.183.1.1            #ism_foo
   trap    public  130.183.1.51           #ism_bar
   ```

2. Edit the **/etc/snmpd.peers** file, and check that the line below has been added to the file (add it if necessary):

   ```
   clsmuxpd   1.3.6.1.4.1.2.3.1.2.1.5  ”clsmuxpd_password” # HACMP
   ```

3. Stop and restart the **snmpd** and **clsmuxpd** processes on the nodes where you made the changes:

   ```
   stopsrc –s clsmuxpd
   stopsrc –s snmpd
   startsrc –s snmpd
   startsrc –s clsmuxpd
   ```

Updating the /etc/hosts and /etc/inittab Files

/etc/hosts

For the **clsmuxpd** daemon to correctly register with **snmpd**, you must include `loopback` as an alias to the `127.0.0.1` entry in the **/etc/hosts**.

On each cluster node, edit the **/etc/hosts** file and add the following line:

```
127.0.0.1      loopback
```

Note: The **/etc/hosts** file usually contains a “`127.0.0.1 localhost`” line. If present, do not remove it. (In any case, add the “`127.0.0.1 loopback`” line as explained above.)

/etc/inittab

You should update **/etc/inittab** (on all cluster nodes) so that the **clsmuxpd** and **snmpd** daemons automatically start at node startup time.

Checking that the MIB Definition Files are Installed

The HACMP MIB is defined in the source file **/usr/sbin/cluster/hacmp.my**. Its compiled version is **/usr/sbin/cluster/hacmp.defs**.

On each cluster node, check that these two files are actually installed in the **/usr/sbin/cluster** directory.

Checking SNMP Operation

To check SNMP operation, request locally and from an ISM station the **clsmuxpd** daemon using the **snmpinfo** command. For example:

1. From a cluster node, enter the following command to request locally the dump of the `clsmuxpd` group of the HACMP MIB tree:

   ```
   snmpinfo -o /usr/sbin/cluster/hacmp.defs -m dump -v clsmuxpd
   ```

   The above command should display the dump of the `clsmuxpd` group of the HACMP MIB tree. It resembles the following:

   ```
   clsmuxpdGets.0 = 0
   clsmuxpdGetNexts.0 = 2
   clsmuxpdSets.0 = 0
   nodeId.1 = 1
   clsmuxpdTraps.0 = 15
   ...
   ```

2. From the ISM station, enter the following command:

   ```
   snmpinfo -h <NodeName> -o /usr/sbin/cluster/hacmp.defs
             -m dump -v clsmuxpd
   ```

   This command should also display the dump of the `clsmuxpd` group.

Notes:

- If nothing is displayed, check your installation as indicated in this section.
- Repeat this verification step for each cluster node.

Installing the Needed Files on the ISM Manager Station

Installing the Fileset Dedicated to ISM

The BullCluster software package includes specific filesets dedicated to ISM. For their installation, refer to the *Software Release Bulletin* (SRB) that comes with the software. This bulletin includes installation instructions and indicates possible prerequisites.

Once you have installed the appropriate fileset(s), install the HACMP MIB within ISM as explained below.

Installation Directory

On an ISM 4.54 station, the files are installed in the **$ISMROOT/var/mibspack/snmp/enterprise/ibm/hacmp** directory.

Provided Files

The fileset dedicated to ISM provides notably the following:

- The **hacmp.mib** file, which is a compiled version of the HACMP MIB, ready to use with ISM.

- The **hacmp.ext/default.mnc** files, which define for each object the associated icon and its animation mode.
- The **hacmp.icons** directory, which include some icon files.

For a complete list, you can examine the contents of the installation directories.

Installing the HACMP MIB within ISM

Warning: The HACMP MIB customization files are located in the **hacmp.ext/default.mnc** directory. When installing the HACMP MIB, the HACMP customization will automatically be added to the default configuration files of the ISM_Monitor application, located in the **$ISMROOT/var/config/ISM_Monitor/default.mnc** directory.

It is recommended to save your default ISM_Monitor configuration before installing the HACMP MIB, using the ISM_Configurator/ISM_Monitor application.

We suppose here you have started ISM on the ISM manager station. To install the HACMP MIB within ISM (repeat this procedure on each ISM manager station you will use to monitor the cluster):

1. From the ISM Application Board, execute the ISM Configurator.
2. Once the configurator is displayed, bring up the **MIB Schema** menu and select **LAUNCH**.
3. In the MIB window that appears, click on **Install/Remove private MIB**, then enter the name of the HACMP MIB:

   ```
   hacmp
   ```

   Follow the messages that are displayed (if you need details, refer to your ISM documentation).
4. Once the MIB installation is complete, stop ISM by selecting the **Exit** option from the **File** menu of the Application Board, then stop the framework by entering the **fmkstop** command.
5. Restart ISM and the framework.

Once these preliminary setup steps are performed, you can draw ISM pictures to represent and monitor the cluster.

Drawing a Basic ISM Picture Representing a Cluster

This section gives some hints that may help you build an ISM picture in order to remote monitor a cluster. These are hints only, since ISM is highly customizable and your needs may vary.

Building an ISM picture basically consists of the following steps:

- Discovering the Networks and Devices Related to the Cluster
- Drawing a Picture Including the Discovered Devices
- Animating the Icon Display

After having drawn a basic picture, you can refine it by adding details and defining specific animation criteria. These tasks are described in subsequent sections, beginning with “Browsing the HACMP MIB and Adding Objects to the Picture”, on page 11-11.

Note: You may want to follow these steps as an exercise, to learn about the particularities involved when integrating a cluster into ISM. Then, it will be up to you to determine your needs and to design appropriate pictures accordingly.

Discovering the Networks and Devices Related to the Cluster

This step consists in making the cluster nodes, adapters and networks, known to ISM, through the use of ISM IP Discovery.

Preliminary Remarks

How ISM IP Discovery Works

ISM IP Discovery identifies systems through their IP address. To each IP address found, corresponds one object and one icon which represents it. For each network or system discovered, ISM IP Discovery creates an object (IP network or IP device) in the ISM CMIS database.

Cluster Specifics

Cluster nodes have at least two network adapters: a service adapter and a standby adapter, whose addresses correspond to two distinct subnetworks. In addition, if the cluster uses IP address takeover (this is the usual case), the service adapter is associated to two IP addresses: the adapter boot address (used at boot time and until cluster services are started) and the adapter service address (used once the cluster services are started on the node).

IP Primary and Secondary Addresses

To each IP address it discovers, ISM IP Discovery associates one icon (whose title is the IP address). Thus, a service adapter is seen with 2 icons, one when discovered through its service address, the other when discovered through its boot address.

When the IP discovery process is performed through the network to which the service adapters are connected: the service adapter (service or boot address) is seen as the **IP primary address**; the standby adapter is seen as the **IP secondary address**.

Procedure

1. On each cluster node you want to “discover”, check that the **snmpd** daemon is enabled. If necessary, start it using the `startsrc -s snmpd` command.
2. From your ISM station, start up the ISM IP Discovery application. Choose the **Network** option. IP Discovery displays the networks it has discovered.

Discover the Service Adapters of the Cluster Nodes

3. From the network list displayed by ISM IP Discovery, select the network that is used by the service/boot adapter of the nodes. Run the discovery process, by selecting the **On Selected Networks** from the **Devices** menu.

   When cluster services are running on a node, the node uses its service address (instead of its boot address). Thus, IP Discovery will discover the service addresses of the nodes.

   Once the discovery process is complete (this can take several minutes), the discovered IP devices are displayed. Among them, you should see those that correspond to the service adapter labels of the cluster nodes.

Discover the Boot Adapters of the Cluster Nodes

4. If you can stop HACMP on the nodes, stop it. The nodes will then use their boot address, which can be discovered by IP discovery. This step is optional, as the HACMP objects of the **clsmuxpd** daemon will be accessed using the service IP addresses discovered in the previous step, but it allows to draw the boot adapters in the monitoring map.

Discover the Standby Adapters of the Cluster Nodes

5. (Note: This step may be accomplished whether or not cluster services are running on the cluster nodes). From ISM IP Discovery, select the subnetwork that is used by the standby adapters of the nodes (this is not the same subnetwork as the previous one, since service/boot adapters and standby adapters use distinct subnetworks). Run the discovery process, by selecting the **On Selected Networks** from the **Devices** menu.

   Once the discovery process is complete (this can take several minutes), the discovered IP devices are displayed. Among them, you should see those that correspond to the standby adapter labels of the cluster nodes.

Once these steps are performed, you can begin to draw a picture that includes the discovered devices, as explained below.

Drawing a Picture Including the Discovered Devices

This step consists in drawing pictures including the discovered networks and cluster devices, through the use of ISM Monitor. These pictures will be used to remote monitor the cluster behavior graphically.

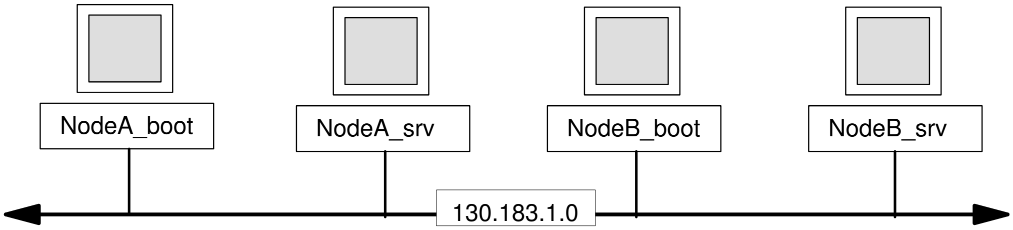
You will now create a picture representing the different adapter addresses of the cluster nodes (see the sample figures on page 11-7). The instructions below assume you are dealing with a simple 2-node cluster. Use these instructions as a guide and adapt them to your specific needs and environment.

Drawing the Boot and Service Adapter Labels

1. From the IP Devices window of ISM IP Discovery, select the boot and service adapter labels. For example, for a 2-node cluster that uses 2 service labels, you have to select 4 lines: 2 boot labels and 2 service labels. Bring up the **Control** menu and choose **Draw Selection**. A new window is displayed.
2. Choose the **New** option so that the selection will be drawn on a new picture, and the **Bus** drawing algorithm (actually, you can choose other options, since you may want to add the adapter labels to an existing picture, and/or to use other drawing algorithms).
3. Click on **Do** to draw the picture. ISM Monitor brings up a new picture, containing the IP Devices you have selected. For the moment, the picture represents only the boot and service adapter labels.

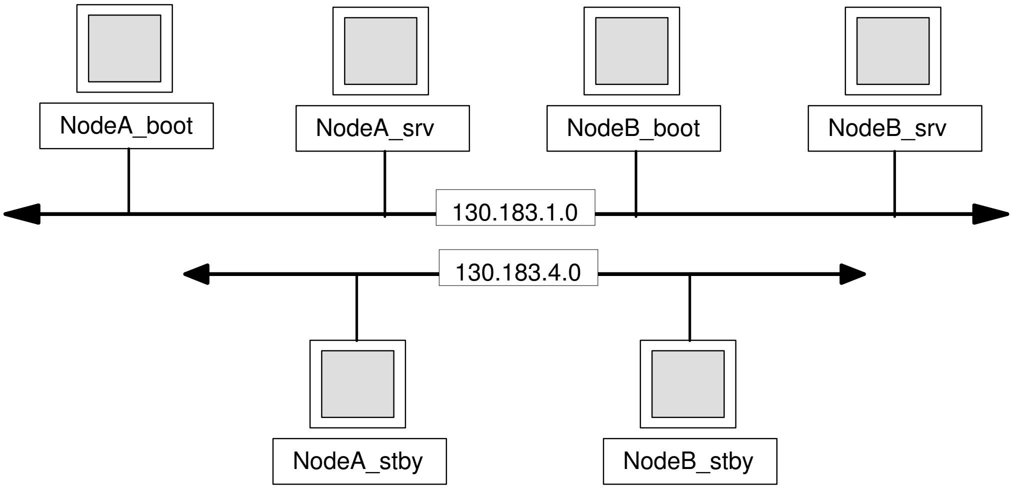
Note: Only draw the Service Adapters if you have not discovered the Boot Adapters. It is possible that the adapter labels are represented with firewall or gateway icons. Because the cluster nodes have several network adapters on distinct networks, ISM may consider the nodes are firewall or gateway machines. However, you can change these icons as you want, according to your preferences.

Once you have completed this task, your picture resembles the following:



Drawing the Standby Adapter Labels

Now, you may want to add the standby adapter labels, in order to obtain a picture resembling the following:



To add the standby adapter labels, follow the steps below:

1. From the IP Devices window of ISM IP Discovery, select the standby adapter labels (i.e., for a 2-node cluster you have to select 2 lines). Bring up the **Control** menu and choose **Draw Selection**. A new window is displayed.
2. ISM will not allow you to draw the standby adapter labels in the current picture because the same hosts are already represented through the other adapter labels (we will have to use the clipboard, as explained below). So, choose the **New** option to draw them in a new picture. Choose the **Bus** drawing algorithm.
3. Click on **Do** to draw the picture. ISM Monitor brings up a new picture, containing the network and the standby adapter labels you have selected. You have to copy them to the previous picture, as explained below, in order to obtain a picture including all the adapter labels.
4. Select the standby adapters and the corresponding network, using the **Select All** option from the **Selection** menu of ISM Monitor. From the Edit menu, choose the **Copy** option: this brings a copy of the selected objects to the clipboard.
5. Open the clipboard, using the **Clipboard** option of the **Edit** menu. Then, bring up the previous picture (which, for the moment, contains only the boot and service adapter label icons).
6. From the clipboard, select the network you have just copied, then choose the **Paste** option of the **Edit** menu. Click on the picture at the location you want to paste the network.
7. Repeat the step above several times to paste the different remaining objects: the standby adapter label icons, and the network links. You cannot select several objects in the clipboard, hence this repetitive operation. Concerning the network links, note that you must drop them twice on the picture, on the two objects you want to link (network icon and adapter label icon).

Animating the Icon Display

This step consists in animating the icon display so that changes in the state of the represented objects will be signaled by color changes in the icons.

Your picture being displayed on an ISM Monitor window, choose **Select All** from the **Selection** menu to select all the icons. From the **Animate** menu, select the **Live** option.

Where You Go From Here

At this point, you may want to save your picture. (You can discard the picture that was temporarily used to hold the standby adapter labels).

In addition, you may want to read the hints below, before proceeding with "Browsing the HACMP MIB and Adding Objects to the Picture" on page 11-11, in order to refine your picture.

Hints For Interpreting Some Basic Information

Interpreting Icon Colors

Each adapter label icon is animated according to the state of the IP primary and secondary addresses represented. (Concerning the IP primary and secondary addresses, see the remarks on page 11-6.)

- For an icon representing a service address, the primary address is the service address.
- For an icon representing a boot address, the primary address is the boot address.
- In all cases, the secondary address is the standby address.

The icon adopts the red, or green color, according to the following rules:

- **green**: when both the primary and secondary addresses are up.
- **red**: when both primary and secondary addresses are down.

Note: The icon may also be red if the **snmpd** daemon is not enabled on the concerned system.

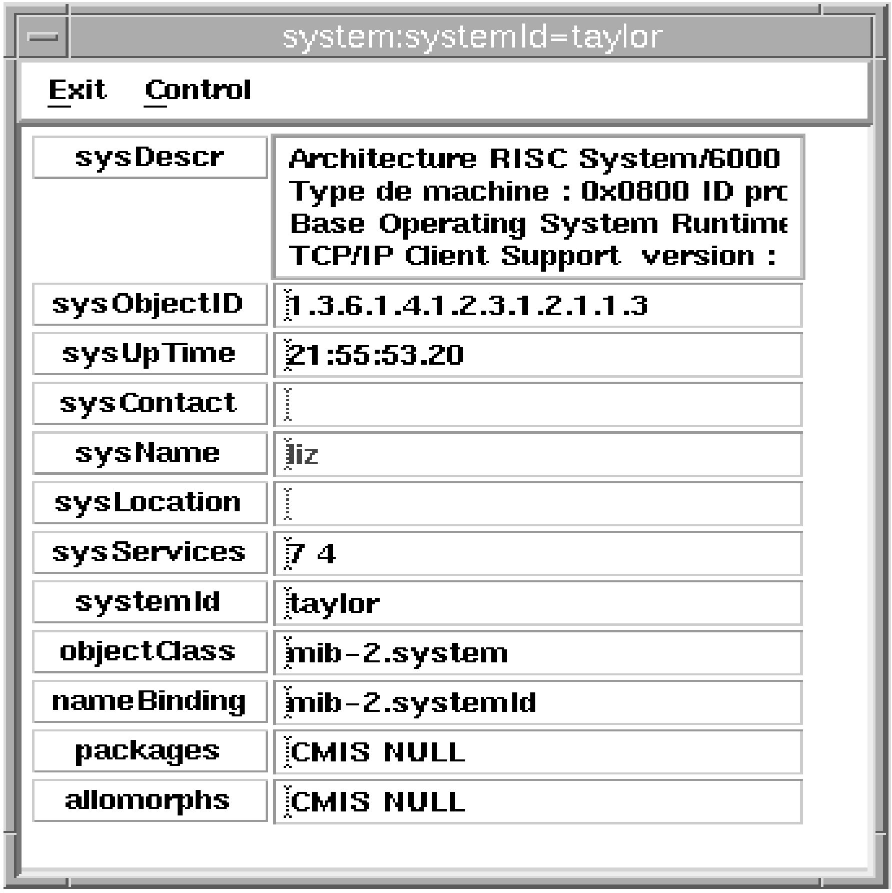
Finding the Node that Holds a Given Service Address

When an IP address takeover occurs within a cluster, one node assumes the address of another node that left the cluster.

When monitoring a cluster, you can know which node assumes a given service address, by browsing through the MIB as follows:

1. From the displayed picture, select the icon of the concerned service address.
2. Bring up the **Control** menu and choose **Detail**. Then click on the MIB–2 button and on the system button.
3. Look at the **sysName** field: it indicates which host actually holds the service address you have selected.

In the sample figure below, we see that **sysName** is the liz host.

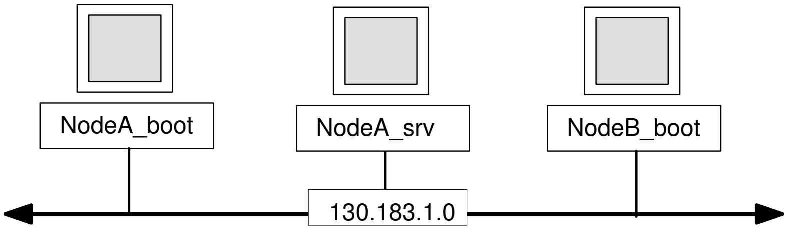


sysDescr	Architecture RISC System/6000 Type de machine : 0x0800 ID prc Base Operating System Runtime TCP/IP Client Support version :
sysObjectID	1.3.6.1.4.1.2.3.1.2.1.1.3
sysUpTime	21:55:53.20
sysContact	
sysName	liz
sysLocation	
sysServices	7 4
systemId	taylor
objectClass	mib-2.system
nameBinding	mib-2.systemId
packages	CMIS NULL
allomorphs	CMIS NULL

Sample Scenario

This section discusses a sample scenario involving a 2-node cluster. We assume here that one resource group is defined, attached to the node A, and configured in rotating standby.

Suppose that we have a picture including the objects shown below:



Now suppose that the initial conditions are as follows:

- The node A holds the service address:
  - In this situation, the NodeA_srv icon is green, while NodeA_boot is red.
  - if you display the detail of NodeA_srv, the **sysName** field shows that the host name is NodeA_boot.
- For the moment, node B uses its boot address. It is ready to takeover resources in the event of a failure on node A. The NodeB_boot icon is green.

In the event of a failure of node A, an IP address takeover occurs (assuming that the cluster implements the IP takeover capability). After the takeover, the conditions are as follows:

- Now, the node B holds the service address:
  - The `NodeA_srv` icon remains green, while `NodeB_boot` becomes red.
  - If you display the detail of `NodeA_srv`, the **sysName** field shows that the host name is `NodeB_boot` (it is no longer `NodeA_boot`).
- The `NodeA_boot` icon becomes green.

The table below summarizes these changes:

	NodeA_boot	**NodeA_srv**	**NodeB_boot**
Before IP address takeover	Red	Green (node A holds the service address)	Green
After IP address takeover	Green	Green (node B holds the service address)	Red

Browsing the HACMP MIB and Adding Objects to the Picture

Once you have drawn a basic picture with adapter labels, you can browse the HACMP MIB and add other objects, in order to refine your picture.

Prerequisite: Checking for clsmuxpd Operation

ISM obtains information from the cluster through the Cluster SMUX peer daemon, **clsmuxpd**. Thus, be sure that this daemon is running on each cluster node.

Notes:

- Normally, the **clsmuxpd** daemon is automatically started when you start cluster services on a node. If necessary, use the **Start Cluster Services** SMIT HACMP menu (fast path: **smit clstart**).
- Another method for starting **clsmuxpd** consists in issuing the following command:

```
startsrc –s clsmuxpd
```

- See “Setting Up the SNMP Environment on the Cluster Nodes”, on page 11-3, for related information.

Browsing the HACMP MIB

You can browse an HACMP MIB as follows:

1. Your picture being displayed in an ISM Monitor window, select an icon representing a service adapter.update: 1 paragraph modified
2. Bring up the **Control** menu and choose **Detail**. A new window appears, showing current values in the SNMP MIB, and including buttons that allow to access additional information.
3. Click on the private button and then on the **risc6000clsmuxpd** button to access the HACMP MIB. A new window is displayed, with a button for each HACMP managed object (address, application, clinfo...).
4. Click on any button to see the attributes of the corresponding object.

Drawing HACMP MIB Objects

You can draw HACMP MIB objects to refine your picture:

1. Your picture being displayed in an ISM Monitor window, browse the MIB as explained above. It is important to note that, **when you want to browse the MIB in order to draw objects, you must start with a service adapter label** (rather than a boot or standby one). This is because the service address is normally always valid when HACMP runs on the cluster.

   Consequently, before browsing, you should first select an icon representing a service adapter.

2. Continuing to browse the MIB, go to the object you want to draw and display its attributes. For example, you can go to a “cluster” object, by clicking on the **cluster** button.
3. Bring up the **Control** menu and choose **Draw**, then click on the picture at the location where you want to draw the object (for example: a “cluster”, if you have selected such an object).

HACMP MIB Objects of Particular Interest

The service, boot and standby adapters you have previously drawn, are `snmpSystem` objects.

In addition, for cluster monitoring purposes, objects of particular interest include:

- `cluster` object
- `node` object

Note: To see the object type associated to an icon, select the icon, execute the **Detail** command and look at the **objectClass** value.

The section “Understanding the HACMP MIB Objects”, on page 11-13, will help you to understand how the things work.

Choosing Attributes For Icon Animation

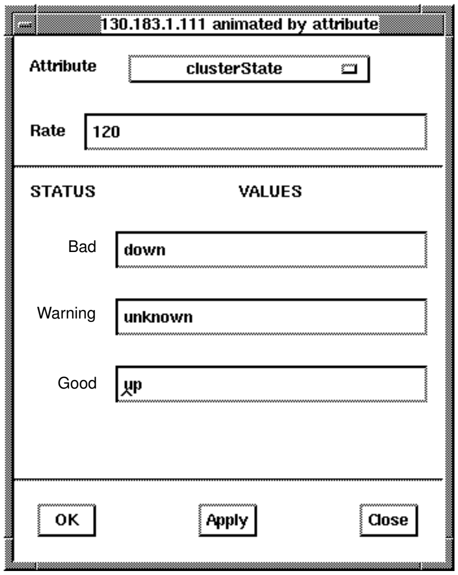
Once you have drawn an object, you can use the **Live** ISM menu option to animate the object according to the default animation mode. Note that by default, a `cluster` object is animated according to its `clusterState` attribute.

You can decide to animate an object according to the attribute you want. An interesting example consists in drawing a `cluster` object and to animate it according to its `clusterSubState` attribute. This attribute is of interest when monitoring a cluster (the next section, “Understanding the cluster and node Objects”, on page 11-13, explains why).

Procedure

To create a `cluster` object animated according to its `clusterSubState` attribute:

1. Your picture being displayed in an ISM Monitor window, select an icon representing a service label adapter, browse the HACMP MIB, go to the `cluster` object, and draw it on the picture.
2. Select the icon you have just drawn. Bring up the **Animate** menu, and use the **Animate By Attribute** feature to change the animation criteria: animate the `cluster` object according to its `clusterSubState` attribute (instead of the default `clusterState`).

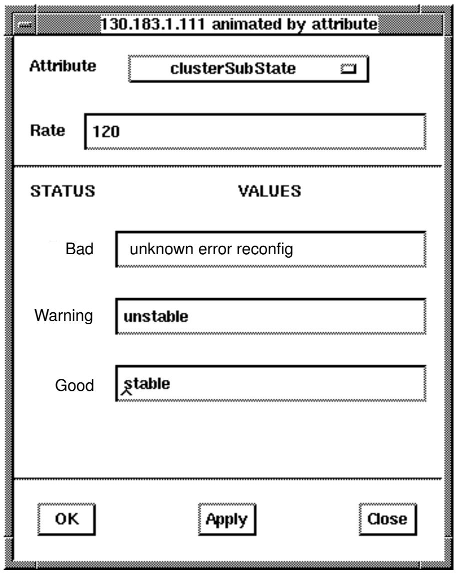


130.183.1.111 animated by attribute

Attribute: clusterState

Rate: 120

STATUS	VALUES
Bad	down
Warning	unknown
Good	up



130.183.1.111 animated by attribute

Attribute: clusterSubState

Rate: 120

STATUS	VALUES
Bad	unknown error reconfig
Warning	unstable
Good	stable

Note: Icons can also be animated according to received SNMP traps. This is explained later in this chapter.

Understanding the HACMP MIB Objects

As stated above, HACMP MIB objects of particular interest include the `cluster` and `node` objects. This section summarizes information concerning these two object types. Other objects are briefly discussed on page 11-15.

Preliminary Note on the hacmp.ext Directory

When you draw an HACMP MIB object, the icon and its animation mode are by default set according to the definitions given in **hacmp.ext**.

Once you are familiar with cluster monitoring with ISM, you may want to specify other default icons for cluster objects, or other default animation criteria. Use the ISM configurator application (menus Applications/Monitor).

cluster Object

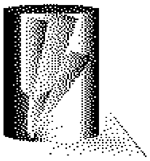
An HACMP cluster is a group of processors that cooperate to provide a highly available environment.

For the `cluster` object, the attributes of particular interest are `clusterState` and `clusterSubState`.

Using cluster Objects in an ISM Picture

Since both attributes are significant when monitoring a cluster, you may want to add to your ISM picture two cluster icons: one animated according to `clusterState`, and the other animated according to `clusterSubState`.

To access a `cluster` object, browse the HACMP MIB and click on the **cluster–nf** button (then you can draw it). The default icon for a `cluster` object, as defined in the **hacmp.ext** file, is shown below:



clusterState Attribute

The table below shows the possible values of the `clusterState` attribute and their meaning. It indicates also which color is used to display the icon label, according to the default definitions in the **hacmp.ext** directory.

clusterState	**Meaning**	**Icon Color**
up	At least one node in the cluster is up, and a primary is defined.	Green
down	At least one node in the cluster is up, but a primary is not yet defined.	Red
unknown	**clsmuxpd** is unable to communicate, or is not yet communicating with an active Cluster Manager.	Orange

clusterSubState Attribute

The table below shows the possible values of the `clusterState` attribute .

clusterSubState	**Meaning**	**Icon Color**
stable	The cluster is stable (no re-configuration is occurring).	Green
unstable	The cluster is unstable (a change in topology is occurring).	Orange
error	A script has failed, the cluster has been in the process of configuration (unstable) for too long.	Red
unknown	**clsmuxpd** is unable to communicate with a Cluster Manager.	Red
reconfig	The cluster is in the process of reconfiguration.	Red

Node Object

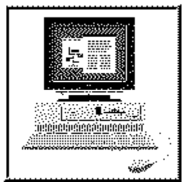
A node is one of the servers that make up the cluster. Each node runs the Cluster Manager, and the SNMP and Cluster SMUX Peer Daemon programs.

For the `node` object, the attribute of particular interest is `nodeState`.

Using node Objects in an ISM Picture

You may want to add to your ISM picture the node icons representing the nodes of the HACMP cluster.

To access the node objects, browse the HACMP MIB and click on the **node** button, then ”detail” the table instance and from here click on the **node Entry** button. You obtain a table of all the cluster nodes that you will be able to draw in your picture. The default icon for a node object is shown below:



nodeState Attribute

The table below shows the possible values of the nodeState attribute and their meaning. It indicates also which color is used to display the icon label.

nodeState	Meaning	Icon Color
up	The node is up and running.	Green
down	The node is down.	Red
joining	This node is in the process of joining this cluster.	Orange
leaving	This node is in the process of leaving this cluster.	Orange

Other HACMP MIB Objects

In addition to the cluster and node objects, the HACMP MIB maintains other objects (application, clinfo, cllockd...).

Because we consider these objects are not very useful when monitoring a cluster, they are not discussed in this document. It is up to you to determine if these objects may be of interest, according to your specific needs.

For more details on these objects, you can read the **hacmp.mib** file (which includes comments), or browse the HACMP MIB from an ISM picture. As a hint, here is a summary of the different object types maintained in the HACMP MIB:

address Describes the attributes of the network addresses of the nodes (gives for example the address role: standby, boot, address).

application Maintains information about the applications registered with the HACMP Cluster SMUX Peer.

clinfo Maintains information on the **clinfo** (cluster information) daemon.

cllockd Maintains information on the **cllockd** (cluster lock manager) daemon.

clsmuxpd Maintains statistical information about the **clsmuxpd** (cluster SMUX Peer) daemon.

clstrmgr Maintains information on the **clstrmgr** (cluster manager) daemon.

cluster Describes the attributes of a cluster.

event Maintains information on the set of 1000 most recent cluster events.

network Describes the attributes of the networks that support the cluster.

node Describes the attributes of the nodes within a cluster.

Dealing With Traps

Prerequisites

clsmuxpd concerns

The Cluster SMUX peer daemon, **clsmuxpd**, should run on each cluster node. If in doubt, check for its operation, as explained on page 11-11.

When a change occurs within a cluster, the cluster manager notifies the **clsmuxpd** daemon. In this event, **clsmuxpd** emits traps to propagate the information. For the ISM station manager to receive these traps, the **/etc/snmpd.conf** file must be set up appropriately on each cluster node. If needed, refer to “Setting Up the snmpd Configuration Files” on page 11-3.

clinfo concerns

If you plan to run the **clinfo** (cluster information) HACMP utility on your ISM manager station, note that you cannot enable it to receive traps. In other words, do not run clinfo with the **–a** option, to avoid trap conflicts between ISM and **clinfo** on the ISM station.

This is because only one SNMP manager can run on a given network station (only one TCP/IP program at a time can listen on a particular port).

Understanding HACMP Traps

ISM is able to handle traps that come from the **clsmuxpd** daemons. You can decide to animate some ISM picture icons according to the HACMP traps ISM receives. The “ISM Alarm” application is used to display the received traps.

Use the ISM Configurator (trap custom menu) to view the HACMP trap definitions.

Alarms: a Note on Terminology

According to ISM terminology, traps are referred as “alarms”.

The HACMP Traps

The **clsmuxpd** daemon may generate the traps listed below. The object associated to the trap is in brackets.

TrapClusterState	(clsmuxpd)
TrapClusterSubState	(clsmuxpd)
TrapNodeState	(node)
TrapNetworkState	(network)
TrapAddressState	(address)
TrapNewPrimary	(clsmuxpd)
TrapAppState	(application)
TrapAdapterSwap	(address)
TrapAddressTakeover	(address)

All these traps have been assigned a ”major” severity.

These traps indicate a change in the state of the main objects maintained in the HACMP MIB. The traps of particular interest are:

- `TrapAddressTakeover`, which indicates that an IP takeover occurred.
- `TrapAdapterSwap`, which indicates that an adapter swap occurred.

  It is recommended to assign the two traps above a ”critical” severity.

A series of 27 other HACMP traps exists, with an undetermined severity, all associated to the `node` object. Among them, those listed below indicate failures in the HACMP cluster behavior and have been assigned a ”critical” severity.

```
trapFailNetwork
trapFailNode
trapClusterConfigToolong
trapClusterUnstableToolong
trapEventError
```

All the HACMP traps have been customized with a specific problem. This specific problem is an OID aliased with the HACMP trap additional text. The specific problem allows you to filter each alarm specifically. For example: to animate an icon which will turn to red only on the occurrence of an adapter swapping.

To visualize customization proposed for HACMP traps (Alarm Severities and Specific Problem), use the ISM Configurator Trap Custom and Alias Server tools.

Viewing Received Traps

Displaying the Alarm List

To view the traps (i.e. the alarms) received by ISM:

1. In the ISM Application Board, select the ISM Alarm application.
2. Bring up the **Execute** menu, and choose the **from file...** option. A list of configuration files is displayed, including the **default.ALcfg** file.
3. Select the **default.ALcfg** file and click on **OK** to load it. An Alarm window appears, that lists in a table the alarms received by ISM manager. In the table, each line corresponds to a set of occurrences of the same alarm from the same system.

Changing the Default Presentation of the Alarm List

In the table, the **EventType** and **ProblemType** columns are displayed at the left side of the table, so you see them first. However, all the HACMP traps belong to the `processing` **EventType** and `sfwrEnvironmental` **ProblemType** (this is defined in the **hacmp.trapdb** file). As a consequence, the information presented in these two columns does not discriminate, and you may find it more convenient to change the default presentation.

So we suggest you change the presentation by bringing the **ProblemText** column at the left side of the table. Displaying the **ProblemText** in the configuration file(s) allows you to easily distinguish HACMP traps from other `processing` / `sfwrEnvironmental` traps emitted by other specific SNMP agents.

To reorder the displayed columns, use the **Preferences...** option of the **Control** menu from the ISM Alarm window.

Generating HACMP Traps for Testing

In order to test the features described above, you may want to provoke the generation of HACMP traps.

To do so, change something in your cluster state: for example, disconnect a service adapter, or stop cluster services on one node. A few seconds later, you will see HACMP traps in the ISM Alarm window.

Animating ISM Picture Icons According to Received Traps

You may want to improve your picture by adding objects that are animated according to the received HACMP traps.

The sample procedure below may serve as a guideline.

1. On your ISM picture, draw a new icon, for example an object of the `cluster` type. Select it.
2. Bring up the **Animate** ISM Monitor menu, then the **Animate By** submenu. In the new menu that appears, select the **Alarms** checkbox and deselect the **Attribute** checkbox.
3. Bring up the **Animate** menu again. Choose successively **Animate By**, **Configure** and **Filtered Alarms**. A configuration panel is displayed.
4. The configuration panel allows you to filter the alarms that will animate the icon initially selected (in our example, a `cluster` object). Various filtering criteria can be specified.

   For example, to only take into account the HACMP traps that come from a given cluster, use the Source Object list editor and add lines having the following form:

   ```
   snmpSystemId=node_name/privateId=node_name/risc6000clsmuxpdId=nod
   e_name/*
   ```

   Thus, assuming the concerned cluster includes two nodes named **node_foo** and **node_bar**, you have to add the following lines:

   ```
   snmpSystemId=node_foo/privateId=node_foo/risc6000clsmuxpdId=node_
   name/*
   snmpSystemId=node_bar/privateId=node_bar/risc6000clsmuxpdId=node_
   name/*sysNameId=node_foo/risc6000clsmuxpdId=node_foo/*
   sysNameId=node_bar/risc6000clsmuxpdId=node_bar/*
   ```

Note: An easy way to determine the *node_name* consists in locating HACMP alarms in an ISM Alarm window (see ISM Alarm on-line help).

   If you want to animate the icon on the occurrence of a particular HACMP trap or a subset of HACMP traps, use the Specific Problem filtering criteria.

5. Now, animate the icon: bring up the **Animate** menu and choose **Live**. The label of the icon becomes green or red.

Icon Label Color

- If the icon label is green, no HACMP traps have been received yet.

  If you generate test HACMP traps (by changing something in your cluster), then, a few seconds later, the icon label becomes red.

- The red color indicates that HACMP traps have been received.

  To display the traps, select the involved `cluster` icon, bring up its associated pop-up menu, and select the **ISM Alarm** item. An ISM Alarm window is then displayed, listing all the HACMP traps that have been emitted by the **clsmuxpd** daemons that run on the different cluster nodes.

Monitoring Disk–Related Errors

To monitor disk-related errors, use teh ISM UNIX agent **alixd**. This agent automatically generates traps on the occurrence of errors i the AIX error log.

Chapter 12. Integration of a PCI Drawer

This chapter uses a typical configuration to show how to integrate a PCI drawer in an Escala EPC rack.

PCI Drawer Integration

Example of a Configuration

Cluster node with 3 Ethernet interfaces (service, standby, administration) and 2 SCSI-DE buses.

The basic configuration is as follows:

Ethernet en0	Service interface
Ethernet en1	Standby interface
Ethernet en2 (native)	Administration interface
scsi0	
scsi1	

The adapters **Ethernet Standby** and **scsi1** are to be moved and installed in the PCI drawer.

Note: In order that the ODM database does not become ”polluted” by the Defined interfaces, and that there is creation of a supplementary interface (ent3 ..., scsi2 ...), all the Ethernet and disk interfaces must be removed before attempting this integration procedure.

Integration Procedure

The steps are as follows:

1. Stop HACMP on the node which is to receive the PCI Drawer, with ”takeover” on the other node, so that there is no interruption of service. At the same time invalidate HACMP automatic restart at reboot.

2. Using the following command:

   ```
   lsattr –E –l scsi1
   ```

   then note the scsi ID number of the interface.

3. Stop the Ethernet and disk interfaces and unconfigure them as follows:

   a. `smit chinet`
      select en <x> for interface en0, en1, en2
      `Current STATE     [detach]`

   b. `rmdev –dl en` <x> for interface en0, en1, en2

   c. `rmdev –dl  et` <x> for interface et0, et1, et2

   d. `rmdev –dl ent` <x> for interfaces ent0, ent1, ent2

   e. `rmdev –dl hdisk` <x> for hdisk x, y ...

   f. `rmdev –dl atf` <x> for atf x, y ...

   g. `rmdev –dl sp` <x> for sp x, y ...

   h. `rmdev –dl scsi1` (adapter SCSI which will be moved).

   Do not unconfigure the scsi0 adapter, which will stay in place; otherwise remember to update its scsi ID number after having isolated the corresponding bus, as described above.

4. Stop the node: `shutdown –F` and power–off.

5. Disconnect the SCSI bus from the interface to be moved (scsi1). Then disconnect the Ethernet cables from the adapters.
6. Withdraw the adapters to be moved and install the interface adapter in the PCI drawer.
7. Install the PCI drawer in the rack.
8. Connect all the Ethernet cables. Then connect a terminator on the rear connector of the SCSI board placed in the PCI drawer.
9. Declare the PCI drawer (using Menu 9 of the ”sbb” menu). The following response can be expected

   ```
   CONFIGURATION in progress
   WAIT #
   UNIT FOUND : 1
   ```

10. Power-up the node and its PCI drawer and reboot AIX.
11. Re-configure the Ethernet interface using **smit chinet**

    select en <x>
    x = 0 for boot interface
    x = 1 for standby interface
    x = 2 for administration interface

    ```
    INTERNET ADDRESS        [@IP]
    Network MASK            [user defined]
    Current STATE           [up]
    ```

12. Update routing table, if necessary, using **smit mkroot**:

    ```
    DESTINATION Address     [default]
    GATEWAY Address         [@IP_gateway]
    METRIC                  [user-defined]
    Network MASK            [user-defined]
    ```

13. For the SCSI interface change the scsi ID with the value found at step 2 using **smit chgscsi**:
    select scsi1

    ```
    Adapter card SCSI ID    [x]
    ```

14. De-configure the configuration for SCSI, atf and sp, as follows:
    a. `rmdev –dl  atf<x>`
    b. `rmdev–dl sp<x>`
15. Stop the machine: `shutdown –F` and power–off.
16. Connect the common SCSI bus in place of the terminator.
17. Power-up the node and reboot AIX.
18. Check the presence of:

    disks
    array drivers
    volume group
    corresponding pv –id
    network

    using the following commands:

`lsdev –Cc disk`	to check ”hdisk x, y, ...”
`lsdev –Cc array`	to check ”sp x, y, ...”
`lsdev –Cc driver`	to check ”atf x, y, ...”
`lspv`	to check pvid and their associated volume groups
`netstat –r`	to check Ethernet interface and gateway

19. When everything is working correctly, start HACMP on the node.

Appendix A. Implementing Console Solutions

This appendix discusses the different solutions that can be implemented to handle the consoles of the cluster nodes. It includes the following sections:

- Console Solutions Overview.
- Console Solutions for HA Solutions, on page A-2.
- Console Solutions for Escala Powerclusters (including the PowerConsole), on page A-5.
- Setting Up and Using the Console Concentrator, on page A-12.
- Setting Up and Using the pwcons Utility, on page A-19.

Console Solutions Overview

A cluster node, as any ordinary UNIX server, requires a system console. For example, if you have a cluster comprising 4 nodes, then 4 distinct consoles are used in the cluster.

To handle the consoles of the nodes of a cluster, several solutions can be implemented, based on different hardware components.

The ordinary console implementation consists in attaching one ASCII terminal to each node. However, if the cluster includes many nodes, you may think that all these terminals clutter your computer room. In addition, using multiple terminals to control a cluster (which constitutes from the administrative standpoint a single computing entity), may be regarded as unpractical. Moreover, ASCII terminals are not very attractive, and many system managers prefer to use graphical displays.

For these reasons, alternative console solutions are proposed.

Console Solutions for HA Solutions

For HA Solutions, the following solutions are proposed to handle the consoles:

- Conventional Console Solution, see page A-2.
- Serial Multiport Adapter -based Solution, see page A-3.

Console Solutions for Escala Powerclusters

To handle the consoles of the nodes of a Powercluster, several solutions can be implemented, based on different hardware components. The most powerful solution, referred to as the PowerConsole, is based on an AIX workstation and a console concentrator. The AIX workstation may be an Escala S100 Series platform. The PowerConsole and its associated ClusterAssistant graphic software package provide the cluster administrator with a very attractive single point of control. Alternative solutions are also proposed.

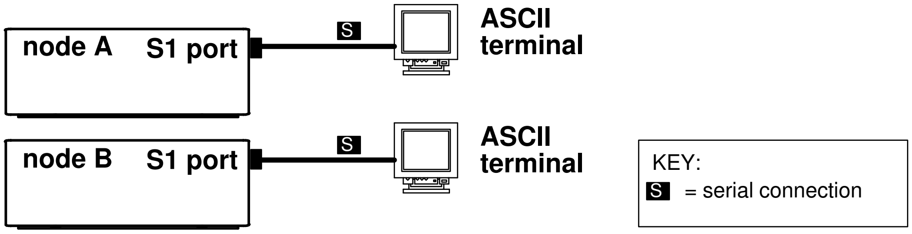
- For details, refer to “Console Solutions for Escala Powerclusters”, on page A-5.

Console Solutions for HA Solutions

Conventional Console Solution

With this conventional solution, each node has its own console directly attached. Each console is usually an ASCII terminal attached to the S1 console port of the node to control. You implement this console solution as you would if you had to install several independent servers.

The figure below shows a sample conventional console solution:



Variants of the conventional console solution

Graphical Consoles

Some HA Solutions are based on servers equipped with graphical consoles instead of ASCII terminals. In that case, the servers handle the console input and output through the graphical console and the S1 console port is unused.

PC Computers

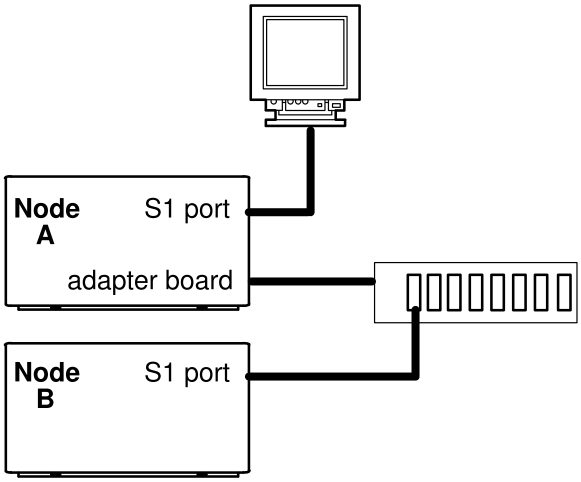
Instead of using ordinary terminals (ASCII or graphical consoles) as system consoles, you can rely on PC computers connected to the S1 serial ports of the cluster nodes. Then, on the screen of your PC computer(s), you can display several terminal emulation windows, each window acting as the system console of one node. You can use the terminal emulation programs that come with MS-Windows 3.1 and with MS-Windows 95, or other third-party programs as well.

This PC-based console solution can be implemented through the “PC Operator’s Console Kit” which includes appropriate cables and documentation. For detailed information, refer to the *PC Operator’s Console Facility User’s Guide* (86A294AT).

Serial Multiport Adapter-Based Console Solutions

If you want to limit the number of console terminals used in your configuration, you can implement a configuration where only one or two terminals handle multiple consoles. In such a configuration, one of the nodes is equipped with a serial 8-port adapter through which the S1 console ports of the other nodes are handled. The two figures below depict such configuration examples.

Example 1: The figure below shows a sample configuration similar to the first one, except it is used for a 2-node cluster.

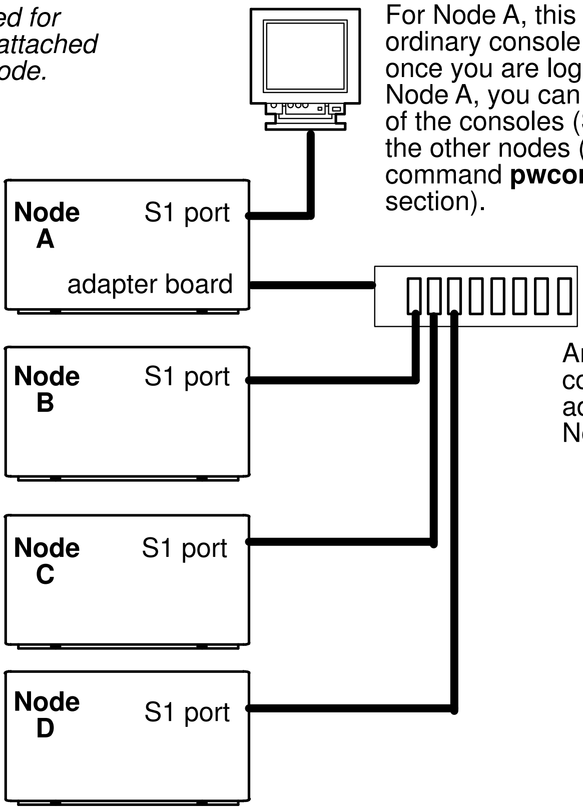


Example 2: The figure below shows a sample configuration including 4 nodes and a single physical terminal.

A single terminal is used for the whole cluster. It is attached to the S1 port of one node.

For Node A, this terminal is an ordinary console. In addition, once you are logged in to Node A, you can gain control of the consoles (S1 ports) of the other nodes (see the command **pwcons** later in this section).

Only one node has an attached terminal. This node is equipped with an 8-port adapter.

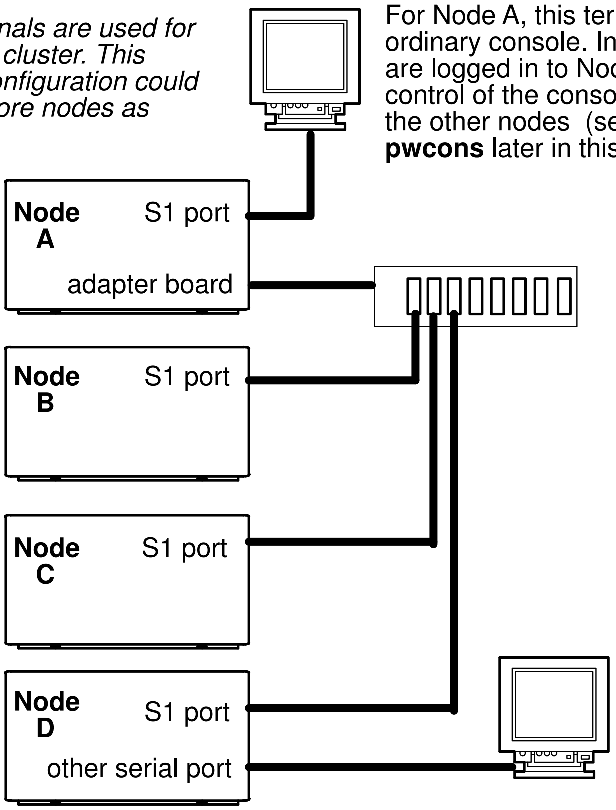


An 8-port adapter connected to its adapter board on Node A.

The other nodes have no terminal attached.
Their S1 ports are connected to the 8-port adapter of the node that has a terminal attached.

Note: On any Escala model, the S1 serial port is dedicated to console connections.

Example 3: The figure below shows a sample configuration similar to the previous one, except it uses two physical terminals.



Two terminals are used for the whole cluster. This sample configuration could include more nodes as well.

For Node A, this terminal is an ordinary console. In addition, once you are logged in to Node A, you can gain control of the consoles (S1 ports) of the other nodes (see the command **pwcons** later in this section).

One node (called the “**reference node**”), here Node A, is equipped with a terminal and a multiport adapter that fans out links to S1 ports of other nodes.

The second terminal is attached to any other node.

The second terminal is used as an ordinary console for Node D. It cannot be used to gain access to S1 ports of other nodes.

Benefits and Drawbacks of Serial Multiport Adapter -Based Solutions

Configurations where a single terminal handles multiple consoles (as in the three examples above) allow you to gain space in the computer room, and to get rid of many cables everywhere. Thus, implementing such a console configuration can be an advantageous choice.

However, when using configurations that do not have one physical terminal per node, there are some drawbacks. Notably, in the event of a failure of the node that has the multiport attached, then you cannot rely on it to attain the S1 ports (console ports) of the other nodes. In such cases, if you need to access the S1 port of another node, you have to physically connect a terminal to this S1 port. This is somewhat inconvenient. In the other hand, in practice, it is infrequently required to operate a system from its console device.

Using the pwcons Utility

When consoles are handled through a multiport adapter, **pwcons** (a utility that comes with the BullCluster software package), can be used from the single terminal to take control of the consoles (S1 ports) of the different nodes. For details, refer to “Setting Up and Using the **pwcons** Utility”, on page A-19.

Console Solutions for Escala Powerclusters

Components of Console Solutions for Escala Powerclusters

To handle the consoles of the nodes of a Powercluster, several solutions can be implemented, based on different hardware components, summarized below.

- Any Powercluster comes with a conventional ASCII terminal (BQ306 model), referred to as a **system console**, see page A-5
- Alternatively, the ASCII terminal can be replaced with a **Graphics Display**, see page
- In addition to this ASCII terminal, depending on the option you have ordered, Powercluster configurations can optionally be equipped with:
  - either, an X terminal, referred to as a **cluster console**, see page **A-6**.
  - or, an Escala S100 Seriesworkstation, referred to as a **PowerConsole**, see page A-6.
- A **console concentrator** (3Com CS/2600 Communications Server) can also be used:
  - it is provided as standard equipment with configurations that include more than two nodes, and with 2–node configurations that are equipped with a PowerConsole workstation, see page A-6.
- An ethernet **hub** (3Com SuperStack II Hub 10 12–port TP) can be provided with any Powercluster configuration for implementing a dedicated administration network, see page A-7.

Role of Console Solutions Components

Many configurations can be achieved using the console solutions components. Below are indications about some of their possible uses. It is up to you to determine the configuration that matches your requirements. As an illustration, refer to the sample configurations depicted hereafter.

System Console (ASCII Terminal)

The **system console** (ASCII terminal) is offered in the following cluster configurations:

- If the configuration includes a console concentrator, the ASCII terminal can be used as a backup console. In that case, keep the terminal connected to the console concentrator. In the event of a concentrator failure (which would cause loss of access to all the S1 console ports), or of a failure of the graphical console, you can temporarily connect the backup terminal to the S1 console port of a node, and use it as a console until the failed device is repaired.
- If the configuration does not include a console concentrator (discussed below), the ASCII terminal can be used as an ordinary system console of one of the cluster nodes. In that case, connect it to the S1 console port of the desired node.
- Example 1, on page A-3, is a basic solution which applies for a two-node powercluster with no cluster console or PowerConsole.

Graphics Display

As an alternative to the system console, the **graphics display** can be used in the following cluster configurations:

- In a uni-node Powercluster, it can directly replace a system console.
- In the case of a two-node Powercluster, a system console can be connected to one node and a graphics display to the second node. This is the solution in the **Disaster Recovery** architecture, with the graphics display connected to the second node.

Note: No graphics display is available with an EPC800.

Cluster Console (X Terminal)

The **cluster console** (X terminal) is equipped with a serial port which can be operated from a terminal emulation window, thus giving access to system consoles. The X terminal is able to boot from its own resident memory (this capability improves its availability, because the X terminal has not to download its software through the network, and can be started even if the network resources are unavailable). The X terminal may be used in the following ways:

- It can be directly attached through its serial port to the S1 console port of a node.
- For two-node operation an **administration hub** is required.
- If the configuration includes a **console concentrator** (discussed below), the X terminal can be connected through its serial port to the concentrator, which in turn is responsible for handling the S1 console ports of the different nodes.

  In both cases, you will also connect the X terminal to the network, in order to take advantage of its inherent multi–windowing and network computing capabilities. This can be either a dedicated *administration network* implemented using the provided hub, or your enterprise network.

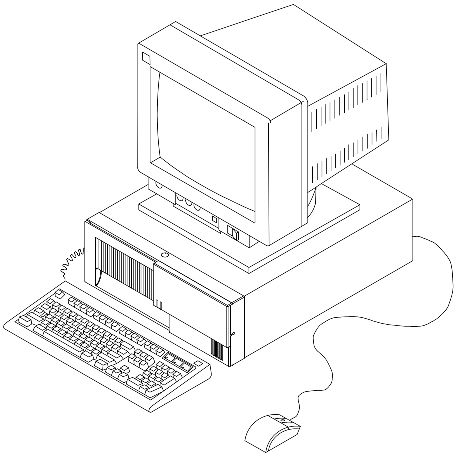
  If there is no administration hub (no dedicated administration network, the console concentrator and cluster console must be connected to your enterprise network.

PowerConsole

The **PowerConsole**, available as an option with Powerclusters, is an AIX workstation that provides a single point of control for managing clusters. It is typically used in conjunction with a console concentrator.

The workstation serves both as the access point (via the concentrator) to the different system consoles, and as a powerful workstation able to run any administration utility. The PowerConsole comes with ClusterAssistant, which is a set of graphic applications that run on the PowerConsole for configuring, managing, and monitoring clusters. ClusterAssistant is integrated in CDE (Common Desktop Environment), which provides a windows/icons/mouse -based framework. This makes operation intuitive and simple.

Escala S Series Workstation



Console Concentrator

The **console concentrator** (3Com CS/2600 Communications Server) provides the means to concentrate the S1 console ports of all the nodes, and thus to make them available from a single access point. This avoids having one ASCII terminal per cluster node, which could be found cumbersome for clusters that include three or more nodes.

Hub

The **hub** (3Com SuperStack II Hub 10 12–port TP) offers twelve 10Base–T (Twisted Pair) ports. It is intended to implement a dedicated *administration network* that interconnects the console-related devices (console concentrator, cluster console, PowerConsole) and the cluster nodes. Having such an independent administration network provides enhanced security. Otherwise the PowerConsole is connected to the customer's public network to access the node for management purposes.

Sample Console Configurations for Powerclusters

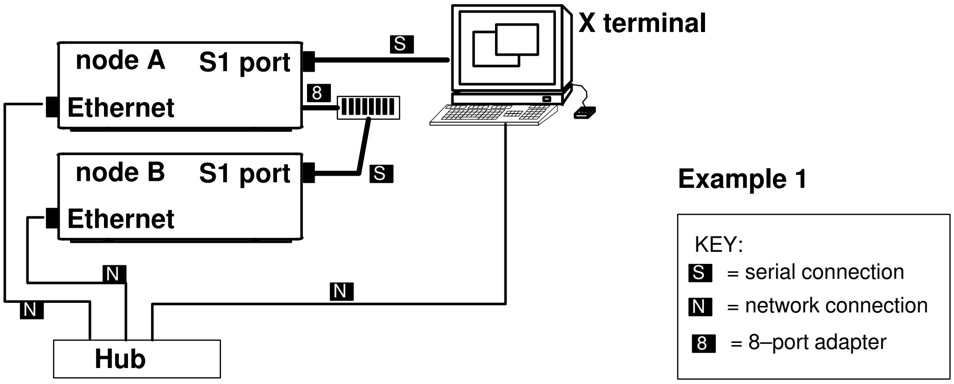
This section presents some typical configurations. Note that these are examples only, and other console configurations can be implemented.

See the *EPC Connecting Guide* for actual console configurations.

Example 1 (2–node cluster, cluster console)

The figure below depicts a cluster with two nodes (A and B), whose consoles are handled through a cluster console (X terminal). In this sample configuration:

- Only node A has a directly-attached terminal. Node A is equipped with an 8–port serial adapter, and the S1 port of node B is connected to this 8–port adapter. (Note: configurations based on an 8-port adapter are also discussed in "Serial Multiport Adapter -Based Console Solutions", starting on page A-3).
- You gain control of the console of node A via the serial line and the terminal emulation window of the X terminal.
- Once you are logged in to node A (through the serial line), you can gain control of the console of node B, by using the `pwcons` command. This command is fully described in "Setting Up and Using the pwcons Utility", starting on page A-19.
- For enhanced security, we have chosen to use the hub to implement an independent *administration network*, which interconnects the nodes and the X terminal. From the latter, you can access both nodes via the network, and run administration utilities as desired, taking advantage of the multi–windowing capability of the X terminal.

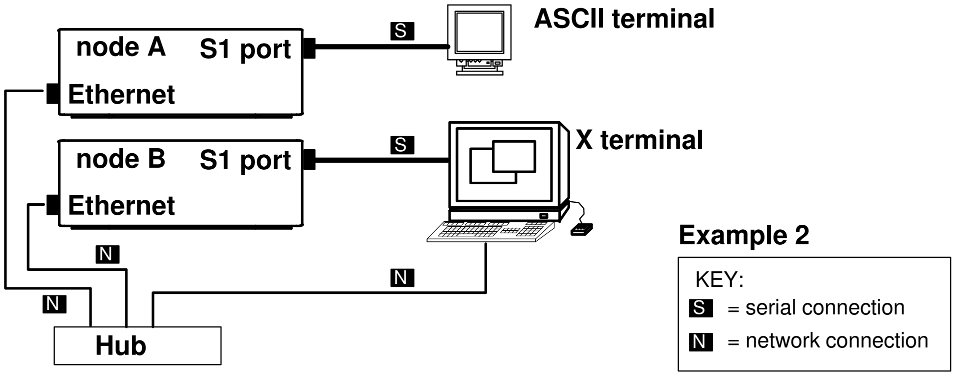


Example 2 (variant of example 1)

The figure below depicts a variant of the example 1, where the ASCII terminal is used. It shows a cluster with two nodes (A and B), whose consoles are handled with a system console (ASCII terminal) and a cluster console (X terminal). In this sample configuration:

- The ASCII terminal allows you to gain control of the console (S1 port) of node A.
- The X terminal, via a serial line and a terminal emulation window, allows you to gain control of the console of node B.

- As in example 1, the hub to implement an independent *administration network*, which interconnects the nodes and the X terminal.



Example 3 – PowerConsole (N–node cluster, AIX workstation, and console concentrator)

The figure below depicts a cluster with three nodes (A, B, and C), whose consoles are handled through a CS/2600 console concentrator and an AIX workstation (PowerConsole). Note that, although this example shows a 3–node cluster, similar configurations can be achieved as well for clusters that have 2 nodes or more than 2 nodes.

In this sample configuration:

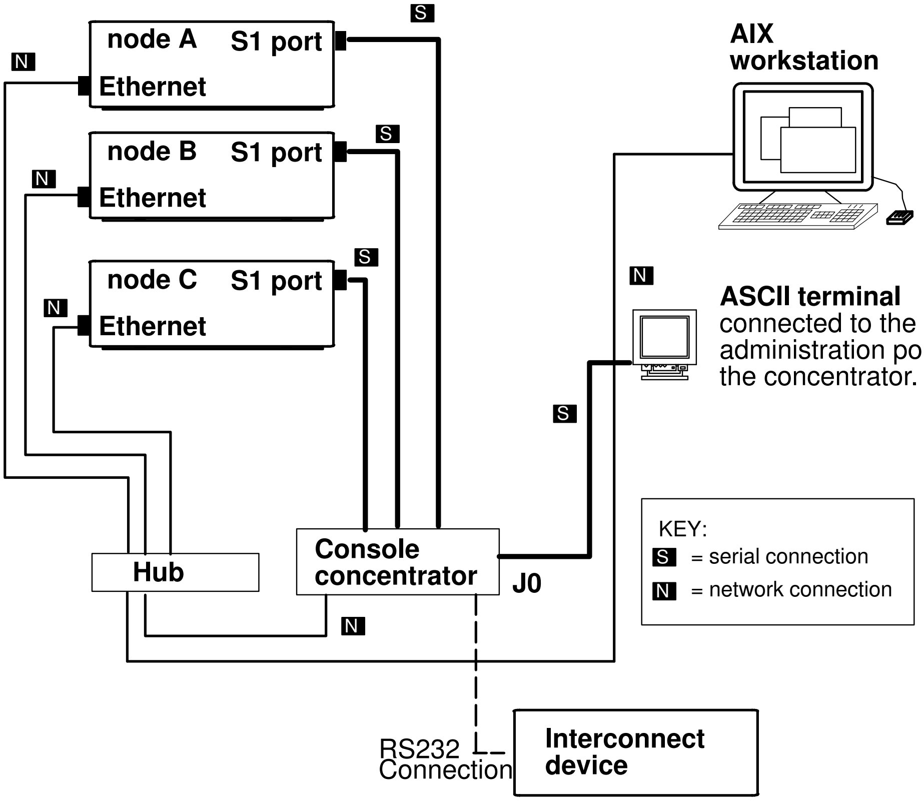
- The workstation is connected to a serial port (J4, for example) of the concentrator. (In this case, the workstation must be near the PowerCluster).
- An ASCII terminal is connected to the J0 serial port of the concentrator. On a CS/2600 concentrator, this J0 port is special in the sense it is used to access the menus and commands to manage the concentrator.
- The S1 console ports of the nodes are connected to the J1, J2 and J3 serial ports of the concentrator. Thus, all these console ports can be accessed via the concentrator.
- From the workstation, you can access at any time, through the concentrator functions, any of the concentrator ports. Thus, you can gain control of the console (S1) of any node. For further information on how to configure and use the concentrator, refer to “Setting up and using the console concentrator”, on page A-12.

In addition, note the following:

- In the event of a concentrator failure (which would cause loss of access to all the S1 console ports), you can temporarily connect the ASCII terminal to the S1 console port of a node, and use it as a console until the failed device is repaired.
- For enhanced security, we have chosen to use the hub to implement an independent *administration network*, which interconnects the nodes, the concentrator, and the workstation. From the latter, you can access the nodes via the network, and run administration utilities as desired, taking advantage of the computing power of the workstation. Without the hub the PowerConsole and Console Concentrator are connected to the ethernet enterprise network.
- If you implement RSF (Remote Services Facilities, a software dedicated to error monitoring), connect the RSF modem to the S2 serial line of the workstation, configure RSF on the workstation and on the nodes.

Note: The Escala S100 Series workstation has an internal modem.

- For further information on how to configure RSF, refer to the RSF documentation.


node A
S1 port
Ethernet
node B
S1 port
Ethernet
node C
S1 port
Ethernet
S
N
AIX workstation
ASCII terminal connected to the J0 administration port of the concentrator.
Hub
Console concentrator
J0
KEY:
S = serial connection
N = network connection
RS232 Connection
Interconnect device

Example 4 (N–node cluster, X terminal, and console concentrator)

You can implement a variant of the example 3 above, consisting in using an X terminal (cluster console) instead of a workstation (PowerConsole). The logical scheme is identical. In this sample configuration:

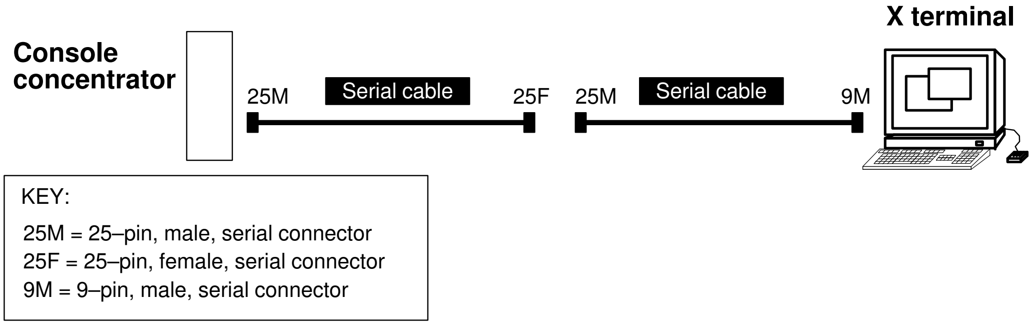
- The X terminal can be connected, via its serial port, to a serial port (J4, for example) of the concentrator.
- An ASCII terminal is connected to the J0 serial port of the concentrator. On a CS/2600 concentrator, this J0 port is special in the sense it is used to access the menus and commands to manage the concentrator.
- The S1 console ports of the nodes are connected to the J1, J2 and J3 serial ports of the concentrator. Thus, all these console ports can be accessed via the concentrator.
- From the X terminal, you can access at any time, through the concentrator functions, any of the concentrator ports. Thus, you can gain control of the console (S1) of any node. For further information on how to configure and use the concentrator, refer to “Setting up and using the console concentrator”, on page A-12.

Guidelines for Concentrator Cabling

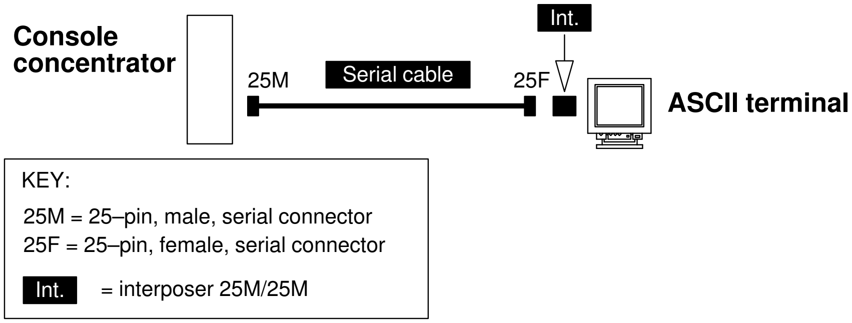
Note: This section only discusses how to connect the concentrator serial ports to the different possible consoles, and how to implement an ethernet connection between the concentrator and the hub. If you need cabling information related to the S1 console ports of the servers (cluster nodes), refer to the server’s hardware documentation and to the *EPC Connecting Guide*.

Note: A serial cable is a straight cable and not a crossed cable.

The figure below shows the cables you need to implement a serial connection between the serial port of a cluster console (X terminal) and a serial port of the concentrator.

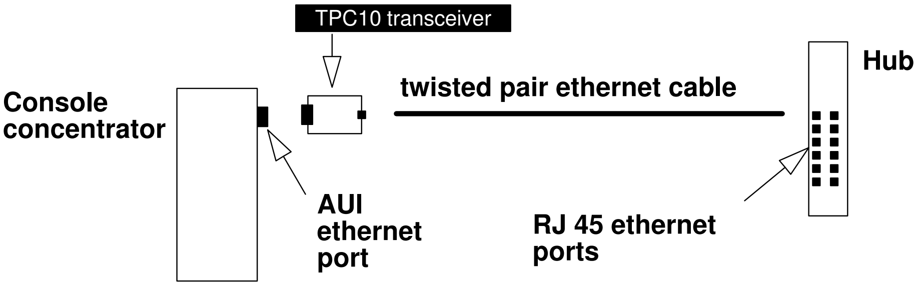

X terminal
Console concentrator
25M
Serial cable
25F
25M
Serial cable
9M
KEY:
25M = 25–pin, male, serial connector
25F = 25–pin, female, serial connector
9M = 9–pin, male, serial connector

The figure below shows the parts you need if you want to implement a serial connection between a BQ306 ASCII terminal and a serial port (J0 typically) of the concentrator.


Int.
Console concentrator
25M
Serial cable
25F
ASCII terminal
KEY:
25M = 25–pin, male, serial connector
25F = 25–pin, female, serial connector
Int. = interposer 25M/25M

Hub Concerns

The console concentrator is equipped with an AUI (thick ethernet) port. If you plan to connect the concentrator to the provided hub, which is equipped with RJ45 (twisted pair ethernet) ports, you must use a TPC10 ethernet transceiver as depicted below.



Rack Mounting Concerns

There is an optional kit for mounting the console concentrator in the rack.

CAUTION:
When installing the console concentrator inside the rack the exterior ambient temperature must not exceed 35 °C.

Setting Up and Using the Console Concentrator

Preliminary Remarks

The procedures below are given as guidelines only. For detailed information on setting up, using and managing the console concentrator, refer to the *CS/2500 Multiprotocol Communications Server* manuals.

In the sample procedures below, we assume that your configuration includes an ASCII terminal (BQ306) connected to the J0 administration port of the concentrator and used as a terminal dedicated to concentrator management. The terminal data transmission rate must be 9600 bauds.

Note: Other, non-standard, configurations are possible (for example, the J0 administrative port could be connected to the AIX workstation).

Provision a diskette (3½", double face/density) in order to copy the original 3Com diskette.

Before proceeding read the *Software Release Bulletin* (SRB) that comes with the PowerConsole software. It includes environment requirements and restrictions, as well as late-breaking news.

PowerConsole Network Connections

Dedicated Administration Network

When located on an independent (dedicated) administration network, the HACMP configuration of the clusters (managed by a PowerConsole or Cluster Console) must be modified to add this network. The network must be declared "public" in HACMP.

Note: It is important that HACMP administration sees the Network attribute as "public".

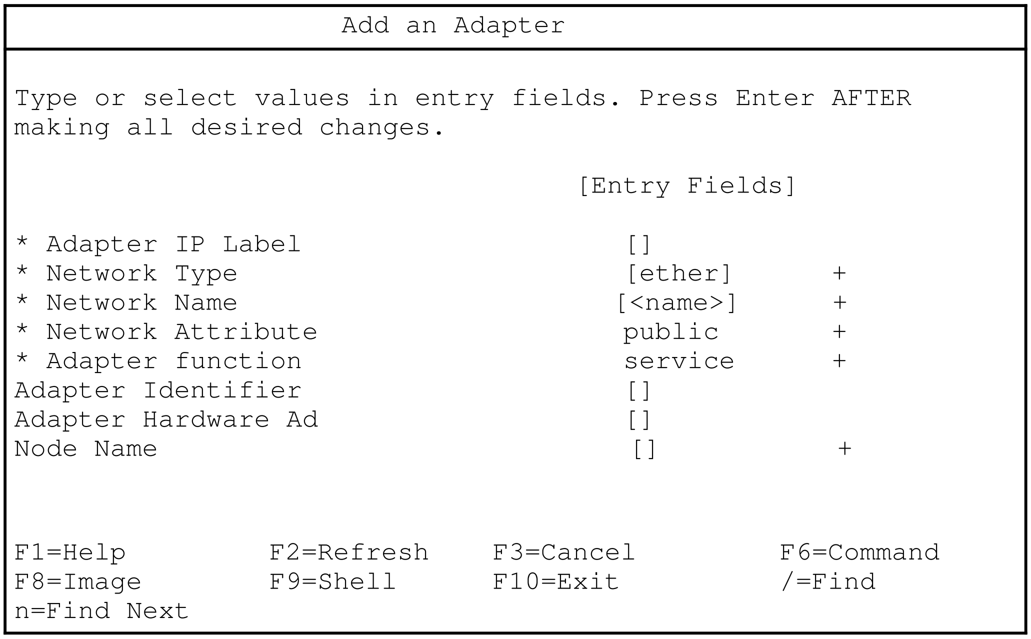
On each node, the ethernet network interface (with the administration network) must be configured as follows:

- AIX Configuration

  Use the command **smit chinet** on each mode to configure the Ethernet interfaces of the dedicated administration network. On each mode, update the */etc/host* file with the new interface names and addresses.

- HACMP Configuration
  - Logon to one of the nodes
  - modify the HACMP configuration to add the adapter used for the administration network on each node:
    - **smit hacmp**
    - **cluster configuration**
    - **cluster Topology**
    - **Configure Adapters**
    - **Add an Adapter**.

– The following window is displayed:



```
                         Add an Adapter

Type or select values in entry fields. Press Enter AFTER
making all desired changes.

                                   [Entry Fields]

* Adapter IP Label                    []
* Network Type                        [ether]      +
* Network Name                       [<name>]      +
* Network Attribute                   public       +
* Adapter function                    service      +
Adapter Identifier                    []
Adapter Hardware Ad                   []
Node Name                             []           +


F1=Help         F2=Refresh    F3=Cancel         F6=Command
F8=Image        F9=Shell      F10=Exit          /=Find
n=Find Next
```

– Make the selections and press Enter.

– Repeat these last operations to add an adapter on each node.

– Synchronize the cluster topology on all the nodes.

Cluster Public Network

IP addresses of the PowerConsole and CS2600 ports must belong to the same logical subnet as the boot and service addresses of the nodes.

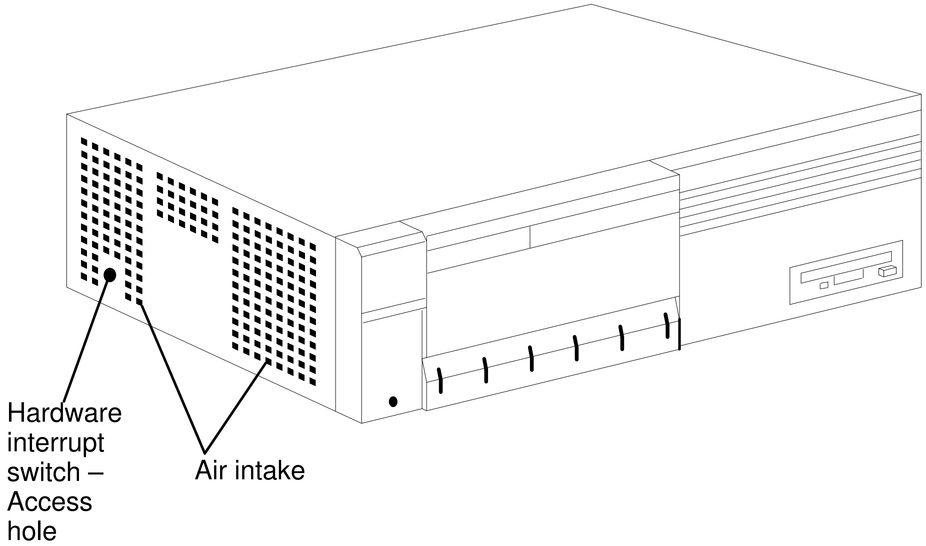
Booting the Console Concentrator

- The ASCII terminal (connected to the J0 administration port of the concentrator) should be properly set up, with the following characteristics: baud rate 9600, 8 bits per character, 1 stop bit, no parity, DTR ignored, full duplex, and echo off.
- Insert the boot diskette in the diskette drive of the concentrator (the diskette has the label “83-0377-004 CS/2600 SW/2500–TO–3270–LOCAL)”.

  Then, power on the concentrator, and wait until the end of self–tests and initialization when the yellow ”self test” LED turns off and wait until the diskette drive LED turns off (at least three minutes).

Setting-up Console Concentrator Base Parameters

- Perform a hardware reset of the concentrator, as explained in the *CS/2500 Series Communications Server Installation Guide*. The access hole for the hardware interrupt switch is shown in figure.



Warning

Use only a non-conductive object, such as a plastic stylus, to press the hardware interrupt switch. Do not use the tip of a pencil. Graphite particles can cause electrical shock to the operator and can damage components on the server's circuit boards.

- Wait a few seconds then press the <Enter> key two or three times (at regular intervals of 1 second), until the "3Com Corporation CS/2600 Series Monitor" and > command prompt, appears on the ASCII console.
- Using the new 3½" diskette, make a copy of the original diskette. At the > monitor prompt, enter:

  ```
  co <Enter>
  ```

  To list available commands, at the > monitor prompt, enter:

  ```
  ?
  ```

  Refer to instructions in the *CS/2600 Installation Guide,* as necessary.

  Store the original diskette in a safe place (from now on, you will use the copy).

- At the > monitor prompt, enter:

  ```
  fc <Enter>
  ```

  The "Firmware Configuration Utility Commands" menu is displayed.

  ```
          C   - Change parameters
          D   - Display parameters
          Esc - Exit to monitor.
  ```

  Select Change parameters, at the ? monitor prompt, enter:

  ```
  C
  ```

  This brings up the "Change parameters" menu. Modify the parameters as follows:

  ```
          5. Monitor              Disabled
          6. Initial boot source  Local Floppy
          A. Boot protocol        TFTP Boot
          B. Change IP/TFTP parameters
  ```

Select the menu **B** that appears when you select the TFTP Boot protocol. Modify the parameters as follows:

Addr discover protocol	**Local Information**
Client Ip address	CS2600 Ethernet address for example 1.0.0.10
Gateway address	for example 1.0.0.10 (if necessary, public network administrator)
Subnet mask value	subnet mask of administration network for example: 255.0.0.0

The “Addr discover protocol” is in accordance with the protocol in use at your site.

The “Client IP address” specifies the IP address you have chosen to assign to the concentrator.

The “Subnet mask value” information, specifies the subnet mask of the administration LAN network (dedicated or public).

- Press the Escape key several times to exit from the menus and to go back to the > monitor prompt.
- Perform a soft reset of the concentrator by pushing the reset button located on the front panel. Wait until the end of self–tests when the ”self test” LED and ”boot state” LED turn off (at least one minute).

  Wait another two minutes until the diskette drive LED turns off.

  Press the Return key several times, until the following prompt appears:

  ```
  Welcome to the 3Com Communication
  [1] CS>
  ```
- To perform further configuration, the “Network Manager” privilege level must be changed. Enter the following command:

  ```
          [1] CS> set pri = nm <Enter>
  ```

  The concentrator asks for a password. Press <Enter> (since initially no password is in effect).

  A new prompt appears, showing that you are at the Network Manager privilege level:

  ```
  [2] cs#
  ```
- From now on, you can configure the base parameters of the concentrator (date, time, system name, password...), as explained in the *CS/2600 Multiprotocol Communications Server* documentation. Once the settings are as desired, update the list of allowed services by entering the following command:

  ```
          cs# set cs = all <Enter>
  ```
- A new prompt appears:

  ```
  [3] cs#
  ```
- Declare again the Client IP address you have assigned (above) to the concentrator, for example:

  ```
          cs# setd -ip net = 1.0.0.10 <Enter>  (CS/2600 Ethernet address)
  ```

Configuring the Console Concentrator Ports

You must now configure the serial ports of the concentrator (J1, J2, J3...) that are connected to the S1 or COM1 console ports of the cluster nodes.

Note: When entering a command at the `cs#` prompt, to designate a concentrator port, you must specify an expression of the form **`!n`** where the number **`n`** represents the number of the port. For example, the `!2` expression designates the J2 serial port of the concentrator.

- The J0 port of the concentrator must be configured as a **terminal** port. On the other hand, the serial ports (J1, J2, J3...) that are connected to the S1 or COM1 ports of the cluster nodes or interconnect devices, must be defined as **host** ports.

  By default, all ports are configured as **terminal ports**. You have to change only the ports that are connected to the cluster nodes or interconnect devices.

  By default, the baud rate parameter is set to **autobaud**. Host ports must be configured to operate at 9600 bauds.

  As an example, the following two commands must be entered to configure the J1 port as a host port, with the appropriate baud rate:

  ```
  [3] cs# setd !1 -term dv = host <Enter>
  [4] cs# setd !1 -term baud = 9600 <Enter>
  [5] cs# setd !1 -term autd = <x>  <Enter>
  ```

  where `<x>` = 1 to 16,000 minutes. This sets the time of auto-disconnection when the port is no longer used.

  Repeat these commands for the other ports (!2, !3...) to complete the configuration.

- Assign an IP address to each serial port connected to a cluster node or ”switch”. For example, to set the IP address for the J1 port to the 1.0.0.11 value, enter:

  ```
  [7] cs# setd !1 -tcpappl porm = <@IP1> <Enter>
  ```

  Where <@IP1> is the IP address for port J1 connected to port S1 or COM1 of node #1, for example: 1.0.0.11.

  Repeat this step for each port connected to the S1 or COM1 console port of a node or interconnect devices.

Checking the Console Concentrator Settings

- To check the IP addresses you have assigned to the concentrator ports, enter:

  ```
  [8] cs# sh -tcpappl porm <Enter>
  ```

- To check the settings of the J1 host port, enter:

  ```
  [9] cs# sh !1 dp <Enter>
  ```

  Similarly, for the other ports (`!2`, `!3`...) whose settings are to be checked.

Checking the Console Concentrator Network Connection

To check that the concentrator is able to communicate through the network. invoke the concentrator **`ping`** command.

For example, use the IP address of the PowerConsole workstation (assuming it is currently up) as an argument of the `ping` command:

```
[10] cs# ping @IP_<name> <Enter>   (for example, ping 1.0.0.20)
```

If the network connection is OK, the ping command returns the following message:

```
pinging... 1.0.0.20 is alive
```

Checking the LISTEN State on Console Concentrator Ports

To check that the concentrator ports, connected to S1 on nodes, are in the LISTEN state, enter:

```
[6] cs# sh –term all <Enter>
```

Port 0 is in the ”command” state. Other ports, connected to S1, must be in the ”LISTEN” state. If this is not the case, disconnect the ports with the **logout** command:

```
[10] cs# logout !1 <Enter>
```

for port J1, for example.

Updating the /etc/hosts file on the PowerConsole

On the PowerConsole workstation, update the */etc/hosts* file by adding all the IP addresses you have defined to the concentrator. These include the IP address of the concentrator itself, as well as the different IP addresses you have assigned to the different host ports (J1, J2...) of the concentrator that are connected to the nodes.

For example, if you are setting up an HACMP cluster with 3 nodes (whose nodes are named for example `jazz`, `java`, and `bebop`), you have assigned IP addresses to the J1, J2, and J3 concentrator ports. In which case, you must add appropriate entries in the `/etc/hosts` file, as in the example below (adapt names and addresses to your requirements and preferences):

```
120.154.33.10 concentrator
120.154.33.54 EstrellaLink
120.154.33.51 jazz_console
120.154.33.52 java_console
120.154.33.53 bebop_console
```

Note: Depending on your requirements, you may want to update */etc/hosts* on other machines as well, or to update the name server in use at your site (if any).

Gaining Access to Node Console Ports

Once you have set up the PowerConsole components as explained in this chapter, the PowerConsole is ready to access the nodes through their console serial port (via the console concentrator). Thus, you are able to gain access to the consoles of the nodes for further configuration tasks.

To gain access to the console port of a node, invoke the **`telnet`** command with the desired IP address as its argument.

Using Telnet

- For example, to gain access to the console of the node which is connected to the J1 concentrator port, assuming that you have assigned the 120.154.33.51 IP address to the J1 port, enter the following command:

  ```
          telnet 120.154.33.51 <Enter>
  ```

  When this command is issued, you gain access to the system console of the node, as you would do from a conventional ASCII terminal directly attached to the S1 console port of the node.

- To disconnect a telnet session, simultaneously hit the **Ctrl** and **]** keys. Once the session is disconnected, you return to the `telnet>` prompt. From here, to quit telnet, type:

  ```
  q <Enter>
  ```

  or

  ```
  quit <Enter>
  ```

Setting Up and Using the pwcons Utility

pwcons Overview

The **pwcons** utility comes with the BullCluster software package. This utility can be useful for configurations equipped with a single terminal where consoles are handled through a multiport adapter (see sample configurations in “Serial Multiport Adapter-Based Console Solutions”, on page A-3, and in “Example 1”, on page A-7).

pwcons is a shell script, which is installed in the **/usr/sbin** directory, and which relies on configuration files that you must set up as explained in “Setting Up pwcons Configuration Files” below.

Usage Scenario and Comments

Assume that Node A is the node that has a terminal attached, and that the S1 ports of the other cluster nodes (Node B, Node C, etc.) are connected to a multiport adapter on Node A. Then, to gain access to consoles of other nodes:

1. From the console terminal of Node A, log in to Node A.
2. Once logged in, you can type a command having the following form:

   ```
   pwcons -c ttyx
   ```

   Where `ttyx` is the device name corresponding to the serial port (typically on the multiport adapter) which is connected to the S1 port of the node you want to attain.

   > *For example, the command* `pwcons -c tty4` *indicates you want to use the* `tty4` *device. If* `tty4` *corresponds to the port (on the multiport adapter) to which is connected the S1 port of Node C, then, this command connects you to Node C.*

   Instead of specifying a `ttyx` device name, you can specify a symbolic name (an arbitrary mnemonic string, typically a node name). This assumes, however, you have set up the **/etc/pwcons.conf** file to declare the correspondence between device names and symbolic names. This is explained below.
3. Assume you have entered `pwcons -c tty4` , and thus, you are connected to the S1 port of a particular node, say Node C. From now on, everything happens as if you were directly connected to the system console of Node C:
   - You see the messages destined to Node C console (although you have no means to see the messages that occurred before you connected to this console).
   - If Node C is running, its login banner is displayed, prompting you to enter your name and password. Thus, you can log in to Node C if desired.
4. Once you have finished your work on Node C, disconnect from its console:
   - If you were logged in, log out.
   - Enter the `~.` (tilde and dot) key sequence, then Return, to terminate the connection.
5. Once the connection is terminated, you revert to the original console (Node A). Then, if desired, you can again use the **pwcons** command to connect to another node.

   Note that, you can use **pwcons** only from the node that has the console attached (Node C in our example). You cannot, for example, gain control of the Node C console, then from here use **pwcons** to gain control of a third node console. You must revert to Node A before being able to establish connection with another node console.

See below for details on how to set up and use **pwcons**.

Setting Up pwcons Configuration Files

First configure TTY with the login capability disabled.
Enter:

```
smit maktty
```

The set of tty parameters is displayed.

1. Specify **disable** in the **Enable LOGIN** field.
2. All other fields remain with default parameters.
3. Press Enter to apply settings.

Setting Up /etc/uucp/Devices

- If you have ordered a Powercluster equipped with a multiport adapter, the serial ports (TTYs) of its multiport adapter have been configured at factory. In addition, the **/etc/uucp/Devices** file has also been set up at factory. However, you may want to read the discussion below in order to check that **/etc/uucp/Devices** is appropriately set up, or to understand how it works.
- For other configurations, refer to the discussion below to properly set up **/etc/uucp/Devices**.

/etc/uucp/Devices (which is fully described in the standard AIX documentation) contains information about the devices on the local system that can establish a connection to a remote computer using utilities such as **cu**. The **pwcons** shell script relies on **cu**, and thus on this configuration file.

The **/etc/uucp/Devices** file must contain a description of each TTY device that is used to connect to the S1 ports of other nodes. So, check that this file is set up as desired (create it if it does not exist, modify it if desired).

Each TTY description must be specified with a line in the format:

```
Direct ttyx – 9600 direct
```

where *ttyx* is a device name corresponding to a serial port (typically on the multiport adapter) which is connected to the S1 port of a node you want to attain.

Note that the file may include comments (comment lines begin with a #) and blank lines (which are ignored).

Example

Assume you are implementing a 4-node cluster, where Node A is a node equipped with a multiport adapter used to connect to the three other nodes S1 ports (as illustrated in Example 2, on page A-3). Assume that the multiport adapter is configured so that its ports that are connected to other nodes correspond to the device names `tty4`, `tty5` and `tty6`.

Then, **/etc/uucp/Devices** must include the following three lines:

```
Direct tty4 – 9600 direct
Direct tty5 – 9600 direct
Direct tty6 – 9600 direct
```

Setting Up /etc/pwcons.conf

This task is not mandatory. The **/etc/pwcons.conf** file is just a convenience that allows you to use symbolic names instead of device names (such as `tty4`) when connecting to a node console with **pwcons**.

The **/etc/pwcons.conf** file describes the correspondence between a symbolic name and a TTY device name that is used to connect to the S1 ports of other nodes. It must include a line for each correspondence you want to define (create the file if it does not exist, modify it if desired).

Each line has the following format:

```
mnemonic ttyx
```

Where `mnemonic` is any arbitrary name you want be interpreted by the **pwcons** command as synonymous for the `ttyx` device. Typically, you could specify a `mnemonic` corresponding to the name of the node that can be reached through *ttyx*.

Note that the file may include comments (comment lines must begin with a #).

Example

Assume you are implementing a 4-node cluster, where `alpha` is a node equipped with a multiport adapter used to connect to the three other nodes S1 ports. Assume that the multiport adapter is configured so that its `tty4`, `tty5` and `tty6` connect to the S1 ports of the nodes `beta`, `gamma` and `delta`.

If **/etc/pwcons.conf** does not exist or is not set up, you can connect to the `beta`'s console by using the **pwcons –c tty4** command. That works, but tty numbers are hard to remember.

Now, if you set up **/etc/pwcons.conf** as follows:

```
beta  tty4
gamma tty5
delta tty6
```

Then you can connect to the `beta`'s console by using the **pwcons –c beta**. This is more convenient, since node names are easier to remember than tty numbers.

pwcons Command Syntax

We assume here that you know the purpose of **pwcons**, which is explained above.

Syntax

/usr/sbin/**pwcons** [–c *ttyx* | *mnemonic*] | [–s *ttyx* | *mnemonic*] | [–S] |
[–d *ttyx* | *mnemonic*] | [–D]

Flags

[–c *ttyx* | *mnemonic*]

Connects to a specific tty, identified by either its `ttyx` device name or by its corresponding *mnemonic* as defined in the **/etc/pwcons.conf** file (described above).

When a connection is opened, a lock file is created to declare the tty as locked. The lock file mechanism is identical to the **cu**'s lock file (**cu** is a program of the uucp family, see your AIX documentation for details).

Once connected, you can terminate the connection at any moment, by entering the `~.` (tilde and dot) key sequence, then Return. (If you were logged in, log out before terminating the connection).

[**–s *ttyx* | *mnemonic***]

Shows the status of a specific tty. It is either opened or closed.

When a tty is closed, it is not locked, so you can connect to it.

When a tty is opened, it is locked. You cannot connect to a locked tty until you force a disconnection to unlock the tty (see `-d` and `-D` flags). If a tty is opened, the `-s` flag displays the process numbers of the commands that have opened it.

[–S] Shows the status of all the tty listed in /**etc**/**pwcons.conf**.

[–d *ttyx* | *mnemonic*]

Forces the disconnection of a specific tty, so that it becomes closed and unlocked. You may want to use this feature if you do not succeed in connecting to a tty because it is locked. Normally, a tty is locked only when someone is connected and uses this tty. However, it may also be locked because a communication session stopped abnormally, without properly closing the connection.

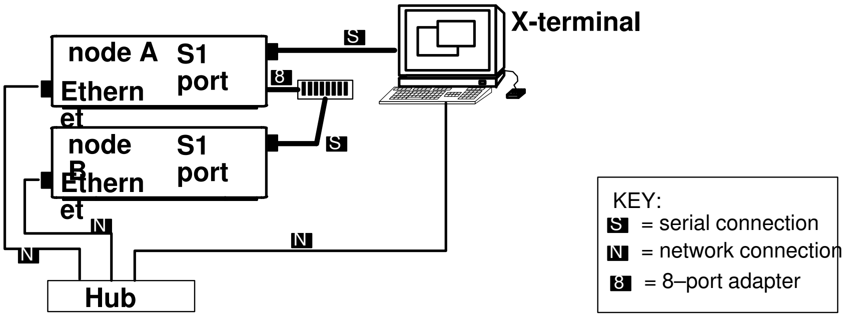
[–D] Forces the disconnection of all the tty's listed in **/etc/pwcons.conf**.

Appendix B. Implementing Cluster Console Solution

This appendix explains the installation of the X-terminal.

X-Terminal Installation Procedure

The X-terminal is connected as shown in the figure.



CAUTION:
Do not plug-in the power cables to the X-terminal box and to the monitor front side before being asked to do so.

1. Install the memory extension and the PCMCIA board which has been previously write-enabled

   See Section 5 of the ”*Installing Your Explora Family System*” documentation.

2. Connect the video cable between the X-terminal box and the monitor.
3. Connect the X-terminal to the Superstack II Hub 10 using a RJ45/RJ45 cable (CBF5410)
4. Connect the X-terminal box:

   to the console concentrator using RS232 cables (CBL 1912, CBLG106–2000),

   otherwise to the S1 plug of a Powercluster node using RS232 cables (CBL 1912, CBLG105–1800).

5. Connect the keyboard and the mouse on the X-terminal box.
6. Plug the power cable to the X-terminal box and on the monitor.
7. Power ON the X-terminal box.
8. Power ON the monitor (Green LED is swiched on). Refer to ”*17 Professional Color Monitor – User's Guide*” for further information on LEDs and command switches.
9. You can stop the automatic starting of the X-terminal by typing the ESC key after Power–up tests have completed.
10. Once the prompt > appears, type the command `se` and press ENTER to get the main menu.
11. Select « keyboard » in order to set the type of keyboard (IBM PS/2 or N101).
12. Select « Monitor » to set the resolution frequency of the monitor.

CAUTION:
Selecting a wrong screen resolution can damage your monitor. Make sure your selection is supported by your monitor. See *”17 Professional Color Monitor – User's Guide”***.**

13. Select « Network » then set the following parameters:

```
Get IP Address from:            NVRAM
Terminal IP Address:            @IP of XT
Subnet Mask:                    Subnet Mask
```

14. Select « Boot » then set the following characteristics as follows:

```
TFTP Boot Directory:    /local/
TFTP Order:             Disabled
NFS Order:              Disabled
MOP Order               Disabled
Local Order             1
```

15. Select « Done » then « Reboot ».

16. Once the prompt > is displayed, if the Boot is not automatic then type:

    >BL and press ENTER.

17. Two or three windows appear after starting has completed:

    a window of Setup and Configuration (upper left)

    a telnet window

    a system console window corresponding to the serial line RS232 (S1 plug of a Powercluster node) provided that the X-terminal is directely wired to a node's S1 plug.

18. Inside the Setup and Configuration window select « Window Manager » and run « NCD Window Manager » clicking on the associated icon.

19. In the telnet window, or after opening "New Telnet" in "TERMINAL" of the Setup window:

    type in the Service field the IP address of another node and hit OK to establish the telnet session

    Open as many New Telnet windows as they are left nodes: go to the Setup & Configuration window, select Terminal then New Telnet and type the node IP address in the service field.

20. In order to have a full automatic boot when powered up, select the SETUP menu in the window of Setup and Configuration then select:

    ```
    Change Setup Parameters
    ```

    ```
    Booting
    ```

    validate « Boot automatically at power-up.

    Save the configuration by choosing « Apply ».

21. In order to have an automatic establishment of additional telnet session to other nodes, select in the Setup window:

    ```
    SETUP
    ```

    ```
    Change Setup Parameter
    ```

    ```
    Commands and Startup
    ```

    then for each node add the IP address within the command line:

    ```
    term –ctype telnet <@IP> –geometry ...–title <node name>...–n<node abbr>...
    ```

    and save the configuration by hitting « Apply ».

22. At the next power–up put back the PCMCIA board in write unable.

Index

Vos remarques sur ce document / Technical publication remark form

Titre / Title : Bull ESCALA EPC Series EPC & HA Solutions Setup Guide

Nº Reférence / Reference Nº : 86 A2 79HX 05

Daté / Dated : September 1999

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 ELECTRONICS ANGERS
CEDOC
34 Rue du Nid de Pie – BP 428
49004 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 ELECTRONICS ANGERS
CEDOC
ATTN / MME DUMOULIN
34 Rue du Nid de Pie – BP 428
49004 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 # Nº Référence CEDOC	**Qty** Qté	**CEDOC Reference #** Nº Référence CEDOC	**Qty** Qté	**CEDOC Reference #** Nº 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 ELECTRONICS ANGERS
CEDOC
34 Rue du Nid de Pie – BP 428
49004 ANGERS CEDEX 01
FRANCE

ORDER REFERENCE
86 A2 79HX 05

Bull

Utiliser les marques de découpe pour obtenir les étiquettes.
Use the cut marks to get the labels.

ESCALA EPC Series
AIX
EPC & HA
Solutions
Setup Guide

86 A2 79HX 05

ESCALA EPC Series
AIX
EPC & HA
Solutions
Setup Guide

86 A2 79HX 05

ESCALA EPC Series
AIX
EPC & HA
Solutions
Setup Guide

86 A2 79HX 05