

VMS

digital

VMS VAXcluster Manual

VMS VAXcluster Manual

Order Number: AA-LA27B-TE

June 1990

This manual describes procedures for configuring and managing VAXcluster systems.

Revision/Update Information: This manual supersedes the *VMS VAXcluster Manual, Version 5.0*.

Software Version: VMS Version 5.4

June 1990

The information in this document is subject to change without notice and should not be construed as a commitment by Digital Equipment Corporation. Digital Equipment Corporation assumes no responsibility for any errors that may appear in this document.

The software described in this document is furnished under a license and may be used or copied only in accordance with the terms of such license.

No responsibility is assumed for the use or reliability of software on equipment that is not supplied by Digital Equipment Corporation or its affiliated companies.


Restricted Rights: Use, duplication, or disclosure by the U.S. Government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.227-7013.

© Digital Equipment Corporation 1990.

All Rights Reserved.
Printed in U.S.A.

The postpaid Reader's Comments forms at the end of this document request your critical evaluation to assist in preparing future documentation.

The following are trademarks of Digital Equipment Corporation:

CDA	DEQNA	MicroVAX	VAX RMS
DDIF	Desktop-VMS	PrintServer 40	VAXserver
DEC	DIGITAL	Q-bus	VAXstation
DECdtm	GIGI	ReGIS	VMS
DECnet	HSC	ULTRIX	VT
DECUS	LiveLink	UNIBUS	XUI
DECwindows	LN03	VAX	
DECwriter	MASSBUS	VAXcluster	

The following is a third-party trademark:

PostScript is a registered trademark of Adobe Systems Incorporated.

ZK4477

Production Note

This book was produced with the VAX DOCUMENT electronic publishing system, a software tool developed and sold by Digital. In this system, writers use an ASCII text editor to create source files containing text and English-like code; this code labels the structural elements of the document, such as chapters, paragraphs, and tables. The VAX DOCUMENT software, which runs on the VMS operating system, interprets the code to format the text, generate a table of contents and index, and paginate the entire document. Writers can print the document on the terminal or line printer, or they can use Digital-supported devices, such as the LN03 laser printer and PostScript printers (PrintServer 40 or LN03R ScriptPrinter), to produce a typeset-quality copy containing integrated graphics.

Contents

PREFACE	xiii
----------------	-------------

CHAPTER 1 INTRODUCTION TO VAXCLUSTER SYSTEMS	1-1
---	------------

1.1 SHARED RESOURCES	1-2
1.1.1 Disk Storage _____	1-2
1.1.2 Batch and Print Job Processing _____	1-2

1.2 INTERCONNECT DEVICES	1-3
---------------------------------	------------

1.3 SOFTWARE COMPONENTS	1-4
--------------------------------	------------

1.4 CONFIGURATION TYPES	1-5
1.4.1 CI-Based VAXcluster Systems _____	1-6
1.4.2 Local Area VAXcluster Systems _____	1-6
1.4.2.1 High-Availability Configurations with Multiple System Disks • 1-8	
1.4.2.2 High-Availability Dual-Host Configurations • 1-9	
1.4.3 Mixed-Interconnect VAXcluster Systems _____	1-10
1.4.4 Security for Local Area and Mixed-Interconnect VAXcluster Systems _____	1-11

1.5 CONNECTION MANAGEMENT	1-12
1.5.1 The Quorum Scheme _____	1-12
1.5.2 Quorum Disk _____	1-14
1.5.3 State Transitions _____	1-15

1.6 CONFIGURATION PLANNING	1-17
-----------------------------------	-------------

CHAPTER 2 PREPARING THE CLUSTER OPERATING ENVIRONMENT	2-1
--	------------

2.1 DIRECTORY STRUCTURE ON A COMMON SYSTEM DISK	2-2
--	------------

Contents

2.2	INSTALLING THE VMS OPERATING SYSTEM IN THE VAXcluster ENVIRONMENT	2-4
------------	--	------------

2.3	CONFIGURING AND STARTING THE DECNET-VAX NETWORK	2-6
2.3.1	Copying Remote Node Databases	2-8
2.3.2	Enabling VAXcluster Alias Operations	2-9

2.4	COORDINATING STARTUP COMMAND PROCEDURES	2-9
2.4.1	Building Startup Procedures for a Common-Environment Cluster	2-10
2.4.1.1	Procedures for Existing Computers • 2-10	
2.4.1.2	Procedures for Newly Installed Computers • 2-11	
2.4.2	Building Startup Procedures for a Multiple-Environment Cluster	2-11

2.5	COORDINATING SYSTEM FILES FOR A COMMON-ENVIRONMENT CLUSTER	2-12
2.5.1	Coordinating User Accounts	2-12
2.5.2	Preparing the Rights Database	2-13
2.5.3	Preparing the MAIL Database	2-14
2.5.4	Coordinating Shared System Files in Clusters with Multiple Common System Disks	2-15

CHAPTER 3	SETTING UP AND MANAGING CLUSTER DISKS	3-1
------------------	--	------------

3.1	CLUSTER-ACCESSIBLE DISKS	3-1
3.1.1	HSC Disks	3-2
3.1.2	MSCP-Served Disks	3-2
3.1.2.1	MSCP Server Functions • 3-3	
3.1.2.2	MSCP Load Sharing • 3-3	
3.1.3	Dual-Pathed Disks	3-4
3.1.3.1	Dual-Ported HSC Disks • 3-4	
3.1.3.2	Dual-Pathed DSA Disks on Local UDA/KDA/KDB Controllers • 3-5	
3.1.3.3	DSSI-Connected ISAs • 3-5	
3.1.3.4	Dual-Ported MASSBUS Disks • 3-6	

3.2	CLUSTER DEVICE-NAMING CONVENTIONS	3-7
3.2.1	Rules for Specifying Allocation Class Values	3-7
3.2.2	Sample Configurations with Named Devices	3-8

3.3	SHARED DISKS	3-11
3.4	CONFIGURING CLUSTER DISKS	3-12
3.5	REBUILDING CLUSTER DISKS	3-12

CHAPTER 4 SETTING UP AND MANAGING CLUSTER QUEUES 4-1

4.1	CLUSTERWIDE QUEUES	4-1
4.2	CLUSTER PRINTER QUEUES	4-2
4.2.1	Setting Up Printer Queues _____	4-3
4.2.2	Setting Up Clusterwide Generic Printer Queues _____	4-4
4.3	CLUSTER BATCH QUEUES	4-6
4.3.1	Setting Up Executor Batch Queues _____	4-7
4.3.2	Setting Up Clusterwide Generic Batch Queues _____	4-8
4.4	USING A COMMON COMMAND PROCEDURE TO SET UP CLUSTER QUEUES	4-10

CHAPTER 5 BUILDING AND MAINTAINING THE CLUSTER 5-1

5.1	CLUSTER_CONFIG.COM FUNCTIONS	5-1
5.2	DETERMINING LOCATIONS AND SIZES FOR SATELLITE PAGE AND SWAP FILES	5-3
5.3	SELECTING BOOT AND DISK SERVERS	5-3
5.4	DETERMINING ALLOCATION CLASS VALUES IN MIXED-INTERCONNECT CLUSTERS	5-4

Contents

5.5	CONFIGURING THE CLUSTER	5-5
5.5.1	Adding a Computer to the Cluster	5-6
5.5.1.1	Updating Network Data After Adding a Satellite • 5-11	
5.5.1.2	Restoring a Satellite's Network Data • 5-12	
5.5.1.3	Controlling Clusterwide Broadcast Messages on Satellites and Boot Servers • 5-12	
5.5.2	Removing a Computer from the Cluster	5-13
5.5.3	Changing a Computer's Characteristics	5-14
5.5.4	Changing the Cluster Configuration Type	5-19
5.5.4.1	Changing an Existing CI-Only Cluster to a Mixed-Interconnect Configuration • 5-19	
5.5.4.2	Changing an Existing Local Area Cluster to a Mixed-Interconnect Configuration • 5-20	
5.5.5	Converting a Standalone Computer to a VAXcluster Computer	5-21
5.5.6	Creating a Duplicate System Disk	5-21

5.6	RECONFIGURING THE CLUSTER AFTER A MAJOR CHANGE	5-23
5.6.1	Updating MODPARAMS.DAT Files to Adjust Cluster Quorum	5-23
5.6.2	Shutting Down the Cluster	5-23
5.6.3	Changing Allocation Class Values on HSC Subsystems	5-24
5.6.4	Rebooting the Cluster	5-24

5.7	MAINTAINING THE CLUSTER	5-24
5.7.1	Running AUTOGEN with the FEEDBACK Option	5-25
5.7.2	Recording Configuration Data	5-25
5.7.3	Monitoring Ethernet Activity in Local Area and Mixed-Interconnect Clusters	5-26
5.7.4	Restoring Cluster Quorum After an Unexpected Computer Failure	5-26
5.7.5	Selecting Cluster Shutdown Options	5-28
5.7.5.1	The REMOVE_NODE Option • 5-28	
5.7.5.2	The CLUSTER_SHUTDOWN Option • 5-29	
5.7.5.3	The REBOOT_CHECK Option • 5-29	
5.7.5.4	The SAVE_FEEDBACK Option • 5-29	
5.7.6	Rebooting a Satellite with an Operating System on a Local Disk	5-29
5.7.7	Performing Security Functions in Local Area and Mixed-Interconnect Clusters	5-30
5.7.7.1	Maintaining Cluster Security Data • 5-31	
5.7.7.2	Controlling Conversational Bootstrap Operations for Satellites • 5-32	

5.8	GUIDELINES FOR CONFIGURING LARGE CLUSTERS	5-32
5.8.1	Configuring Disk Server Ethernet Adapters and Memory	5-33
5.8.2	Configuring System Disks	5-33
5.8.2.1	Concurrent User Activity • 5-33	
5.8.2.2	Concurrent Booting Activity • 5-34	
5.8.2.3	Boot Time Costs • 5-36	
5.8.2.4	Moving High-Activity Files off System Disks • 5-36	
5.8.2.5	Controlling Dump File Size and Creation • 5-36	
5.8.2.6	Sharing Dump Files • 5-37	
5.8.3	Adding Computers to an Existing Cluster	5-38
5.8.3.1	Running AUTOGEN with FEEDBACK for Initial Configuration • 5-38	
5.8.3.2	Creating a Command File to Run AUTOGEN with FEEDBACK • 5-39	
5.8.4	Setting Up a New Large VAXcluster System	5-40
5.8.5	Defining the VAXcluster Alias	5-41

APPENDIX A CLUSTER SYSGEN PARAMETERS **A-1**

APPENDIX B BUILDING A COMMON SYSUAF.DAT FILE **B-1**

APPENDIX C CLUSTER TROUBLESHOOTING INFORMATION **C-1**

C.1	DIAGNOSING FAILURES OF COMPUTERS TO BOOT OR TO JOIN THE CLUSTER	C-1
C.1.1	Summary of Events for Computers Booting and Joining the Cluster	C-1
C.1.2	CI-Connected Computer Fails to Boot	C-3
C.1.3	Satellite Fails to Boot	C-4
C.1.4	Computer Fails to Join the Cluster	C-6
C.1.5	Startup Procedures Fail to Complete	C-7
C.2	DIAGNOSING CLUSTER HANGS	C-7
C.2.1	Cluster Quorum Is Lost	C-8
C.2.2	A Shared Cluster Resource Is Inaccessible	C-8

Contents

C.3	DIAGNOSING CLUEXIT BUGCHECKS	C-9
<hr/>		
C.4	DIAGNOSING VAXPORT DEVICE PROBLEMS	C-9
C.4.1	VAXport Communication Mechanisms	C-10
C.4.2	Port Failures	C-11
C.4.2.1	Verifying CI Port Functions • C-12	
C.4.2.2	Verifying CI Cable Connections • C-13	
C.4.2.3	Repairing CI Cables • C-16	
C.4.3	Analyzing Error Log Entries for VAXport Devices	C-16
C.4.3.1	Error Log Entry Formats • C-17	
C.4.3.2	Device-Attention Entries • C-17	
C.4.3.3	Logged-Message Entries • C-20	
C.4.3.4	Error Log Entry Descriptions • C-23	
C.4.4	OPA0 Error Messages	C-30

INDEX

EXAMPLES

2-1	Sample Interactive Network Configuration Session	2-7
4-1	Common Procedure to Set Up VAXcluster Queues	4-11
5-1	Sample Interactive CLUSTER_CONFIG.COM Session to Add a CI-Connected Computer as a Boot Server	5-7
5-2	Sample Interactive CLUSTER_CONFIG.COM Session to Add a Satellite with Local Page and Swap Files	5-9
5-3	Sample NETNODE_UPDATE.COM File	5-12
5-4	Sample Interactive CLUSTER_CONFIG.COM Session to Remove a Satellite with Local Page and Swap Files	5-13
5-5	Sample Interactive CLUSTER_CONFIG.COM Session to Enable the Local Computer as a Disk Server	5-16
5-6	Sample Interactive CLUSTER_CONFIG.COM Session to Change the Local Computer's ALLOCLASS Value	5-17
5-7	Sample Interactive CLUSTER_CONFIG.COM Session to Enable the Local Computer as a Boot Server	5-17
5-8	Sample Interactive CLUSTER_CONFIG.COM Session to Change a Satellite's Hardware Address	5-18
5-9	Sample Interactive CLUSTER_CONFIG.COM Session to Convert a Standalone Computer to a Cluster Boot Server	5-21
5-10	Sample Interactive CLUSTER_CONFIG.COM CREATE Session	5-22
5-11	Sample SYSMAN Session to Change the Cluster Password	5-32
C-1	CI Device-Attention Entry	C-17

C-2	Ethernet Device-Attention Entry _____	C-19
C-3	CI Logged-Message Entry _____	C-21

FIGURES

1-1	Typical CI-Based VAXcluster Configuration _____	1-6
1-2	Local Area VAXcluster System with Single Boot Server _____	1-8
1-3	High-Availability Local Area VAXcluster Configuration _____	1-9
1-4	Dual-Host VAXcluster Configuration _____	1-10
1-5	Typical Mixed-Interconnect VAXcluster System _____	1-11
2-1	Directory Structure on Common System Disk _____	2-2
2-2	File Search Order on Common System Disk _____	2-3
3-1	CI-Based Configuration with Shared Disks _____	3-2
3-2	Mixed-Interconnect VAXcluster Segment with Dual-Pathed HSC Disk _____	3-9
3-3	Mixed-Interconnect VAXcluster Segment with Dual-Pathed DSA Disk _____	3-9
3-4	Device Names in a Mixed-Interconnect Cluster _____	3-10
4-1	Sample Printer Configuration _____	4-3
4-2	Printer Queue Configuration _____	4-4
4-3	Clusterwide Generic Printer Queue Configuration _____	4-6
4-4	Sample Batch Queue Configuration _____	4-7
4-5	Clusterwide Generic Batch Queue Configuration _____	4-9
C-1	A Correctly Connected Two-Computer CI Cluster _____	C-13
C-2	Crossed CI Cable Pair _____	C-14

TABLES

2-1	Information Requested for CI-based Configurations _____	2-5
2-2	Information Requested for Local Area and Mixed-Interconnect Configurations _____	2-6
3-1	Specifying Values for MSCP_LOAD and MSCP_SERVE_ALL Parameters _____	3-3
5-1	Summary of CLUSTER_CONFIG.COM Functions _____	5-2
5-2	Data Requested by CLUSTER_CONFIG.COM _____	5-5
5-3	CLUSTER_CONFIG.COM CHANGE Options _____	5-14
5-4	Summary of SYSMAN CONFIGURATION Commands for Cluster Authorization _____	5-31
5-5	System Disk I/O Activity and Boot Time for Single Satellite _____	5-35
5-6	System Disk I/O Activity and Boot Times for Multiple Satellites _____	5-35

Contents

5-7	AUTOGEN Dump File Symbols _____	5-37
A-1	Cluster SYSGEN Parameters _____	A-1

Preface

Intended Audience

This document addresses persons responsible for setting up and managing VAXcluster systems. To use the document as a guide to cluster management, you must have a thorough understanding of VMS system management concepts and procedures, as described in the *Introduction to VMS System Management*, the *Guide to Setting Up a VMS System*, and the *Guide to Maintaining a VMS System*.

Document Structure

The *VMS VAXcluster Manual* contains five chapters and three appendixes.

Chapter 1 describes the VAXcluster environment.

Chapter 2 explains how to prepare the cluster operating environment.

Chapter 3 discusses disk management concepts and procedures.

Chapter 4 discusses queue management concepts and procedures.

Chapter 5 explains how to build a VAXcluster system once the necessary preparations are made, and how to reconfigure and maintain the cluster.

Appendix A lists and defines cluster SYSGEN parameters.

Appendix B provides guidelines for building a cluster common user authorization file.

Appendix C provides VAXcluster troubleshooting information.

Associated Documents

This document is not a one-volume reference manual. The VMS utilities and commands discussed are described in detail in separate VMS Utility Reference Manuals and in the *VMS DCL Dictionary*.

For additional information on the topics covered in this manual, refer to the following documents:

- *Introduction to VMS System Management*
- *Guide to Setting Up a VMS System*
- *Guide to Maintaining a VMS System*
- *Guide to VMS File Applications*
- *VMS Networking Manual*
- *VMS Volume Shadowing Manual*
- VMS Utility Reference Manuals

Preface

Conventions

The following conventions are used in this manual:

Ctrl/x	A sequence such as Ctrl/x indicates that you must hold down the key labeled Ctrl while you press another key or a pointing device button.
Return	In examples, a key name is shown enclosed in a box to indicate that you press a key on the keyboard. (In text, a key name is not enclosed in a box.)
...	In examples, a horizontal ellipsis indicates one of the following possibilities: <ul style="list-style-type: none">• Additional optional arguments in a statement have been omitted.• The preceding item or items can be repeated one or more times.• Additional parameters, values, or other information can be entered.
.	A vertical ellipsis indicates the omission of items from a code example or command format; the items are omitted because they are not important to the topic being discussed.
()	In format descriptions, parentheses indicate that, if you choose more than one option, you must enclose the choices in parentheses.
[]	In format descriptions, brackets indicate that whatever is enclosed within the brackets is optional; you can select none, one, or all of the choices. (Brackets are not, however, optional in the syntax of a directory name in a file specification or in the syntax of a substring specification in an assignment statement.)
red ink	Red ink indicates information that you must enter from the keyboard or a screen object that you must choose or click on. For online versions of the book, user input is shown in bold .
boldface text	Boldface text represents the introduction of a new term or the name of an argument, an attribute, or a reason. Boldface text is also used to show user input in online versions of the book.
<i>italic text</i>	Italic text represents information that can vary in system messages (for example, Internal error <i>number</i>).
UPPERCASE TEXT	Uppercase letters indicate that you must enter a command (for example, enter OPEN/READ), or they indicate the name of a routine, the name of a file, the name of a file protection code, or the abbreviation for a system privilege.

numbers

Hyphens in coding examples indicate that additional arguments to the request are provided on the line that follows.

Unless otherwise noted, all numbers in the text are assumed to be decimal. Nondecimal radices—binary, octal, or hexadecimal—are explicitly indicated.

Introduction to VAXcluster Systems

A VAXcluster system is a highly integrated organization of VAX computers. As members of a VAXcluster system, these computers can share processing resources, data storage, and queues under a single VMS security and management domain, and they can boot or fail independently.

Using procedures described in Chapter 2, system managers can tailor the cluster operating environment to create a **common-environment** or a **multiple-environment** VAXcluster system.

- In a common-environment VAXcluster system, the same resources are available on all computers. User accounts are identical, the same known images are installed, the same logical names are defined, and mass storage devices and queues are shared.
- In a multiple-environment VAXcluster system, a group of computers shares one set of resources, while another group shares a different set. Alternatively, an individual computer can perform a specialized function using restricted resources, while other computers perform general timesharing work.

Although most cluster resources can be shared, user processes and memory are computer specific. When a process is created on a VAXcluster computer, the process must complete on that computer, using local memory. If the computer fails before the process completes, the process is terminated. However, users can recover from such a failure more quickly than on a standalone computer, because they need not wait until the computer is rebooted. Typically, they can log in on another VAXcluster computer to create a new process and continue working, provided that the resources required by the process (such as images and global sections) are available on that computer.

This chapter describes VAXcluster operating features and components, including the following:

- Shared resources
- Interconnect devices
- Software components
- Configuration types
- Connection management
- Configuration planning

Be sure you understand these topics before you attempt to perform any cluster setup operations.

Introduction to VAXcluster Systems

1.1 Shared Resources

1.1 Shared Resources

In any VAXcluster system, users can share computing, disk storage, and batch and print job processing resources. The ability to share resources facilitates workload balancing, because work can be distributed across the cluster. To keep pace with user demand, resources can be added without disrupting normal cluster operation.

1.1.1 Disk Storage

A major advantage of VAXcluster systems is the ability to make disk storage resources accessible to all VAXcluster computers. Storage devices such as DIGITAL Storage Architecture (DSA) disks, RF-series integrated storage assemblies (ISAs), and ESE20 solid state disks can be configured for local or clusterwide access. A **cluster-accessible** disk can be used by any active computer in the cluster that successfully mounts it. A disk that is not cluster accessible can be accessed only by the local computer.

Cluster-accessible disks offer the following advantages:

- More efficient use of mass storage, because more than one computer can use the same disk.
- Access by users to their default work disks when logging in to any computer on which the disks are accessible.
- Clusterwide file sharing. Because computers can share common versions of files, updates to a file are made only once to a single copy of the file.
- Implementation of clusterwide job-controller queues. Batch and print jobs can be processed on any computer that has access to the disks.

Some VAXcluster systems include hierarchical storage controller (HSC) subsystems. These are self-contained, intelligent mass storage subsystems that enable VAXcluster computers to share DSA disks and, in the VAXcluster configurations described in Section 1.4.1, DSA tapes. Because the HSC subsystem is an intelligent controller, it optimizes physical disk and tape operations and supports many combinations of standard disk interfaces (SDIs) and standard tape interfaces (STIs), which connect disks and tapes. HSC disk configurations provide flexibility, expansion potential, online maintenance and backup capability, and the capability for using controller-based (HSC-based) volume shadowing. (For detailed information on volume shadowing, see the *VMS Volume Shadowing Manual*.)

Procedures for setting up and managing cluster disks are described in Chapter 3.

1.1.2 Batch and Print Job Processing

System managers control how jobs share batch processing and printer resources by setting up and maintaining clusterwide generic queues. The strategy for setting up and managing these queues determines how well workloads are matched to available resources.

Introduction to VAXcluster Systems

1.1 Shared Resources

All clusterwide queues are controlled by a single job-controller queue file (JBCSYSQUE.DAT), which must be accessible to the computers participating in the clusterwide queue scheme. This file makes queues available across the cluster and enables jobs to execute on any queue from any computer, provided that the necessary mass storage volumes can be accessed by the computer on which the job executes.

Procedures for setting up and managing cluster queues are described in Chapter 4.

1.2 Interconnect Devices

Interconnect devices used to configure a VAXcluster system include the following:

- **Computer Interconnect (CI).** This high-speed, dual-path interface links computers and HSC subsystems in a computer room environment. A Computer Interconnect consists of several components such as CI port controllers (adapters), the star coupler, star coupler expander (CISCE), and the high-bandwidth CI cables themselves.
- **CI port controllers.** Port controllers like the CI780, CIBCA, CIBCI, and CIXCD (CI to XMI adapter) are microcoded, intelligent adapters that connect computers to CI cables. Each interface connects to one pair of transmitter and one pair of receiver cables.

Under normal operating conditions, both pairs of cables are available to meet traffic demands. If one path fails, all traffic uses the remaining path. The VMS operating system periodically tests a failed path. As soon as a failed path is restored, it is automatically used for normal traffic.

- **Star coupler and star coupler expander (CISCE).** The star coupler and star coupler expander provide a common connection point for CI-connected computers and HSC subsystems. Both coupler devices connect all CI cables from computers and HSC subsystems, creating a radial or “star” arrangement that has a maximum radius of 45 meters. These devices support the physical connection and disconnection of individual computers and HSC subsystems without affecting other computers or HSC subsystems.

The star coupler and CISCE are dual-pathed devices that contain separate components for each path. The star coupler is a passive device; the CISCE consists of redundant amplifiers. Both devices are designed so that all CI cables are transformer coupled and independent of earth ground reference. These attributes help to ensure signal integrity.

- **DIGITAL Storage System Interconnect (DSSI).** Like the Computer Interconnect, the DSSI bus permits multiple computers to communicate directly with storage devices. The DSSI bus connects as many as 6 ISAs between DSSI controllers.

Introduction to VAXcluster Systems

1.2 Interconnect Devices

- **Ethernet.** The Ethernet is a bus that uses digital baseband signaling. The Ethernet is used both for DECnet-VAX transmissions and, in some VAXcluster systems, for cluster communications. The Ethernet must be configured according to requirements specified in the VAXcluster *Software Product Description* (SPD).

Depending on how a VAXcluster system is configured, VAXcluster software can use multiple interconnects for cluster communications.

1.3 Software Components

The software components used to implement VAXcluster communication and resource-sharing functions always run on each computer in the cluster. Thus, if one computer fails, the VAXcluster system continues operating, because the components still run on the remaining computers. These software components are as follows:

- **System Communications Services (SCS)** software implements intercomputer communication, according to the DIGITAL System Communications Architecture (SCA).
- **VAXport drivers** (for example, PADRIVER and PEDRIVER) control the communication paths between local and remote ports.
- The **connection manager** dynamically defines the VAXcluster system and coordinates participation of computers in the cluster. The connection manager uses SCS to provide an acknowledged message delivery service for higher VMS software layers. The connection manager also maintains cluster integrity when computers join or leave the cluster—that is, when cluster **state transitions** occur. (State transitions are discussed in Section 1.5.3.)
- The **distributed file system** allows all computers to share mass storage, whether the storage device is connected to an HSC subsystem or to a computer. A local disk can be made available to the entire cluster. All cluster-accessible disks appear as if they are local to every computer.

The distributed file system and VMS Record Management Services (VMS RMS) provide the same access to disks and files across the cluster that is provided on a standalone computer. VMS RMS files may be shared to the record level.

- The **distributed lock manager** is used for synchronization functions by the distributed file system, job controller, device allocation, and other cluster facilities. It is available to users for developing cluster applications. The distributed lock manager implements the \$ENQ and \$DEQ system services to provide clusterwide synchronization of access to resources by allowing the locking and unlocking of resource names. (For detailed information on system services, refer to the *VMS System Services Volume*.) The distributed lock manager also provides a queueing mechanism so that processes can be put into a wait state until a particular resource is available. As a result, cooperating processes can synchronize their access to shared objects such as files and records.

Introduction to VAXcluster Systems

1.3 Software Components

If a VAXcluster computer fails, all locks that it holds are released. This mechanism allows processing to continue on the remaining computers. The distributed lock manager also supports clusterwide deadlock detection.

- The **distributed job controller** makes queues available across the cluster. VAXcluster computers can share batch and print queues. Users can submit jobs to any queue in the cluster, provided that the necessary mass storage volumes and peripheral devices are accessible to the computer on which the job executes. System managers can also set up generic batch queues that distribute batch processing workloads among computers. For detailed information on VAXcluster queues, see Chapter 4.
- The **mass storage control protocol (MSCP) server** implements the MSCP protocol, which is used to communicate with a controller for local MASSBUS or UNIBUS disks, or for DSA disks, such as RA-series disks. In conjunction with one or both of the disk class drivers (DUDRIVER, DSDRIVER), the MSCP server implements this protocol on a computer, allowing the computer to function as a storage controller. The computer submits I/O requests to locally accessed disks, such as UNIBUS, MASSBUS, and UNIBUS Disk Adapter (UDA) disks, and accepts the I/O requests from any computer in the cluster. In this way, the MSCP server makes locally connected disks available across the cluster. In the local area and mixed-interconnect VAXcluster systems described in Section 1.4.2 and Section 1.4.3, the MSCP server can also make HSC disks accessible over the Ethernet.

In addition to these components, all VAXcluster systems require **DECnet-VAX software**, which ensures that system managers can access all VAXcluster computers from a single terminal, even if terminal switching facilities are unavailable.

In local area and mixed-interconnect VAXcluster systems, DECnet-VAX software is required both for system management functions and cluster communications, such as remote booting operations.

In these systems, DECnet and SCS software coexist on the same extended Ethernet local area network (LAN). They share the same data link and physical link protocols, which are implemented by the Ethernet data link drivers, the Ethernet adapters, and the Ethernet itself.

1.4 Configuration Types

While processing needs and available hardware resources must determine how individual VAXcluster systems are configured, sites can choose from the following configuration types:

- CI-based VAXcluster systems
- Local area (Ethernet-based) VAXcluster systems
- Mixed-interconnect VAXcluster systems

Introduction to VAXcluster Systems

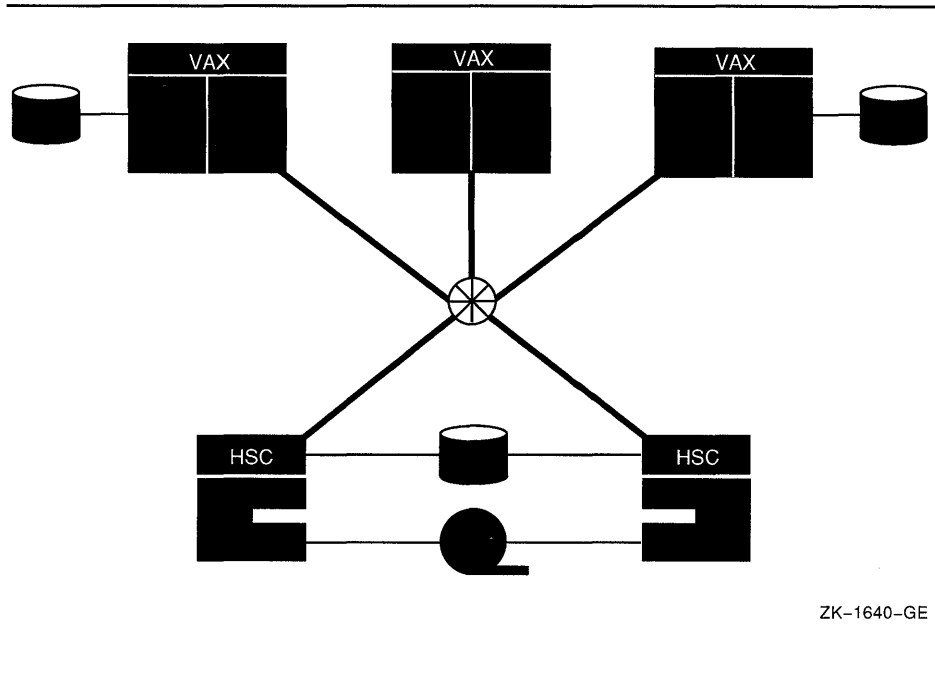
1.4 Configuration Types

These configuration types are described in Section 1.4.1, Section 1.4.2, and Section 1.4.3, respectively. For complete information on currently supported configurations, refer to the VAXcluster SPD.

1.4.1 CI-Based VAXcluster Systems

A CI-based VAXcluster system uses CI components for cluster communications, with a star coupler or CISCE as a common connection point for computers and HSC subsystems. Figure 1-1 shows how the components are typically configured. Note that if you want to add workstations to a CI-based VAXcluster system, you must convert it to a mixed-interconnect system. Refer to Section 5.5.4 for instructions.

Figure 1-1 Typical CI-Based VAXcluster Configuration



ZK-1640-GE

1.4.2 Local Area VAXcluster Systems

In a local area VAXcluster system, cluster communications are carried out over the Ethernet by a VAXport driver that emulates certain CI port functions. Because HSC subsystems require CI connections, local area VAXcluster systems do not include HSC subsystems.

A single extended Ethernet LAN can support multiple local area VAXcluster systems, each system identified and secured by a unique **group number** and a **cluster password**. (For information on cluster security, see Section 1.4.4.)

Introduction to VAXcluster Systems

1.4 Configuration Types

Computers in a local area cluster are generally configured either as **servers** (boot servers and disk servers) or as **satellites** (computers without a local VMS system disk). Using MSCP server software, the servers make their locally connected disks available to satellites over the Ethernet.

Boot servers are disk servers that downline load the VMS operating system to satellites by means of DECnet Maintenance Operation Protocol (MOP). When a satellite requests an operating system load, a boot server sends an image to the satellite that allows the satellite to load the VMS operating system and join the cluster. Because a boot server *must* serve its system disk to the cluster, a boot server always runs MSCP server software.

Typically, a boot server is both a management center for the cluster and a major resource provider. Its system disk contains the cluster common files for startup, authorization, and queue setup, as well as the root directories from which the satellites are booted and in which their specific system files reside. These root directories, one for each satellite, are created when system managers add satellites to the cluster using the CLUSTER_CONFIG.COM command procedure described in Chapter 5.

Boot and disk servers make user and application data disks available across the cluster. These servers should be the most powerful computers in the cluster and should use the highest bandwidth Ethernet adapters in the cluster.

Satellites are booted remotely from a boot server's system disk. Generally, satellites are consumers of cluster resources, though they may also sometimes provide disk-serving and batch-processing facilities. If satellites are equipped with local disks, they may enhance performance by using such local disks for paging and swapping.

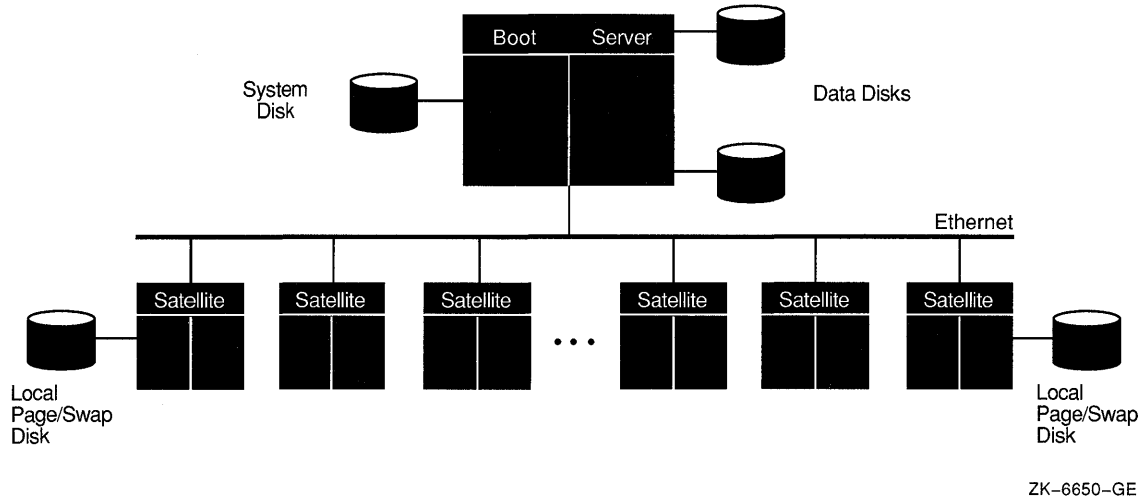
Any local area VAXcluster system can be converted to a mixed-interconnect system. Refer to Section 5.5.4 for instructions.

Figure 1-2 shows a local area VAXcluster system with a single boot server. Note that because all computers in this configuration rely on the boot server's system disk, the boot server (or its system disk) is a single point of failure.

Introduction to VAXcluster Systems

1.4 Configuration Types

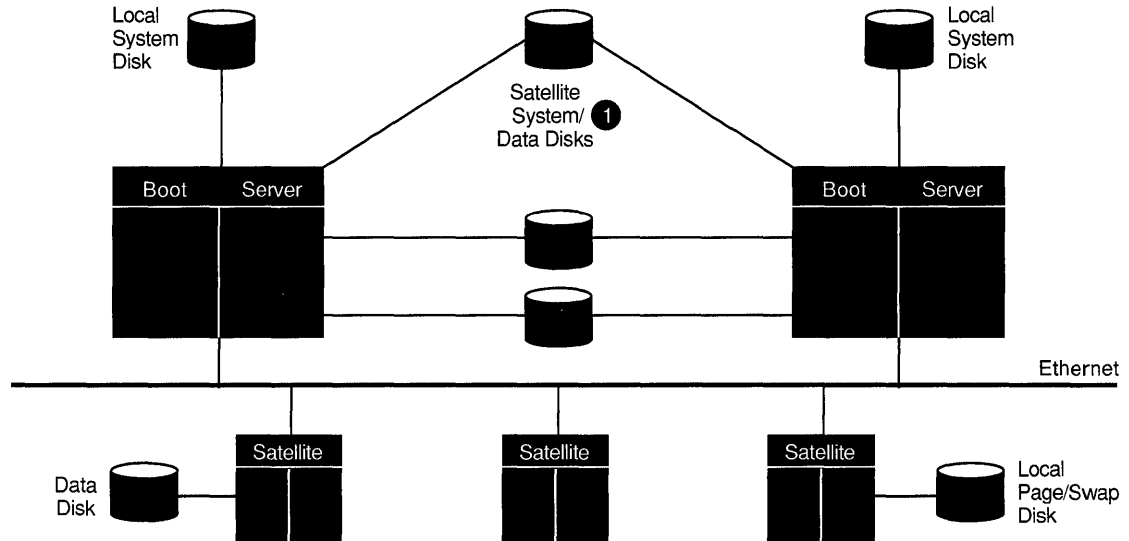
Figure 1-2 Local Area VAXcluster System with Single Boot Server



1.4.2.1 High-Availability Configurations with Multiple System Disks

You can configure a high-availability local area VAXcluster system using two boot servers with locally connected system disks. A satellite system disk can be dual ported between the servers. Figure 1-3 shows such a configuration. The boot servers boot from their locally connected system disks, and the satellites boot from the satellite system disk, which contains their root directories. Thus, if one boot server fails, satellites can access their system disk through the other server. If the satellite system disk fails, it can be restored from the remaining boot server, and the satellites can then be rebooted.

Figure 1-3 High-Availability Local Area VAXcluster Configuration



1 System Disk for Booting Satellites

ZK-1877A-GE

1.4.2.2 High-Availability Dual-Host Configurations

Compared with configurations using two boot servers with local system disks (Figure 1-3), dual-host VAXcluster configurations using Integrated Storage Assembly (ISA) devices and the DSSI bus offer greater flexibility, growth potential, and ease of system management. Although dual-host configurations do not provide some of the features of CI-based VAXcluster systems (such as controller-based volume shadowing capabilities), they do provide high availability at the low end without the need for multiple system disks.

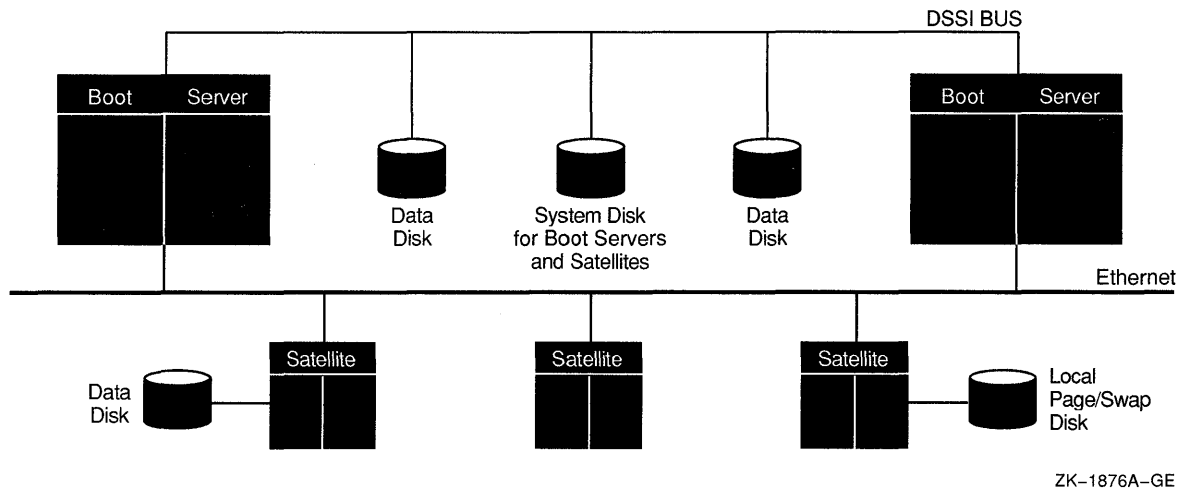
Both boot servers in a dual-host VAXcluster configuration can access a common system disk and all data disks directly and serve them to satellites. Satellites (and users connected through terminal servers) can access any disk through either boot server. If one of the boot servers fails, applications on satellites continue to run because disk access fails over to the other server. Although applications running on nonintelligent devices, such as terminal servers, are interrupted, users of terminals can log in again and restart their jobs.

In the dual-host configuration shown in Figure 1-4, the two boot servers and all satellites boot from a common system disk.

Introduction to VAXcluster Systems

1.4 Configuration Types

Figure 1-4 Dual-Host VAXcluster Configuration



In addition to the two boot servers, a dual-host VAXcluster configuration typically includes several DSSI-connected ISAs and a TK70 tape serving the configuration. Using additional DSSI cables, you can connect either one or two **storage expansion boxes** containing additional ISAs. If you decide to use a storage expansion box, it is a good idea to place a common system disk and critical data disks in the expansion box, which has a dedicated power supply. Thus, if one boot server fails, the other server and satellites can still access the disks.

Note that there are specific rules for various dual-host configurations; that is, the setup for a dual-host MicroVAX 3800 system is different from that for a dual-host VAX 4000 system. These rules are provided in the appropriate installation and user manuals.

1.4.3 Mixed-Interconnect VAXcluster Systems

A mixed-interconnect VAXcluster system can include VAX computers, MicroVAX-class satellites and workstations, and HSC subsystems. Because MSCP server software and disk class drivers allow CI-connected computers to serve HSC disks to satellites, the satellites can access the large amount of storage that is available through HSC subsystems.

Mixed-interconnect VAXcluster systems combine the following advantages of CI-based and local area VAXcluster systems:

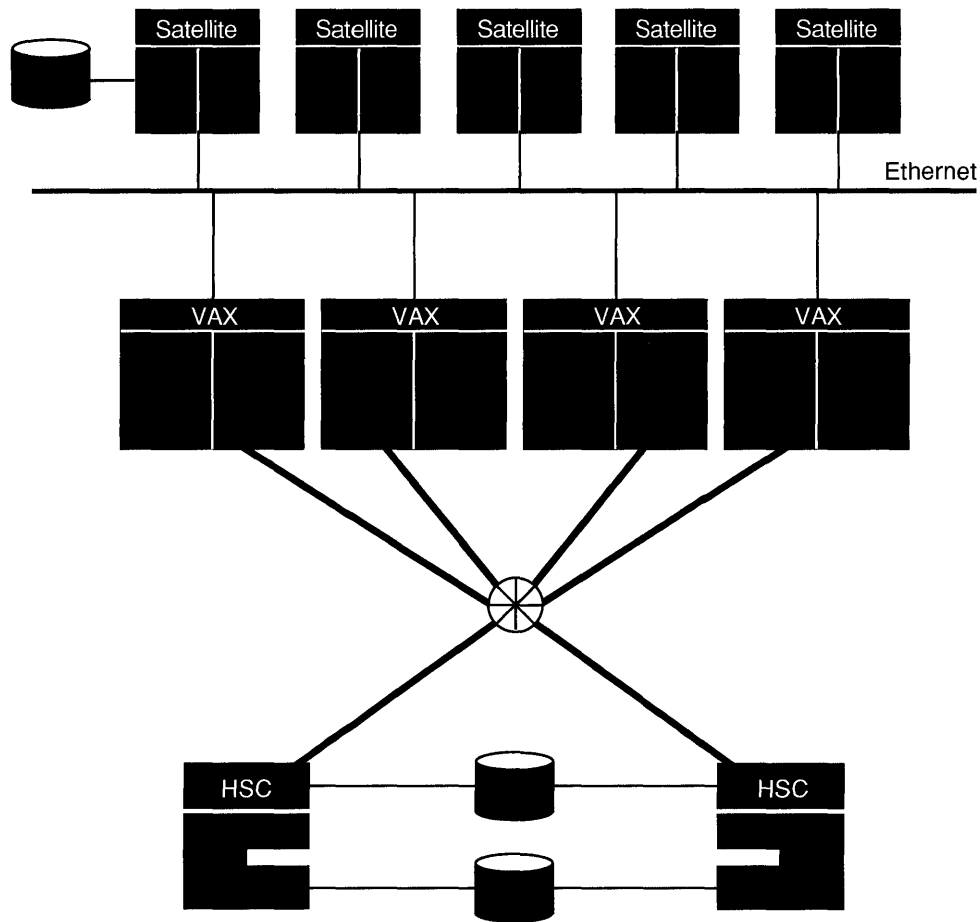
- Use of HSC subsystems for mass storage
- Support for MicroVAX-class computers and workstations
- High availability of system resources
- Centralized cluster management

Introduction to VAXcluster Systems

1.4 Configuration Types

Figure 1-5 shows a typical mixed-interconnect VAXcluster system.

Figure 1-5 Typical Mixed-Interconnect VAXcluster System



ZK-6659-GE

1.4.4 Security for Local Area and Mixed-Interconnect VAXcluster Systems

Local area and mixed-interconnect VAXcluster systems use a **group number** and a **cluster password** to allow multiple independent VAXcluster systems to coexist on the same extended Ethernet LAN and to prevent access to a cluster by unauthorized computers.

Introduction to VAXcluster Systems

1.4 Configuration Types

- The group number uniquely identifies each mixed-interconnect and local area VAXcluster system on an extended Ethernet LAN. This number must be in the range from 1 to 4095 or from 61440 to 65535. Note that if you plan to have more than one of these clusters at your site, you must coordinate the assignment of group numbers among system managers.
- The cluster password serves as an additional check to ensure the integrity of individual clusters on the same extended Ethernet LAN that accidentally use identical group numbers. (If each cluster's password is unique, the clusters will form independently.) The password also prevents an intruder who discovers the group number from joining the cluster. The password must be from 1 to 31 alphanumeric characters in length, including dollar signs (\$) and underscores (_).

Security data is maintained in the cluster authorization file, `SY$COMMON:[SYSEXE]CLUSTER_AUTHORIZE.DAT`. This file is created during installation of the VMS operating system, if you indicate that you want to set up a local area or mixed-interconnect cluster. The installation procedure then prompts you for the cluster group number and password. Cluster security functions are described in detail in Chapter 5. (If you convert a CI-based cluster to a mixed-interconnect configuration, the file is created when you execute the `CLUSTER_CONFIG.COM` command procedure, as described in Chapter 5.)

1.5 Connection Management

The integrity of a VAXcluster system is controlled by a software component called the **connection manager**, which determines and coordinates membership in the cluster. The connection manager creates a cluster when the first active computers are booted and then reconfigures the cluster when computers join or leave it.

VAXcluster computers can share various data and system resources, such as disk volumes. To achieve the coordination that is necessary to maintain resource integrity, the computers must share a clear sense of cluster membership, which is maintained by the connection manager.

Within a single VAXcluster system, the VMS operating system guarantees the integrity of shared resources by carefully coordinating their use. However, because use of shared resources is not coordinated between computers in separate clusters—a condition known as **cluster partitioning**—the connection manager prevents this condition using a scheme called **quorum**.

1.5.1 The Quorum Scheme

The quorum scheme is based on the arithmetic principle that the whole cannot be divided into multiple parts in such a way that more than one part is greater than half of the whole. (Integer arithmetic is used in this section.)

Introduction to VAXcluster Systems

1.5 Connection Management

The quorum scheme functions as follows:

- In a VAXcluster system, each **voting member** (a computer with a nonzero value for the SYSGEN parameter VOTES) contributes a fixed number of votes toward quorum. On satellites, the VOTES value is always set to zero by default.
- Each active computer in the cluster (including satellites) indirectly specifies an initial quorum value using the SYSGEN parameter EXPECTED_VOTES. This parameter is the sum of all votes held by potential cluster members. It is used to derive an estimate of the correct quorum value for the cluster, according to the following formula:

$$\text{Estimated quorum} = (\text{EXPECTED_VOTES} + 2) / 2$$

- During certain cluster state transitions, the system dynamically computes the cluster quorum to be the *maximum* of the following:
 - The current cluster quorum value
 - The largest of the values calculated from the following formula, where EV is the EXPECTED_VOTES value specified by each computer:

$$(EV+2) / 2$$

- The value calculated from the following formula, where V is the total of the SYSGEN parameter VOTES held by all cluster members:

$$(V+2) / 2$$

The cluster state transitions that cause cluster quorum to be recalculated occur when a computer joins the cluster and when the cluster recognizes a quorum disk. (The role of the quorum disk is discussed in Section 1.5.2.)

- If the current number of votes ever drops below quorum (because of computers leaving the cluster), the remaining cluster members suspend all process activity and all I/O operations to cluster-accessible disks until sufficient votes are added (computers joining the cluster) to bring the total number of votes to a value greater than or equal to quorum.
- As the cluster configuration changes, a computer only raises the cluster quorum value; it never lowers the value. (However, system managers can lower the value; for details, see Section 5.7.4.)

For example, consider a cluster consisting of three computers, each computer having its VOTES parameter set to 1 and its EXPECTED_VOTES parameter set to 3. The connection manager dynamically computes the cluster quorum value to be 2. In this example, any two of the three computers constitute a quorum and can run in the absence of the third computer. No single computer can constitute a quorum by itself. Therefore, there is no way the three VAXcluster computers can be partitioned and run as two independent clusters.

Introduction to VAXcluster Systems

1.5 Connection Management

1.5.2 Quorum Disk

A quorum disk acts as a virtual computer, adding to the total cluster votes. By establishing a quorum disk in configurations with a small number of voting computers, you can increase the availability of the cluster. Such configurations can tolerate the failure either of the quorum disk or of a computer and continue operating.

To use a quorum disk, one or more computers must have a direct (non-MSCP-served) connection to the disk. Such computers are known as **quorum disk watchers**. Computers that cannot access the disk directly rely on the quorum disk watchers for information about the status of votes contributed by the quorum disk.

You should enable as quorum disk watchers any computers that have an active direct connection to the quorum disk or that have the potential for a direct connection. To enable a computer as a quorum disk watcher, use the CLUSTER_CONFIG.COM CHANGE function described in Section 5.5.3. The procedure prompts for the name of the quorum disk and specifies that name as a value for the SYSGEN parameter DISK_QUORUM in MODPARAMS.DAT. The procedure also sets an appropriate value for the QDKSVOTES parameter. The number of votes contributed by the quorum disk is equal to the smallest value of the SYSGEN parameter QDKSVOTES on any quorum disk watcher.

Note: You can also enable the first installed cluster computer as a quorum disk watcher by answering YES when the VMS installation procedure asks whether the cluster will contain a quorum disk.

For the quorum disk's votes to be counted in the cluster votes total, the following conditions must be met:

- On one or more computers capable of becoming watchers, you must specify the same *physical* device name as a value for the SYSGEN parameter DISK_QUORUM. The remaining computers (which must have a blank value for DISK_QUORUM) recognize the name specified by the first quorum disk watcher with which they communicate. A *VAXcluster system can include only one quorum disk.*
- At least one quorum disk watcher must have a direct, active connection to the quorum disk. Thus, the quorum disk may be a dual-ported DSA disk, which has an active direct connection to only one computer at a time.
- The disk must contain a valid format file named QUORUM.DAT in the master file directory (MFD). The QUORUM.DAT file is created automatically after a system specifying a quorum disk has booted into the cluster. This file is used on subsequent reboots. If no quorum disk is enabled when a computer boots, the file is not created on that computer.
- To permit recovery from failure conditions, the quorum disk must be mounted by all disk watchers.

1.5.3 State Transitions

VAXcluster state transitions occur when a computer joins or leaves a VAXcluster system. The connection manager controls these events to ensure the preservation of data integrity throughout the cluster. A state transition's duration and effect on users (applications) are determined by the reason for the transition, the configuration, and the applications in use.

Every transition goes through one or more phases, depending on whether its cause is the addition of a new VAXcluster member or the failure of a current member. If the transition is caused by the addition of a new member, the phases are as follows:

- New member detection

Early in its boot sequence, a computer seeking membership in a VAXcluster system sends messages to current members asking to join the cluster. The first cluster member that receives the membership request acts as the new computer's advocate and proposes reconfiguring the cluster to include the computer in the cluster. While the new computer is booting, no applications are affected.

- Reconfiguration

All current VAXcluster members must establish communications with the new computer. Once communications are established, the new computer is admitted to the cluster. In some cases, the lock database is rebuilt.

If the transition is caused by the failure of a current VAXcluster member, the phases are as follows:

- Failure detection

The duration of this phase depends on the cause of the failure and on how the failure is detected.

During normal cluster operation, messages sent from one computer to another are acknowledged when received. If a message is not acknowledged within a period determined by VAXcluster communications software, the repair attempt phase begins.

If a cluster member is shut down or crashes, the VMS operating system causes datagrams to be sent from the computer shutting down to the other members. These datagrams state the computer's intention to sever communications and stop sharing resources. Because sending these datagrams is virtually the last activity of a dying computer, they are called "last gasp" datagrams. If any current cluster member receives a last gasp datagram, the "gasping" computer is removed from the cluster. The failure detection and repair attempt phases are bypassed, and the reconfiguration phase begins immediately.

Introduction to VAXcluster Systems

1.5 Connection Management

- Repair attempt

If the communication path to a VAXcluster member is broken, attempts are made to repair the path. Repair attempts continue for an interval specified by the SYSGEN parameter RECNXINTERVAL. (System managers can adjust the value of this parameter to suit local conditions.) Thereafter, the path is considered irrevocably broken, and steps must be taken to reconfigure the VAXcluster system so that all computers can once again communicate with each other, and so that any computers that cannot communicate are removed from the cluster.

- Reconfiguration

When a VAXcluster member fails, the cluster must be reconfigured. One of the remaining computers acts as coordinator and exchanges messages with all other cluster members to determine the configuration of an **optimal subcluster** with the most members and the most votes. This phase, during which all user (application) activity is blocked, usually lasts less than 1 second.

- VAXcluster system recovery

Recovery includes the following stages, some of which can take place in parallel:

- I/O completion

When a computer is removed from the cluster, VAXcluster software ensures that all I/O operations that are started by the old configuration complete before I/O operations that are generated by the new configuration start. There is usually little or no effect on applications.

- Lock database rebuild

Because the lock database is distributed among all members, some portion of the database may need rebuilding. A rebuild is always performed when a computer leaves the cluster, but only in certain cases when a computer is added.

- Disk mount verification

This stage occurs only when the failure of a voting member causes quorum to be lost. To protect data integrity, all I/O activity is blocked until quorum is regained. Mount verification is the mechanism used for this purpose.

- Quorum votes validation

If, when a computer is removed, the remaining members can determine that it has shut down or crashed, the votes contributed by the quorum disk are included without delay in quorum calculations that are performed by the remaining members. However, if they cannot determine that the computer has shut down or crashed (for example, if a console halt, power failure, or communications failure has occurred), the votes are not included for a period equal to four times the value of QDSKINTERVAL seconds. This period is sufficient to determine that the failed computer is no longer using the quorum disk.

— Disk rebuild

If the transition is the result of a computer rebooting after a crash, the disks are marked as improperly dismounted, and they must be rebuilt before they can be remounted. The rebuild reclaims space that was cached by the failed computer but never returned, as with a normal dismount. A rebuild makes the disk briefly inaccessible to users unless the rebuild is deferred by using the /NOREBUILD qualifier to the MOUNT commands in the system startup files. (See Section 3.5 for more information on rebuilding disks.)

- Application recovery

When assessing the effect of a state transition on application users, consider that the application recovery phase includes activities such as replaying a journal file, cleaning up recovery units, and users logging in again because the terminal server has failed over to another computer.

1.6 Configuration Planning

The process of setting up a VAXcluster system requires careful preparation. In planning your configuration, you must determine the following:

- *Configuration type.* If you want to include MicroVAX-class computers or workstations in your VAXcluster system, you must set up a local area or mixed-interconnect configuration. These configurations are described in Section 1.4.2 and Section 1.4.3.
- *Operating environment (common or multiple).* These environments are described at the beginning of this chapter and in Chapter 2, which provides information on configuring the DECnet-VAX network and on preparing the startup, user authorization, and other files that define the operating environment.
- *Disk storage configuration.* Chapter 3 provides information about disk storage configurations, including descriptions of disk types, rules for specifying disk names in VAXcluster systems, and sample disk configurations.
- *Queue configuration.* Chapter 4 provides information about VAXcluster queues and includes a sample queue setup command procedure.
- *Computer configuration.* Procedures for configuring VAXcluster computers are described in Chapter 5. That chapter also includes a detailed discussion of the cluster configuration command procedure, `SYSS$MANAGER:CLUSTER_CONFIG.COM`.

Once you have planned your configuration, installed the necessary hardware, and checked hardware devices for proper operation, you can set up the cluster using various system software facilities. Setup procedures are typically as follows:

- Installing or upgrading the VMS operating system on the first VAXcluster computer. Follow instructions in the installation and operations guide for your computer.

Introduction to VAXcluster Systems

1.6 Configuration Planning

- Installing required software licenses. Follow instructions in the *VMS License Management Utility Manual*.
- Configuring and starting the DECnet-VAX network. Follow instructions in Chapter 2. For more detailed information on network operations, refer to the *VMS Networking Manual*.
- Preparing files that define the cluster operating environment and that control disk and queue operations. Follow instructions in Chapters 2, 3, and 4.
- Adding computers to the cluster. Follow instructions in Chapter 5.

Depending on various factors, the order in which these operations are performed can vary from site to site.

2

Preparing the Cluster Operating Environment

By setting up appropriate startup and other system files, you can prepare the VAXcluster operating environment on the first installed computer before adding other computers to the cluster. Depending on your processing needs, you can prepare either a **common-environment** or a **multiple-environment** cluster.

In a common-environment cluster, the operating environment is identical on each VAXcluster computer because the computers are run from common system files. The computers are set up with identical user accounts, the same known images are installed, the same logical names are defined, and mass storage devices and queues are shared. In effect, users in a common-environment cluster can log in to any computer and work in the same operating environment.

In a multiple-environment cluster, the environment varies from computer to computer, and users can work in environments that are specific to the computer they are logged in to. A multiple-environment cluster is effective when you want to share data among computers but want certain computers to serve specialized needs. For example, you might want to set up a three-computer cluster, in which the timesharing environments on two computers are the same, while the third computer is set up exclusively for batch processing of large inventory jobs. In this case, the timesharing computers are set up with a common environment, sharing users, queues, and access to mass storage devices, while the third computer runs in its own restricted environment.

This chapter concentrates on the steps necessary to prepare a common-environment cluster. Approaches for preparing a multiple-environment cluster are also described, but are presented as general guidelines.

Topics include the following:

- Directory structure on a common system disk
- Installing the VMS operating system in the VAXcluster environment
- Configuring and starting the DECnet-VAX network
- Coordinating startup command procedures
- Coordinating system files for a common-environment cluster

Once you have prepared the cluster operating environment as described in this chapter and determined your disk and queue configurations using the information in Chapter 3 and Chapter 4, you can build the cluster following instructions in Chapter 5.

Preparing the Cluster Operating Environment

2.1 Directory Structure on a Common System Disk

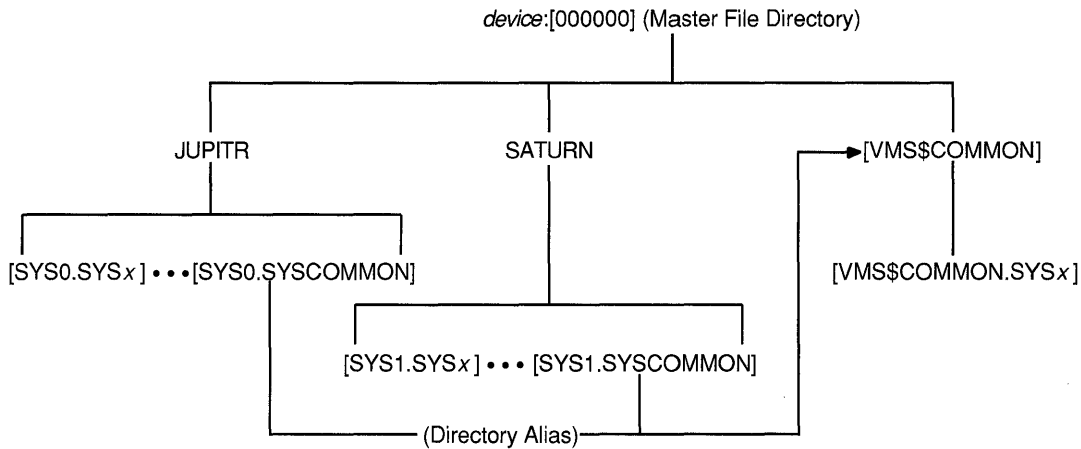
2.1 Directory Structure on a Common System Disk

The VMS installation or upgrade procedure generates a *common system disk*, on which most operating system and optional product files are stored in a common root directory. The entire directory structure—that is, the common root plus each computer's local root—is stored on the same disk. After the installation or upgrade completes, you use the CLUSTER_CONFIG.COM command procedure described in Chapter 5 to create a local root for each new computer and boot it into the cluster.

Each local root contains, in addition to the usual system directories, a [SYSx.SYSCOMMON] directory that is an alias for [VMS\$COMMON], the cluster common root directory in which cluster common files actually reside. When you add a computer to the cluster, CLUSTER_CONFIG.COM creates the alias.

Figure 2-1 illustrates the directory structure set up for computers JUPITR and SATURN, which are run from a common system disk. The disk's master file directory (MFD) contains the local roots (SYS0 for JUPITR, SYS1 for SATURN) and the cluster common root directory, [VMS\$COMMON].

Figure 2-1 Directory Structure on Common System Disk



SYS\$SPECIFIC = device:[SYSn.]

SYS\$COMMON = device:[SYS n.SYSCOMMON.]

SYS\$SYSROOT = device:[SYS n.], device:[SYS n.SYSCOMMON.]

Key: n = System Root

x = System Subdirectory

ZK-6658-GE

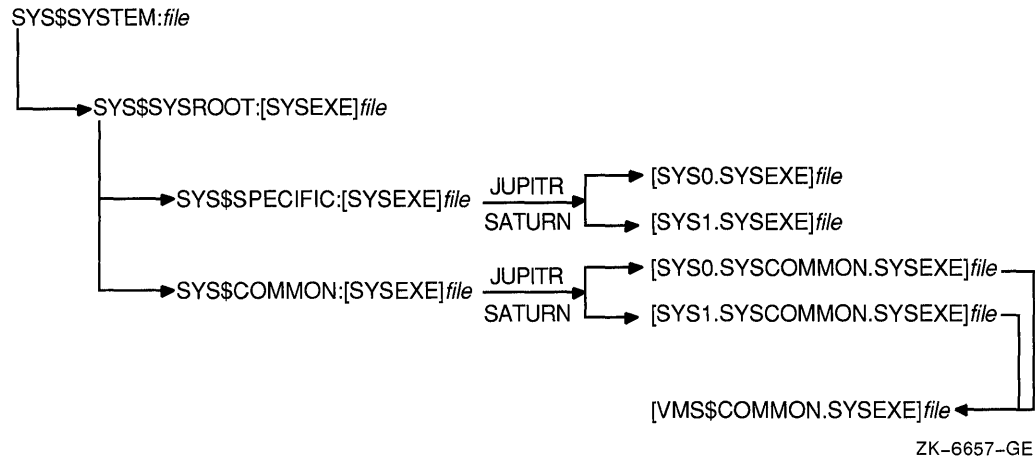
The logical name SYS\$SYSROOT is defined as a search list that points to a local root first (SYS\$SPECIFIC) and then to the common root (SYS\$COMMON). Thus, the logical names for the system directories (SYS\$SYSTEM, SYS\$LIBRARY, SYS\$MANAGER, and so forth) point to two directories: a local root (for example, SYS\$SPECIFIC:[SYSEXE]) and a common root (for example, SYS\$COMMON:[SYSEXE]). Figure 2-2

Preparing the Cluster Operating Environment

2.1 Directory Structure on a Common System Disk

shows how directories on a common system disk are searched when the logical name SYS\$SYSTEM is used in file specifications.

Figure 2-2 File Search Order on Common System Disk



It is important to keep this search order in mind when manipulating system files on a common system disk. Computer-specific files must always reside and be updated in the appropriate computer's system subdirectory. For example, MODPARAMS.DAT must reside in SYS\$SPECIFIC:[SYSEXE], which is [SYS0.SYSEXE] on JUPITR, and [SYS1.SYSEXE] on SATURN. Thus, to create a new MODPARAMS.DAT file for JUPITR when logged in on JUPITR, you would enter the following command:

```
$ EDIT SYS$SPECIFIC:[SYSEXE]MODPARAMS.DAT
```

Once the file is created, you could use the following command to modify it:

```
$ EDIT SYS$SYSTEM:MODPARAMS.DAT
```

Note that if a MODPARAMS.DAT file does not exist in JUPITR's SYS\$SPECIFIC:[SYSEXE] directory when you enter this command, but there is a MODPARAMS.DAT file in the directory SYS\$COMMON:[SYSEXE], the command edits the MODPARAMS.DAT file in the common directory. If there is no MODPARAMS.DAT file in either directory, the command creates the file in JUPITR's SYS\$SPECIFIC:[SYSEXE] directory.

To modify JUPITR's MODPARAMS.DAT when logged in on any other computer that boots from the same common system disk, you enter the following command:

```
$ EDIT [SYS0.SYSEXE]MODPARAMS.DAT
```

If you want to modify records in the cluster common system authorization file in a cluster with a single cluster common system disk, you could enter the following commands on any computer:

```
$ SET DEFAULT SYS$COMMON:[SYSEXE]
$ RUN AUTHORIZE
```

Preparing the Cluster Operating Environment

2.1 Directory Structure on a Common System Disk

But if, for example, you have set up a computer-specific system authorization file (SYSUAF.DAT) for computer JUPITR and you want to modify records in that file when logged in on another computer that boots from the same cluster common system disk, you must, before invoking AUTHORIZE, set your default directory to JUPITR's computer-specific [SYSEXEXE] directory. For example:

```
$ SET DEFAULT [SYS0.SYSEXEXE]
$ RUN AUTHORIZE
```

2.2 Installing the VMS Operating System in the VAXcluster Environment

You must perform the installation or upgrade once for each system disk in the cluster. Because, however, several computers normally run from the same cluster common system disk, you need not perform the installation or upgrade on each computer.

You may want to set up a cluster that has a combination of one or more common system disks and one or more individual system disks. Again, you must do the installation or upgrade once for each system disk. For example, if your cluster consists of 10 computers, 4 of which share one common system disk, 4 of which share a second common system disk, and each of the other 2 has its own system disk, you would do the installation or upgrade 4 times. Note that if your cluster includes multiple common system disks, *you must later coordinate system files to define the cluster operating environment, as described in Section 2.5.4.*

To perform the installation, follow instructions in the installation and operations guide for your computer. However, before you start the installation, be sure you have determined which *VAXcluster system configuration type* you want to create (CI-based, local area, or mixed-interconnect), because the installation procedure requests configuration-specific information. (Configuration types are described in Section 1.4.)

Table 2-1 lists the information requested for CI-based configurations; Table 2-2 lists the information requested for local area and mixed-interconnect configurations. Typical responses are explained in the tables. Note that initial questions are the same for all configuration types.

If your system disk is on an HSC subsystem, you must obtain the HSC subsystem's *disk allocation class* value before starting the installation, because the installation procedure requests that information. (Allocation classes are discussed in detail in Section 3.2.) To obtain the value, enter a command sequence like the following at the HSC console. The information displayed includes the allocation class value.

```
CTRL/C
HSC> SHOW SYS
15-Jun-1990 14:31:43.41  Boot:   13-Jun-1990 11:31:11.41  Up:    51:00
.
.
.
DISK allocation class = 1          TAPE allocation class = 0
Start command file m Disabled
SETSHO - Program Exit
```

Preparing the Cluster Operating Environment

2.2 Installing the VMS Operating System in the VAXcluster Environment

If you later want to change the allocation class value, follow the instructions in Section 5.6.

Note: While rebooting at the end of the installation procedure, the system displays messages warning that you must install VMS and VAXcluster software licenses. Be sure to install these licenses, as well as the DECnet-VAX license, as soon as the system is available. Procedures for installing licenses are described in the release notes distributed with the software kit and in the *VMS License Management Utility Manual*.

Table 2-1 Information Requested for CI-based Configurations

Item	Response
Will this node be a cluster member (Y/N)?	Enter Y.
What is the node's DECnet node name?	Enter DECnet node name—for example, JUPITR. The DECnet node name can be from 1 to 6 alphanumeric characters in length and cannot include dollar signs (\$) or underscores (_).
What is the node's DECnet node address?	Enter DECnet node address—for example, 2.2.
Will the Ethernet be used for cluster communications (Y/N)?	Enter N. The Ethernet is not used for cluster communications in CI-based VAXcluster systems.
Will JUPITR be a disk server (Y/N)?	Enter Y or N, depending on your configuration requirements. Refer to Section 1.4.3 and Chapter 3 for information on served cluster disks.
Enter a value for JUPITR's ALLOCLASS parameter:	If the system disk is connected to a dual-ported disk, enter a value from 1 to 255 that will be used on both sides. Otherwise, enter 0. (For detailed information on allocation classes see Section 3.2.)
Does this cluster contain a quorum disk [N]?	Enter Y or N, depending on your configuration. If you enter Y, the procedure prompts for the name of the quorum disk. Enter the device name of the quorum disk. (For detailed information on quorum disks, see Section 1.5.2.)

Preparing the Cluster Operating Environment

2.2 Installing the VMS Operating System in the VAXcluster Environment

Table 2–2 Information Requested for Local Area and Mixed-Interconnect Configurations

Item	Response
Will this node be a cluster member (Y/N)?	Enter Y.
What is the node's DECnet node name?	Enter DECnet node name—for example, JUPITR. The DECnet node name may be from 1 to 6 alphanumeric characters in length and cannot include dollar signs (\$) or underscores (_).
What is the node's DECnet node address?	Enter DECnet node address—for example, 2.2.
Will the Ethernet be used for cluster communications (Y/N)?	Enter Y. The Ethernet is required for cluster communications in local area and mixed-interconnect VAXcluster systems.
Enter this cluster's group number:	Enter a number in the range from 1 to 4095 or 61440 to 65535.
Enter this cluster's password:	Enter the cluster password. The password must be from 1 to 31 alphanumeric characters in length and can include dollar signs and underscores.
Reenter this cluster's password for verification:	Reenter the password.
Will JUPITR be a disk server (Y/N)?	Enter Y. In local area and mixed-interconnect configurations, the system disk is always served to the cluster. Refer to Section 1.4.3 and Chapter 3 for information on served cluster disks.
Will JUPITR serve HSC disks (Y/N)?	Enter a response appropriate for your configuration.
Enter a value for JUPITR's ALLOCLASS parameter:	If the system will serve HSC disks, enter the HSC's allocation class value. If the system disk is connected to a dual-ported disk, enter a value from 1 to 255 that will be used on both sides. Otherwise, enter 0. (For detailed information on allocation classes see Section 3.2.)
Does this cluster contain a quorum disk [N]?	Enter Y or N, depending on your configuration. If you enter Y, the procedure prompts for the name of the quorum disk. Enter the device name of the quorum disk. (For detailed information on quorum disks, see Section 1.5.2.)

2.3 Configuring and Starting the DECnet–VAX Network

After you have installed the operating system and required licenses on the first VAXcluster computer, you configure, tailor, and start the DECnet–VAX network. If you locate certain network files in the SYS\$COMMON:[SYSEXE] directory as described in step 3, other computers can share the data when they join the cluster. The process of configuring the network typically entails several operations:

Preparing the Cluster Operating Environment

2.3 Configuring and Starting the DECnet-VAX Network

- Executing the SYS\$MANAGER:NETCONFIG.COM command procedure.
- Making remote node data available clusterwide.
- Optionally defining a VAXcluster alias. You establish an alias using NCP commands like those shown in step 3 for alias SOLAR. (For more information on the VAXcluster alias, refer to the *VMS Networking Manual*.) Note that if you plan to define an alias, you must specify that one computer operate as a *router* node when you execute NETCONFIG.COM. Note further that you must later enable alias operations for other computers, as described in Section 2.3.2.
- Starting the network.

To perform these operations, proceed as follows:

- 1 Log in as system manager and execute NETCONFIG.COM. Enter information about your node when prompted and answer Yes when the procedure asks whether you want network configuration commands to be executed.

Note: When the procedure asks whether you want the network started, answer No if you first want to define a VAXcluster alias.

Example 2-1 shows a typical NETCONFIG.COM session.

Example 2-1 Sample Interactive Network Configuration Session

```
$ @NETCONFIG.COM

      DECnet-VAX network configuration procedure

This procedure will help you define the parameters needed to get DECnet
running on this machine.  You will be shown the changes before they are
executed, in case you want to perform them manually.

What do you want your DECnet node name to be? [JUPITR]: 
What do you want your DECnet address to be?   [2.2]: 
Do you want to operate as a router? [NO (nonrouting)]: YES
Do you want a default DECnet account?        [NO]: 

      Here are the commands necessary to set up your system.

      .
      .
      .

Do you want these commands to be executed? [YES]: 
      .
      .
      .

The changes have been made.
If you have not already registered the DECnet-VAX key, then do so now.
After the key has been registered, you should invoke the procedure
SYS$MANAGER:STARTNET.COM to start up DECnet-VAX with these changes.

(If the key is already registered) Do you want DECnet started? [YES] NO
$
```

Preparing the Cluster Operating Environment

2.3 Configuring and Starting the DECnet-VAX Network

- 2 NETCONFIG.COM creates, in the SYS\$SPECIFIC:[SYSEXE] directory, the permanent remote node database file NETNODE_REMOTE.DAT, in which remote node data is maintained. To make this data available clusterwide, you must rename the file to the SYS\$COMMON:[SYSEXE] directory:

```
$ RENAME SYS$SPECIFIC:[SYSEXE]NETNODE_REMOTE.DAT -  
_ $ SYS$COMMON:[SYSEXE]NETNODE_REMOTE.DAT
```

For information on sharing other network data, see the *VMS Networking Manual*.

- 3 If you want to define a VAXcluster alias, invoke the Network Control Program (NCP) Utility to do so. For example:

```
$ RUN SYS$SYSTEM:NCP  
NCP> DEFINE NODE 2.1 NAME SOLAR  
NCP> DEFINE EXECUTOR ALIAS NODE SOLAR  
NCP> EXIT  
$
```

The information you specify using these commands is entered in the DECnet-VAX permanent executor database and takes effect when you start the network.

- 4 Start the network:

```
$ @SYS$MANAGER:STARTNET.COM
```

- 5 To ensure that the network is started each time a VAXcluster computer boots, add the following line to the appropriate startup command file or files:

```
$ @SYS$MANAGER:STARTNET.COM
```

For information on preparing startup command files, see Section 2.4. For more detailed information on DECnet-VAX configuration issues and procedures, refer to the *VMS Networking Manual*.

2.3.1 Copying Remote Node Databases

Some sites with large networks maintain remote node data in a central database file. If this is the case at your site, and if you want to make the data available clusterwide, you can, *after starting the network*, copy remote node database entries from that central file. For example, if the file resides on node SATURN, you could enter the following NCP commands to copy entries from the permanent database on SATURN to the permanent database on your system disk, and then to update your volatile database:

```
NCP> COPY KNOWN NODES FROM SATURN USING PERMANENT TO PERMANENT  
NCP> SET KNOWN NODES ALL
```

Note that only node names and addresses are copied. See the *VMS Networking Manual* for more information on copying node databases.

Preparing the Cluster Operating Environment

2.3 Configuring and Starting the DECnet-VAX Network

2.3.2 Enabling VAXcluster Alias Operations

If you have defined a VAXcluster alias as described in Section 2.3, you can enable alias operations for other computers *after the computers are up and running in the cluster*. To enable such operations (that is, to allow a computer to accept incoming connect requests directed toward the alias), follow these steps:

- 1 Log in as system manager and invoke the SYSMAN Utility:

```
$ RUN SYS$SYSTEM:SYSMAN
```

- 2 At the SYSMAN> prompt, enter the following commands:

```
SYSMAN> SET ENVIRONMENT/CLUSTER
%SYSMAN-I-ENV, current command environment:
      Clusterwide on local cluster
      Username LAZARUS      will be used on nonlocal nodes
SYSMAN> SET PROFILE/PRIVILEGES=(OPER,SYSPRV)
SYSMAN> DO MCR NCP SET EXECUTOR STATE OFF
%SYSMAN-I-OUTPUT, command execution on node X...
.
.
.
SYSMAN> DO MCR NCP DEFINE EXECUTOR ALIAS INCOMING ENABLED
%SYSMAN-I-OUTPUT, command execution on node X...
.
.
.
SYSMAN> DO @SYS$MANAGER:STARTNET.COM
%SYSMAN-I-OUTPUT, command execution on node X...
.
.
.
```

2.4 Coordinating Startup Command Procedures

You must coordinate your site-specific SYSTARTUP and SYLOGIN command procedures according to the type of cluster operating environment you want to prepare. For a common-environment cluster, these procedures should perform the same system startup and login functions for each computer. For a multiple-environment cluster, you may want some startup commands to remain specific to certain computers, as described in Section 2.4.2.

In a common-environment cluster, you can prepare SYSTARTUP procedures using one of the following methods:

- In each computer's SYS\$SPECIFIC:[SYSMGR] directory, set up a SYSTARTUP_V5.COM procedure that performs computer-specific startup functions and then invokes a common SYSTARTUP procedure, typically named SYSTARTUP_COMMON.COM. This procedure is usually located in the SYS\$COMMON:[SYSMGR] directory on a common system disk but can reside on any disk, provided that the disk is cluster accessible and is mounted when the procedure is invoked.

Preparing the Cluster Operating Environment

2.4 Coordinating Startup Command Procedures

- After setting up computer-specific SYSTARTUP_V5.COM procedures, create a copy of SYSTARTUP_COMMON.COM for each computer. However, if you use multiple SYSTARTUP_COMMON.COM files, you must update all copies whenever you make changes.

To set up a common SYLOGIN procedure, define the logical name SYS\$SYLOGIN on each computer to be the full file specification of the procedure. If the common SYLOGIN file is on a cluster-accessible disk, include the command that defines SYS\$SYLOGIN in your common SYSTARTUP command file. If the computers use separate duplicate copies of SYLOGIN.COM, include the definition in each computer's specific startup procedure.

For example, the following command defines SYS\$SYLOGIN to be the common file [SYSMGR]SYLOGIN.COM on the cluster-accessible disk WORK5:

```
$ DEFINE/SYSTEM/EXEC SYS$SYLOGIN WORK5:[SYSMGR]SYLOGIN
```

Certain startup functions, even in a common-environment cluster, are computer specific. Therefore, you must include commands in the computer-specific startup procedure on each computer to do the following:

- Set up dual-ported and local disks
- Load device drivers
- Set up terminals
- Invoke the common SYSTARTUP command procedure

Section 2.4.1 and Section 2.4.2 present guidelines for using common and computer-specific command procedures to build a cluster environment.

2.4.1 Building Startup Procedures for a Common-Environment Cluster

The first step in preparing a common-environment cluster is to build cluster common SYSTARTUP and SYLOGIN command procedures. In a common-environment cluster, each computer executes the common procedures at startup time to define the same operating environment. Because each computer is set up with the common procedure, users can work in the same operating environment on any VAXcluster computer.

2.4.1.1 Procedures for Existing Computers

To build procedures for a cluster in which existing computers are to be combined in a VAXcluster system, you should compare both the computer-specific SYSTARTUP and SYLOGIN command procedures on each computer and make any adjustments required. For example, you can compare the procedures from each computer and include commands that define the same logical names in your common SYSTARTUP command file.

An easy method of comparing the existing procedures and creating common versions is to log in to each computer (in the single-computer environment) and print the existing SYSTARTUP and SYLOGIN command procedure files. You can then use the file listings to compare

Preparing the Cluster Operating Environment

2.4 Coordinating Startup Command Procedures

the procedures. After you have chosen which commands to make common, you can build the common procedures on one of the VAXcluster computers.

2.4.1.2 Procedures for Newly Installed Computers

The strategy for clusters being formed from newly installed VMS systems is basically the same as that used for clusters that are to include previously installed systems: include common elements in a common command procedure file (for example, SYSTARTUP_COMMON.COM). With newly installed systems, however, the SYSTARTUP and SYLOGIN command procedure files are empty. Therefore, you must build the common procedures from scratch.

For example, you could build a common startup command procedure named SYSTARTUP_COMMON.COM and include the commands that you want to be common to all computers. You must decide which of the following elements you want to include in the common procedure:

- Commands that install images.
- Commands that define logical names; for example, the logical name that refers to the location of SYLOGIN.COM.
- Commands that set up queues. (See Chapter 4 for information on setting up cluster queues.)
- Commands that set up and mount physically accessible mass storage devices. (See Chapter 3 for information on setting up cluster disks.)
- Commands that perform any other common startup functions. (See the *Guide to Setting Up a VMS System* for more information on startup command procedures.)

To build a common SYLOGIN.COM command file, include in the file commands that define clusterwide logical names and symbols.

You can include commands that mount cluster-accessible storage devices and set up queues in the common SYSTARTUP procedure or in separate command files (such as MSCPMOUNT.COM and STARTQ.COM) that are invoked by the common procedure. However, because such commands are computer specific, they must be executed by the local computer. Therefore, you must use conditional logic to control their execution. Sample command files for mounting storage devices and setting up queues are described in Chapter 3 and Chapter 4, respectively.

2.4.2 Building Startup Procedures for a Multiple-Environment Cluster

To build SYSTARTUP and SYLOGIN command files for a multiple-environment cluster, include in the files elements that you want to remain unique to a computer, such as commands to define computer-specific logical names and symbols. These files must be placed in the SYS\$SPECIFIC root on each computer.

For example, consider a three-member cluster consisting of computers JUPITR, SATURN, and PLUTO. The timesharing environments on JUPITR and SATURN are the same. However, PLUTO runs applications for a specific user group. In this cluster, you would create common

Preparing the Cluster Operating Environment

2.4 Coordinating Startup Command Procedures

SYSTARTUP and SYLOGIN command procedures for JUPITER and SATURN that define identical environments on these computers. But the command procedures for PLUTO would be different; they would include commands to define PLUTO's special application environment.

2.5 Coordinating System Files for a Common-Environment Cluster

To prepare a common VAXcluster user environment, you must coordinate the following system files:

- SYSUAF.DAT
- NETPROXY.DAT
- RIGHTSLIST.DAT
- VMSMAIL_PROFILE.DATA
- JBCSYSQUE.DAT
- NETNODE_REMOTE.DAT

These files, which are part of the VMS operating system, control such functions as user logins, proxy login access, mail, and access to files and job queues. By coordinating these files, you can define either a common-environment or a multiple-environment cluster.

In a common-environment cluster, you use a common version of each system file and place the files in the SYS\$COMMON:[SYSEXE] directory on a common system disk. (In contrast, in a multiple-environment cluster, you would use computer-specific versions of the files and place the files in each computer's SYS\$SPECIFIC:[SYSEXE] directory.)

Section 2.5.1 describes procedures for coordinating user accounts in common SYSUAF.DAT and NETPROXY.DAT files. Section 2.5.2 and Section 2.5.3 describe procedures for preparing the cluster RIGHTSLIST and VMSMAIL_PROFILE database files, respectively. For detailed information on the job-controller queue file, JBCSYSQUE.DAT, refer to Chapter 4. The NETNODE_REMOTE.DAT file is described in Section 2.3.

Note: If you want to set up a common-environment cluster with more than one common system disk, you must coordinate files on each disk and ensure that the disks are mounted with each cluster reboot. Refer to Section 2.5.4 for instructions.

2.5.1 Coordinating User Accounts

In a common-environment cluster, you must coordinate the user accounts from each computer and build common versions of the following files:

- SYSUAF.DAT
- NETPROXY.DAT

Preparing the Cluster Operating Environment

2.5 Coordinating System Files for a Common-Environment Cluster

If you are setting up a common-environment cluster that consists of newly installed systems, you can follow the instructions in the *Guide to Setting Up a VMS System* to build these files. Because the SYSUAF.DAT file on new VMS systems is empty except for the four Digital-supplied accounts, very little coordination is necessary.

However, if the cluster will include one or more computers that have been running with computer-specific SYSUAF.DAT and NETPROXY.DAT files, you must create common versions of the files. Procedures for creating a common SYSUAF.DAT file from computer-specific files are described in Appendix B.

Procedures for creating a common NETPROXY.DAT file are basically the same as those for creating a common SYSUAF.DAT file, except that less coordination is needed when you merge the individual NETPROXY.DAT files. For example, user identification codes (UICs) are not used in the NETPROXY records and therefore need not be coordinated. You should decide which existing proxy login records you want to keep and include these records in the common NETPROXY.DAT file.

Once you have prepared SYSUAF.DAT and NETPROXY.DAT files, you can set up each of them either as a common file on a cluster-accessible disk or as separate duplicate files. Note, however, that if you choose to use duplicate files, you must update all copies whenever you make changes.

If your cluster is running from one common system disk, make sure that SYSUAF.DAT and NETPROXY.DAT are located in the directory SYS\$COMMON:[SYSEXE].

If your cluster is running from any other system disk configuration, you must decide where to locate SYSUAF.DAT and NETPROXY.DAT. Once you have placed these two files in a directory, you must define clusterwide logical names to point to them.

Assume that disk WORK5 is shared by all computers in the cluster and that it contains cluster common SYSUAF.DAT and NETPROXY.DAT files. The following commands define system logical names that point to the location of the common files:

```
$ DEFINE/SYSTEM/EXEC SYSUAF WORK5:[SYSEXE]SYSUAF
$ DEFINE/SYSTEM/EXEC NETPROXY WORK5:[SYSEXE]NETPROXY
```

You must add the DEFINE commands to your common SYSTARTUP command file. After you have copied the files to the appropriate directory on the cluster-accessible disk, you should delete these files from the system disk.

2.5.2 Preparing the Rights Database

The rights database file, RIGHTS.LIST.DAT, associates users of the system or cluster with special names called **identifiers**. This file is the basis of a VMS protection scheme that uses access control lists (ACLs). For a detailed description of this scheme, see the *Guide to VMS System Security*. For information about how the rights database is created, refer to the *VMS Authorize Utility Manual*.

Preparing the Cluster Operating Environment

2.5 Coordinating System Files for a Common-Environment Cluster

The cluster manager or security manager maintains the rights database, adding and removing identifiers as needed. By allowing groups of users to hold identifiers, the manager can create a different kind of group designation than the one based on UICs. This alternative grouping allows the holders of the identifier to make more efficient use of resources. It also permits each user to be a member of multiple overlapping groups.

If your cluster is running from one common system disk, the installation or upgrade procedure places the RIGHTS.LIST.DAT file in the directory SYS\$COMMON:[SYSEXE]. No further action is required on your part.

If your cluster is running from any other system disk configuration, copy SYS\$SYSTEM:RIGHTS.LIST.DAT to the directory in which you placed the SYSUAF.DAT and NETPROXY.DAT files. Then define a clusterwide logical name for the RIGHTS.LIST.DAT file. For example:

```
$ DEFINE/SYSTEM/EXEC RIGHTS.LIST -  
      WORK5:[SYSEXE]RIGHTS.LIST
```

You must also add this DEFINE command to your common SYSTARTUP command file.

2.5.3 **Preparing the MAIL Database**

In a common-environment cluster, you may want to prepare a common mail database to allow users to use the Mail Utility (MAIL) to send and read their MAIL messages from any computer in the cluster.

Each time MAIL executes in a single-system environment, it accesses a database file named SYS\$SYSTEM:VMSMAIL_PROFILE.DAT. To set up VMSMAIL_PROFILE.DAT as a common file, define the logical name VMSMAIL_PROFILE to be the complete file specification of the common file by specifying the DEFINE command in the following format:

```
$ DEFINE/SYSTEM/EXEC VMSMAIL_PROFILE file-spec
```

You must make sure that you define the logical name before you invoke MAIL for the first time. When invoked for the first time, MAIL creates the database file, VMSMAIL_PROFILE.DAT, in SYS\$SYSTEM by default. By defining VMSMAIL_PROFILE to be the location of a common file on a cluster-accessible disk, you cause MAIL to create and use that file.

If your cluster is running from one common system disk, define VMSMAIL_PROFILE to be SYS\$COMMON:[SYSEXE]VMSMAIL_PROFILE and invoke the Mail Utility, by entering the following two commands:

```
$ DEFINE/SYSTEM/EXEC VMSMAIL_PROFILE -  
_ $ SYS$COMMON:[SYSEXE]VMSMAIL_PROFILE  
$ MAIL
```

VMSMAIL_PROFILE.DAT is created in the common system directory. You no longer need to use the logical name or make changes to your common SYSTARTUP command file.

Preparing the Cluster Operating Environment

2.5 Coordinating System Files for a Common-Environment Cluster

If your cluster is running from any other system disk configuration, you must decide where to locate the common `VMSMAIL_PROFILE.DATA` file. (Typically, you would place this file in the same directory in which `SYSUAF.DAT` and `NETPROXY.DAT` reside—for example, `WORK5:[SYSEXE]`.) You then define a logical name for the file and invoke the Mail Utility:

```
$ DEFINE/SYSTEM/EXEC VMSMAIL_PROFILE -  
    WORK5:[SYSEXE]VMSMAIL_PROFILE  
$ MAIL
```

The `DEFINE` command defines `VMSMAIL_PROFILE.DATA` to be a file located in `[SYSEXE]` on the cluster-accessible disk volume `WORK5`. The first time `MAIL` is invoked, `VMSMAIL_PROFILE.DATA` is created in `WORK5:[SYSEXE]`. Subsequently, `MAIL` uses this file as the database. You must also add the `DEFINE` command to your common `SYSTARTUP` command file.

2.5.4 Coordinating Shared System Files in Clusters with Multiple Common System Disks

To prepare a common user environment for a VAXcluster system that includes more than one common system disk, you must coordinate on those disks the system files listed in Section 2.5. In local area and mixed-interconnect clusters, you must also coordinate the file `SYS$MANAGER:NETNODE_UPDATE.COM`, which is described in Section 5.5.1.2.

Proceed as follows:

- 1 Edit the file `[VMS$COMMON.SYSMGR]SYLOGICALS.COM` on each system disk and define logical names that specify the location of the cluster common files. For example, if the files will be located on `1DJA16`, you could define logical names like the following:

```
$ DEFINE/SYSTEM/EXEC SYSUAF -  
    $1$DJA16:[VMS$COMMON.SYSEXE]SYSUAF.DAT  
$ DEFINE/SYSTEM/EXEC NETPROXY -  
    $1$DJA16:[VMS$COMMON.SYSEXE]NETPROXY.DAT  
$ DEFINE/SYSTEM/EXEC RIGHTSLIST -  
    $1$DJA16:[VMS$COMMON.SYSEXE]RIGHTSLIST.DAT  
$ DEFINE/SYSTEM/EXEC VMSMAIL_PROFILE -  
    $1$DJA16:[VMS$COMMON.SYSEXE]VMSMAIL_PROFILE.DATA  
$ DEFINE/SYSTEM/EXEC NETNODE_REMOTE -  
    $1$DJA16:[VMS$COMMON.SYSEXE]NETNODE_REMOTE.DAT  
$ DEFINE/SYSTEM/EXEC NETNODE_UPDATE -  
    $1$DJA16:[VMS$COMMON.SYSMGR]NETNODE_UPDATE.COM
```

- 2 To ensure that the system disks are correctly mounted with each reboot, follow these steps:
 - a. Copy the file `SYS$EXAMPLES:CLU_MOUNT_DISK.COM` to the directory `[VMS$COMMON.SYSMGR]`.

Preparing the Cluster Operating Environment

2.5 Coordinating System Files for a Common-Environment Cluster

- b. Edit SYLOGICALS.COM and include commands to mount, with appropriate volume label, the system disk containing the shared files. For example, if the system disk is \$1\$DJA16, you would include a command like the following:

```
$ @SYS$SYSDEVICE:[VMS$COMMON.SYSMGR]CLU_MOUNT_DISK.COM -  
$1$DJA16: volume-label
```

- 3 In the site-specific file used for queue setup, specify the location of the job-controller queue file (JBCSYSQUE.DAT), using a command like the following:

```
$ START/QUEUE/MANAGER -  
$1$DJA16:[VMS$COMMON.SYSEXE]JBCSYSQUE.DAT
```

When you execute CLUSTER_CONFIG.COM to add computers to a cluster with more than one common system disk, a different device name must be used for each system disk on which computers are added. For this reason, CLUSTER_CONFIG.COM supplies as a default device name the logical volume name (for example, DISK\$MARS_SYS1) of SYS\$SYSDEVICE: on the local system.

Using different device names ensures that each computer added has a unique root directory specification, even if the system disks contain roots with the same name—for example, DISK\$MARS_SYS1:[SYS10] and DISK\$MARS_SYS2:[SYS10].

3

Setting Up and Managing Cluster Disks

A VAXcluster system can include two types of disk and tape devices:

- Restricted-access devices, which are accessible only by the local computer or computers to which they are directly connected
- Cluster-accessible devices, which are accessible by any computer in the cluster

A disk or magnetic tape device connected to a hierarchical storage controller (HSC) subsystem is by design a cluster-accessible device. Any other disk device, such as a MASSBUS, UNIBUS, or BI disk, is a restricted-access device unless you explicitly set it up as a cluster-accessible device. Any other tape device is not accessible to the cluster.

As system manager, you are responsible for planning, organizing, and setting up the proper cluster device configuration for your site. You must decide which disk devices should have access restricted to the local computer and which should be accessible to the cluster. For example, you may want to restrict access to a particular disk to the users on the computer that is directly connected to the device. Alternatively, you may decide to set up a disk as a cluster-accessible device, so that any user on any computer can allocate and use it.

You can use the information in this chapter to plan and set up your disk configuration. Topics include the following:

- Cluster-accessible disks
- Cluster device-naming conventions
- Shared disks
- Configuring cluster disks
- Rebuilding cluster disks

3.1 Cluster-Accessible Disks

A **cluster-accessible disk** is a disk that every computer in the cluster can recognize and access. The following types of disks are cluster accessible:

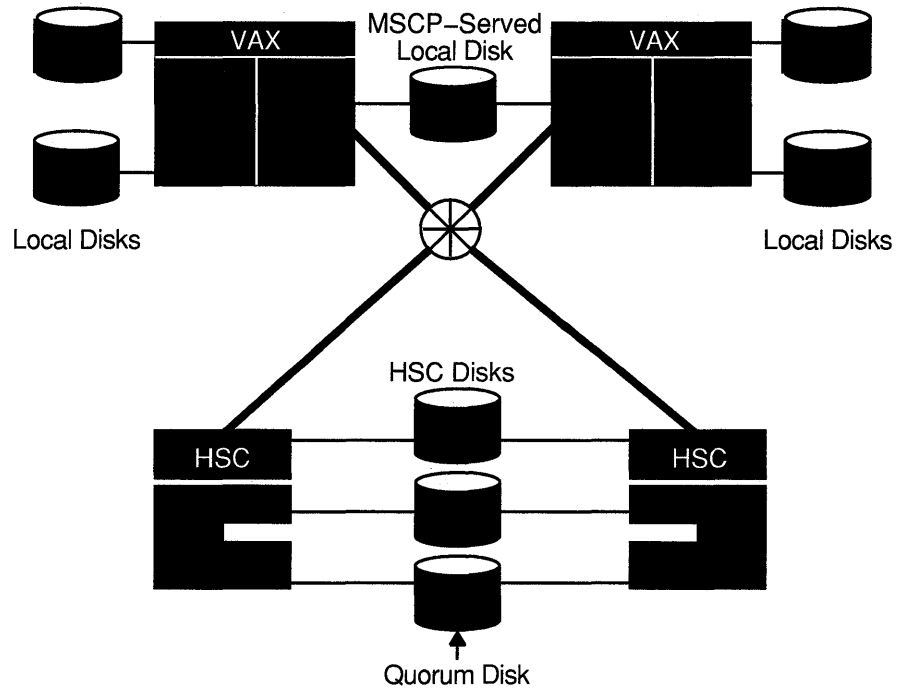
- HSC disks
- MSCP-served disks
- Dual-pathed disks

Figure 3–1 illustrates how disks might be configured in a typical CI-based cluster. The HSC disks and the dual-ported MSCP-served local disk are considered cluster accessible.

Setting Up and Managing Cluster Disks

3.1 Cluster-Accessible Disks

Figure 3-1 CI-Based Configuration with Shared Disks



ZK-1637-GE

3.1.1 HSC Disks

An HSC disk is a DIGITAL Storage Architecture (DSA) disk that is connected to an HSC subsystem. If an HSC subsystem is connected in a cluster, its disks are automatically accessible by any VAXcluster computer. You can set up HSC disks (including a quorum disk) to be dual pathed between two HSC subsystems, as shown in Figure 3-1. Dual-pathed disks are described in Section 3.1.3.

3.1.2 MSCP-Served Disks

The mass storage control protocol (MSCP) server is used to communicate between a computer and a DSA controller. The MSCP server enables a computer to make locally connected disks, such as MASSBUS, UNIBUS, and BI disks, available to all other cluster members.

Unlike HSC devices, controllers for locally connected disks are not automatically cluster accessible. Access to these devices is restricted to the local computer unless you explicitly set them up as cluster accessible using the MSCP server.

Setting Up and Managing Cluster Disks

3.1 Cluster-Accessible Disks

3.1.2.1 MSCP Server Functions

To make a disk accessible to all VAXcluster computers, the MSCP server must be loaded on the local computer, and it must be instructed to make the disk available across the cluster. MSCP server functions are enabled with the SYSGEN parameters `MSCP_LOAD` and `MSCP_SERVE_ALL` (see Table 3-1). By specifying appropriate values for these parameters in a computer's `MODPARAMS.DAT` file, and then running `AUTOGEN` to reboot the computer, you enable the computer to serve all suitable disks to the cluster early in the boot sequence. (You can also use the `CLUSTER_CONFIG.COM CHANGE` function to perform these operations.) The served disks thus become accessible with minimal interruption whenever the serving computer reboots. Further, the MSCP server automatically serves any suitable disk that is added to the system later. For example, if new drives are attached to an HSC subsystem, the disks become available within seconds after the cables are connected.

Table 3-1 summarizes the SYSGEN parameter values you can specify to configure the MSCP server. Initial values are determined by your responses when you execute the VMS installation or upgrade procedure, or when you execute the `CLUSTER_CONFIG.COM` command procedure described in Chapter 5 to set up your configuration. Note that if you later change the values, you must reboot the computer on which you changed the values before the new values can take effect (see Section 5.5.3).

Table 3-1 Specifying Values for `MSCP_LOAD` and `MSCP_SERVE_ALL` Parameters

Parameter	Value	Function
<code>MSCP_LOAD</code>	0	Do not load the MSCP server (default value).
	1	Load the MSCP server with attributes specified by <code>MSCP_SERVE_ALL</code> parameter.
<code>MSCP_SERVE_ALL</code>	0	Do not serve any disks (default value).
	1	Serve all available disks.
	2	Serve only locally connected (non-HSC) disks.

For a sample use of these parameters, see the discussion of Figure 3-4.

3.1.2.2 MSCP Load Sharing

MSCP servers monitor their I/O traffic and periodically calculate a load available rating to indicate available capacity for I/O requests.

Load available is calculated by counting the read and write requests sent to the server and periodically converting this to requests per second and subtracting this calculated value from the server's load capacity (also specified in requests per second).

This information is communicated to the MSCP class driver (`DUDRIVER` and `DSDRIVER`). When a disk is mounted or a failover occurs, the class driver selects the server with the highest load available rating to access the disk.

Setting Up and Managing Cluster Disks

3.1 Cluster-Accessible Disks

Load balancing is enabled and controlled by the SYSGEN parameters `MSCP_LOAD` and `MSCP_SERVE_ALL`. In most cases, the values established by `CLUSTER_CONFIG.COM` are appropriate.

The `MSCP_SERVE_ALL` parameter determines whether the server participates in load sharing. If the parameter is set to 2 (serve only local disks), the server does not monitor its I/O traffic and does not participate in load balancing. Other valid settings for `MSCP_SERVE_ALL` (0 or 1) result in the server monitoring I/O traffic and communicating load available information to the class drivers.

The `MSCP_LOAD` parameter is used to communicate load capacity to the server, in addition to its existing function of controlling the loading of the MSCP server. If the parameter is set to 1, the MSCP server is loaded and its load capacity is set to a default value based on CPU type. If `MSCP_LOAD` is set to a value greater than 1, the server is loaded and its load capacity set to that value. Setting `MSCP_LOAD` to zero disables loading of the MSCP server.

3.1.3 Dual-Pathed Disks

A **dual-pathed disk** is a disk that is accessible to all the computers in the cluster, not just to the computers that are physically connected to the disk. The term dual-pathed refers to the two paths through which computers can access a disk to which they are not directly connected. If one path fails, the disk is accessed over the other path. (Note that with a dual-ported MASSBUS disk, a computer directly connected to the disk always accesses it locally.) Dual-pathed disks can be any of the following:

- Dual-ported HSC disks
- Dual-pathed DSA disks on local UDA/KDA/KDB controllers
- DSSI-connected integrated storage assemblies (ISAs) in dual-host VAXcluster configurations
- Dual-ported MASSBUS disks

3.1.3.1 Dual-Ported HSC Disks

By design, HSC disks are cluster accessible. Therefore, if they are dual ported, they are automatically dual pathed. CI-connected computers can access a dual-pathed HSC disk by way of a path through either HSC subsystem connected to the device. If one HSC subsystem fails, access fails over to the other subsystem.

For each dual-ported HSC disk, you can control failover to a specific port using the port select buttons on the front of each drive. By pressing either port select button (A or B) on a particular drive, you can cause the device to fail over to the specified port.

With the port select buttons, you can select alternate ports to balance the disk controller workload between two HSC subsystems. For example, you can set half of your disks to use port A and set the other half to use port B.

Setting Up and Managing Cluster Disks

3.1 Cluster-Accessible Disks

The port select buttons also enable you to fail over all the disks to an alternate port manually when you anticipate the shutdown of one of the HSC subsystems.

3.1.3.2 Dual-Pathed DSA Disks on Local UDA/KDA/KDB Controllers

A dual-ported DSA disk can be failed over between the two computers that serve it to the cluster, provided that the same disk controller letter and allocation class are specified on both computers and that both computers are running the MSCP server.

Caution: Failure to observe these requirements can endanger data integrity.

However, because a DSA disk can be on line to only one controller at a time, only one of the computers can use its local connection to the disk. The second computer accesses the disk through the MSCP server. If the computer that is currently serving the disk fails, the other computer detects the failure and fails the disk over to its local connection. The disk is thereby made available to the cluster once more.

Note: A dual-ported DSA disk cannot be used as a system disk.

The VMS operating system supports specifying a preferred path for DSA disks, including RA-series disks and disks that are accessed through the MSCP server. If a preferred path is specified for a disk, the MSCP disk class drivers (DUDRIVER and DS DRIVER) use that path as their first attempt to locate the disk and bring it on line with a DCL MOUNT command or failover of an already mounted disk.

In addition, you can initiate failover of a mounted disk to force the disk to the preferred path or to use loadsharing information for disks accessed by MSCP servers.

The preferred path is specified by a \$QIO function (IO\$_SETPRFPTH), with the P1 parameter containing the address of a counted ASCII string (.ASCIC). This string is the node name of the HSC or VMS system that is to be the preferred path. The node name must match an existing node that is known to the local node, and, if it is a VMS system, it must be running the MSCP server. This function does not move the disk to the preferred path. For more information on use of the IO\$_SETPRFPTH function, refer to the *VMS I/O User's Reference Manual: Part I*.

3.1.3.3 DSSI-Connected ISAs

The DSSI bus connects as many as six ISAs between DSSI controllers on boot servers in dual-host VAXcluster systems (see Section 1.4.2). Simultaneously accessible to both servers, these storage assemblies can be served to satellites. If one boot server fails, access fails over to the other server and applications continue to run. Any DSSI-connected ISA can be used as a system disk. Note that, because most failures occur in system enclosures, you should try to locate the system disk in a storage expansion box, which has a dedicated power supply.

Like configurations based on HSC subsystems, DSSI configurations provide multihost access, flexibility, and expansion potential. However, they do not support controller-based (HSC-based) volume shadowing.

Setting Up and Managing Cluster Disks

3.1 Cluster-Accessible Disks

3.1.3.4 Dual-Ported MASSBUS Disks

A dual-ported MASSBUS disk can be connected between two computers, if it has the same controller letter and allocation class on both.

Before mounting the disk, enter the DCL command SET DEVICE in the following format on both computers:

```
SET DEVICE/DUAL_PORT device-name
```

Note: A MASSBUS disk can be used either as a dual-ported disk or as a system disk, but not as both.

In clusters with more than two computers, you can set up a dual-ported MASSBUS disk to be cluster accessible through the MSCP server on either or both computers to which the disk is connected. Be sure, however, *not* to use the SYSGEN commands AUTOCONFIGURE or CONFIGURE to configure a dual-ported MASSBUS disk that is already available on the computer through the MSCP server. Establishing a local connection to the disk when a remote path is already known creates two uncoordinated paths to the same disk. Use of these two paths can endanger data integrity on any disk that is mounted on the drive.

If the local path to the disk is not found during the system bootstrap procedure, the MSCP server path from the remote computer is the only available access to the drive. The local path is not found during a boot if any of the following conditions exist:

- The port select switch for the drive is not enabled for the local computer.
- The disk, cable, or adapter hardware for the local path is broken.
- There is sufficient activity on the other port to “mask” the existence of the port.
- The computer is booted in such a way that the SYSGEN command AUTOCONFIGURE ALL in the site-independent startup procedure (SYS\$SYSTEM:STARTUP.COM) was not executed.

Use of the disk is still possible through the MSCP server path.

Caution: Under these conditions, do not attempt to add the local path back into the system I/O database using the SYSGEN commands AUTOCONFIGURE or CONFIGURE. SYSGEN is currently unable to detect the presence of the disk’s MSCP path and would incorrectly build a second set of data structures to describe it. Subsequent events could lead to incompatible and uncoordinated file operations, which might endanger data integrity.

To recover the local path to the disk, you must reboot the computer connected to that local path.

Note that if the disk is not dual ported or is never MSCP served on the remote computer, this restriction does not apply.

Setting Up and Managing Cluster Disks

3.2 Cluster Device-Naming Conventions

3.2 Cluster Device-Naming Conventions

To manage cluster devices properly, you must understand the conventions used to identify them. Every cluster device is identified by a unique name, which provides a reliable way to access it in the cluster.

Devices that are local to a VAXcluster computer can be accessed by that computer through the traditional device name (for example, DUA1) or through a cluster device name in the format *node-name\$device-name* (for example, JUPITR\$DUA1).

However, a device that is dual pathed between two computers must be identified by a unique, path-independent name that includes an **allocation class**. The allocation class is a numeric value from 0 to 255 that is used to create a device name in the following format:

```
$allocation-class$device-name
```

For example, the allocation class device name \$1\$DJA17 identifies a disk that is dual ported between two computers or HSC subsystems that both have an allocation class value of 1.

Each time a computer that is not directly connected to such a disk tries to access the disk, the choice of which path to take is made arbitrarily; no one path to the disk is ever guaranteed. Because the access path is chosen without regard to the names of the computers or HSC subsystems serving the disk, an allocation class device name is required to identify the disk uniquely.

3.2.1 Rules for Specifying Allocation Class Values

Allocation classes play an important role in determining strategies for configuring and naming disks. In fact, the VMS operating system uses allocation class values above all other available information when determining the configuration of cluster devices.

The following rules apply for specifying allocation class values:

- Computers or HSC subsystems to which a dual-pathed disk is connected must have the same nonzero allocation class value.
- All cluster-accessible disks on computers with a nonzero allocation class value must have unique names. For example, if two computers have the same allocation class value, it is invalid for both computers to have a disk named DJA0. This restriction also applies to HSC subsystems.
- Single-ported disks with an allocation class value of zero can have the same unit number on different computers.

Zero is the default allocation class value. Any computer in a CI-based cluster that is not connected to a dual-pathed disk should be assigned this

Setting Up and Managing Cluster Disks

3.2 Cluster Device-Naming Conventions

value. In a mixed-interconnect cluster, all of the following must have a nonzero allocation class value:

- HSC subsystems
- Computers that serve HSC disks
- Computers connected to dual-pathed disks

Note that if a dual-host configuration is included in a mixed-interconnect cluster, you must specify a unique allocation class value for the computers and ISAs in the dual-host configuration. This value must be different from that of other computers and HSC subsystems in the cluster.

Caution: Failure to set allocation class values correctly can endanger data integrity and cause locking conflicts that suspend normal cluster operations.

To assign an allocation class value to a VAXcluster computer that supports dual-pathed devices, specify the value with the SYSGEN parameter ALLOCLASS. To assign an allocation class for an HSC subsystem, specify the value using the HSC console to enter a command in the following format, where *n* is the allocation class value:

```
SET ALLOCATE DISK n
```

For complete information on HSC console commands, refer to the HSC hardware documentation.

3.2.2 Sample Configurations with Named Devices

Figure 3–2 and Figure 3–3 show how cluster device names are specified for the following:

- Dual-pathed HSC disks
- Dual-pathed DSA disks

Figure 3–2 shows a mixed-interconnect VAXcluster segment with a dual-pathed HSC disk. Note that the allocation class value (1) is the same on VAX computers JUPITR and SATURN, and that the disk's device name (\$1\$DJA17) is constructed using that value. JUPITR and SATURN can access the disk through either of the HSC subsystems VOYGR1 or VOYGR2.

Setting Up and Managing Cluster Disks

3.2 Cluster Device-Naming Conventions

Figure 3-2 Mixed-Interconnect VAXcluster Segment with Dual-Pathed HSC Disk

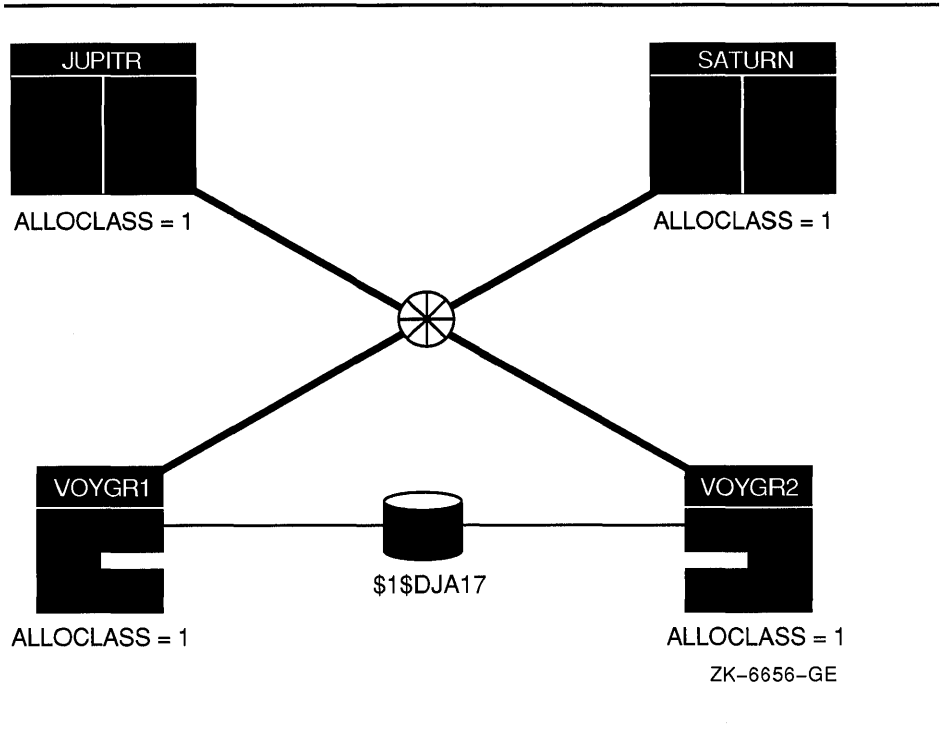
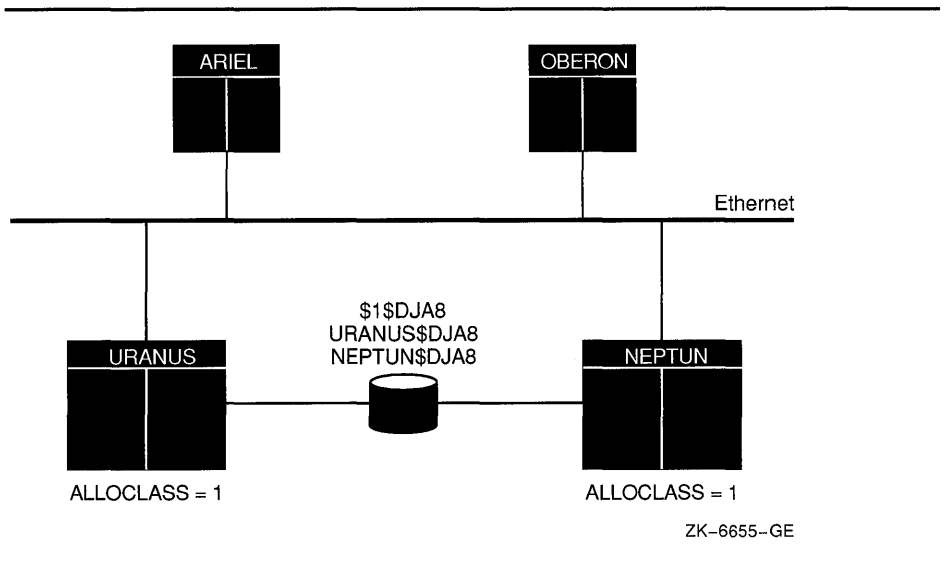


Figure 3-3 shows a mixed-interconnect VAXcluster segment with a dual-pathed DSA disk.

Figure 3-3 Mixed-Interconnect VAXcluster Segment with Dual-Pathed DSA Disk



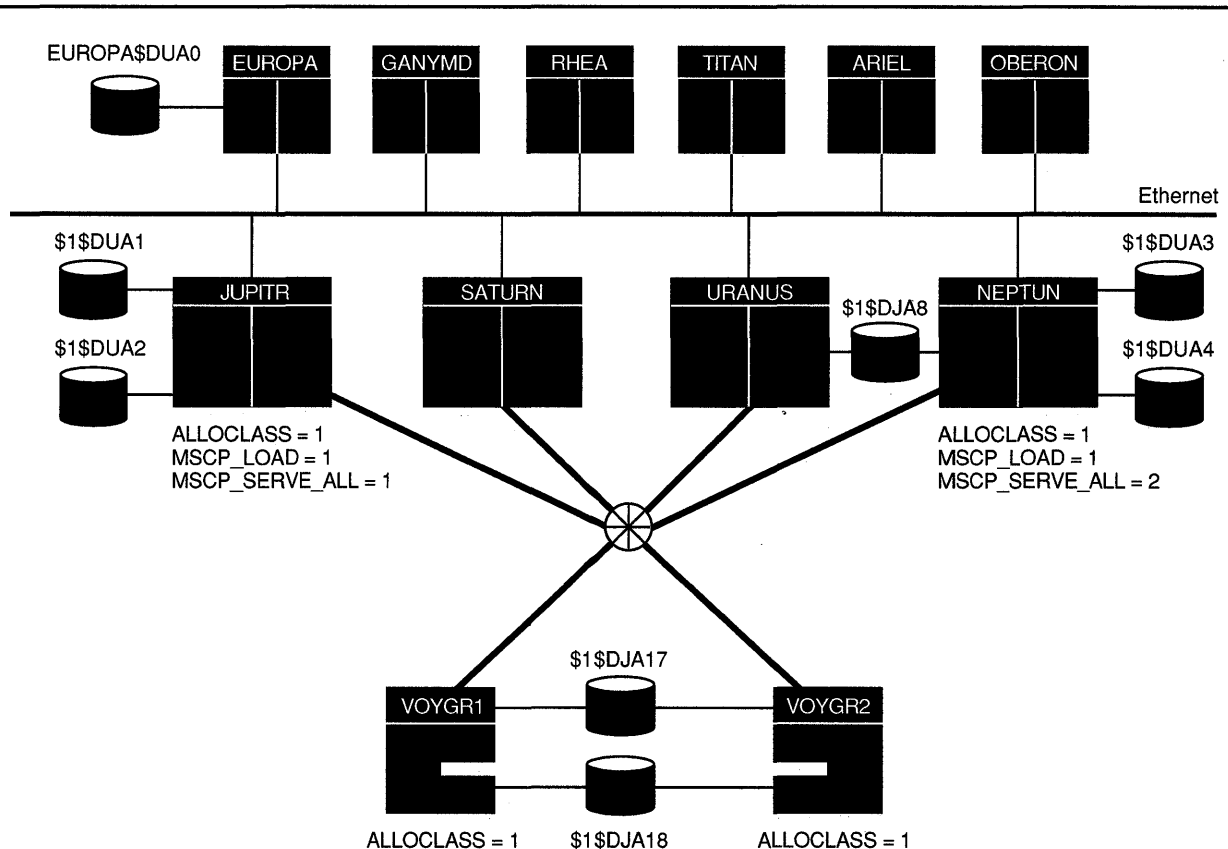
Setting Up and Managing Cluster Disks

3.2 Cluster Device-Naming Conventions

URANUS and NEPTUN can access the disk either locally or through the other computer's MSCP server. However, when satellites ARIEL and OBERON access the disk, they arbitrarily choose a path through either URANUS or NEPTUN. If the satellites try to access the disk by using the computer-specific device name URANUS\$DJA8, and the disk is not currently accessible through URANUS, access will fail. But if the satellites use the allocation class device name \$1\$DJA8, they can access the disk through NEPTUN. As a general rule, you should always use a path-independent, allocation class device name to identify dual-pathed cluster disks.

Figure 3-4 shows how device names are typically specified in a mixed-interconnect cluster. This figure also shows how relevant SYSGEN parameter values are set in each CI-connected computer's MODPARAMS.DAT file. Note that the values shown for JUPITR are the same for SATURN and URANUS, but that NEPTUN has a different value for MSCP_SERVE_ALL.

Figure 3-4 Device Names in a Mixed-Interconnect Cluster



ZK-6660-GE

In this configuration, a set of disks is dual-pathed to the HSC subsystems named VOYGR1 and VOYGR2; these subsystems are connected to JUPITR, SATURN, URANUS, and NEPTUN through the star coupler. The MSCP server is loaded on all four computers (MSCP_LOAD = 1)

Setting Up and Managing Cluster Disks

3.2 Cluster Device-Naming Conventions

and the ALLOCLASS parameter is set to the same value (1) on these computers and on both HSC subsystems. But MSCP_SERVE_ALL is set to 1 only on JUPITR, SATURN, and URANUS. Therefore, only these three computers can serve the disks on VOYGR1 and VOYGR2 to satellites. Because MSCP_SERVE_ALL is set to 2 on NEPTUN, NEPTUN can serve only its local disks.

Disks on the HSC subsystems have allocation class names in the form `1ddcu`. For example, the disk DUA17 is named `1DUA17`. On CI-connected computers, VMS software would also recognize the disk as `JUPITR$DUA17` and as either `VOYGR1$DUA17` or `VOYGR2$DUA17`. On satellites, it would recognize the disk as `JUPITR$DUA17` or as `$1$DUA17`. This example shows why you should always use an allocation class name like `1DUA17` when you configure cluster devices: The allocation class name is the only name that all computers recognize at all times.

Note that, for optimal availability, two or more CI-connected computers should serve HSC disks to the cluster.

3.3 Shared Disks

A **shared disk** is a disk that is mounted on a cluster-accessible device by one or more VAXcluster computers. Shared disks play a key role in common-environment clusters, because when you place system files or command procedures on a shared disk, computers can share a single copy of each common file (see Chapter 2). Note, however, that a shared disk is a single point of failure for data access by the computers sharing the disk.

To mount cluster-accessible disks that are to be shared among all computers, specify the same MOUNT command on each computer or specify the MOUNT command with the /CLUSTER qualifier on one or more computers. (Typically, you specify MOUNT/CLUSTER on all computers.) When you execute MOUNT/CLUSTER on one computer, the disk is mounted on every computer that is active in the cluster at the time the command executes. Note that only system or group disks can be mounted across the cluster. Thus, if you specify MOUNT/CLUSTER without the /SYSTEM or /GROUP qualifier, /SYSTEM is assumed. Also note that each cluster disk mounted with the /SYSTEM, /GROUP, or /SHARED qualifiers must have a unique volume label.

If you want to mount a shared disk on some but not all VAXcluster computers, execute the same MOUNT command (without the /CLUSTER qualifier) on each computer that shares the disk.

For example, suppose you want all the computers in a three-member cluster to share a disk named COMPANYDOCS. To share the disk, each of the three computers can execute identical MOUNT commands, or one of the three computers can mount COMPANYDOCS using the MOUNT /CLUSTER command, as follows:

```
$ MOUNT/SYSTEM/CLUSTER/NOASSIST $1$DUA4: COMPANYDOCS
```

Setting Up and Managing Cluster Disks

3.3 Shared Disks

If you want just two of the three computers to share the disk, those two computers must both mount the disk with the same MOUNT command. For example:

```
$ MOUNT/SYSTEM/NOASSIST $1$DUA4: COMPANYDOCS
```

To mount the disk at startup time, include the mount command either in a common command procedure that is invoked at startup time or in the computer-specific startup command file.

3.4 Configuring Cluster Disks

To configure cluster disks, you can create command procedures to set up and mount them. You may want to include commands that set up and mount cluster disks in a separate command procedure file that is invoked by a site-specific SYSTARTUP procedure. Depending on your cluster environment, you can set up your command procedure in either of the following ways:

- As a separate file specific to each computer in the cluster
- As a common computer-independent file

You can set up the common procedure as a shared file on a shared disk, or you can make copies of the common procedure and store them as separate files. With either method, each computer can invoke the common procedure from the site-specific SYSTARTUP procedure.

The MSCPMOUNT.COM file in the SYS\$EXAMPLES directory on your system is a sample common command procedure that contains commands typically used to mount cluster disks. The example includes comments explaining each phase of the procedure.

3.5 Rebuilding Cluster Disks

To minimize disk I/O operations (and thus improve performance) when files are created or extended, the VMS file system maintains a cache of preallocated file headers and disk blocks.

If a disk is improperly dismounted—for example, if a system crashes or is removed from a cluster without running SYS\$SYSTEM:SHUTDOWN.COM—this preallocated space becomes temporarily unavailable. When the disk is remounted, MOUNT scans the disk to recover the space as it rebuilds the disk.

On a nonclustered computer, the scan operation merely prolongs the boot process. In a VAXcluster system, however, this operation can degrade response time for all user processes in the cluster. While the scan is in progress on a particular disk, most activity on that disk is blocked. User processes that attempt to read or write to files on the disk can experience delays of several minutes or longer, especially if the disk contains a large number of files or has many users.

Setting Up and Managing Cluster Disks

3.5 Rebuilding Cluster Disks

Because the rebuild operation can delay access to disks during the startup of any VAXcluster computer, Digital recommends that procedures for mounting cluster disks use the /NOREBUILD qualifier. When MOUNT /NOREBUILD is specified, disks are not scanned to recover lost space, and users experience minimal delays while computers are mounting disks.

System disks are especially critical in this regard, because most system activity requires access to a system disk. When a system disk rebuild is in progress, very little activity is possible on any computer that uses that disk. Unlike other disks, the system disk is automatically mounted early in the boot sequence. If a rebuild is necessary, and if the value of the SYSGEN parameter ACP_REBLDSYSD is 1, the system disk is rebuilt during the boot sequence. (The default setting of 1 for the SYSGEN parameter ACP_REBLDSYSD specifies that the system disk should be rebuilt.)

In local area and mixed-interconnect clusters, however, the ACP_REBLDSYSD parameter should normally be set to zero on all satellites. This setting prevents them from rebuilding a system disk when it is mounted early in the boot sequence and eliminates delays caused by such a rebuild when satellites join the cluster.

In large clusters, a substantial amount of system disk space (some for each computer) might be preallocated to caches, and if many computers abruptly leave the cluster (for example, during a power failure), this space can become temporarily unavailable. Thus, ACP_REBLDSYSD on boot servers in local area and mixed-interconnect clusters with many computers should be set to the default value of 1, and procedures that mount disks on the boot servers should use the /REBUILD qualifier. While these measures can make boot server rebooting more noticeable, they ensure that system disk space is available after an unexpected shutdown.

Once the cluster is up and running, system managers can submit one or more batch procedures that execute SET VOLUME/REBUILD commands to recover lost disk space. Such procedures can run at a time when users would not be inconvenienced by the blocked access to disks (for example, between midnight and 6 a.m. each day). Because the SET VOLUME/REBUILD command determines whether a rebuild is needed, the procedures can execute the command for each disk that is usually mounted. Note that the procedures run more quickly and cause less delay in disk access if they are executed on powerful computers. Moreover, several such procedures, each of which rebuilds a different set of disks, can be executed simultaneously.

Caution: If either or both MOUNT/NOREBUILD and ACP_REBLDSYSD = 0 are specified when mounting disks, it is essential to run a procedure with SET VOLUME/REBUILD commands on a regular basis to rebuild the disks. Failure to rebuild disk volumes can result in a loss of free space and in subsequent failures of applications to create or extend files.

4

Setting Up and Managing Cluster Queues

On a standalone computer, print and batch job processing is limited to local devices. In VAXcluster systems, however, computers can share device and processing resources. This ability to share resources allows for better workload balancing because batch and print jobs can be distributed across the cluster.

You control how jobs share device and processing resources in a cluster by setting up and maintaining cluster queues. The strategy you use to set up and manage these queues determines how well workloads are matched to your cluster's resources.

This chapter explains how to set up and manage VAXcluster queues. Topics include the following:

- Clusterwide queues
- Cluster printer queues
- Cluster batch queues
- Using a common command procedure to set up cluster queues

Because queues in a VAXcluster system are established and controlled with the same commands used to manage queues on a standalone computer, the discussions in this chapter assume some knowledge of queue management on a standalone system, as described in the *Guide to Setting Up a VMS System*.

4.1 Clusterwide Queues

Clusterwide queues are controlled by the clusterwide job-controller queue file JBCSYSQUE.DAT. This file makes queues available across the cluster and allows jobs to execute on any queue from any computer, provided that the necessary mass storage volumes can be accessed by the computer on which the job executes.

For clusterwide queues, only one job-controller queue file can be used, and that file must be located on a disk that is accessible to the computers participating in the clusterwide queue scheme. Note that performance can be enhanced by locating the file on a shared disk that has a low level of activity.

Setting Up and Managing Cluster Queues

4.1 Clusterwide Queues

All VAXcluster computers that share clusterwide queues must use the same queue file. You control which computers share these queues by specifying the location of the common job-controller queue file with a command in the following format:

```
START/QUEUE/MANAGER device:[directory]JBCSYSQUE.DAT
```

You can include the command either in each participating computer's specific SYSTARTUP command file or in a common queue startup command file that is shared by all computers participating in the clusterwide queue scheme. A sample common procedure for setting up a VAXcluster batch and print system is provided in Section 4.4.

4.2 Cluster Printer Queues

To establish printer queues, you must determine the type of queue configuration that best suits your VAXcluster system. You have several alternatives that depend on the number and type of print devices you have on each computer and on how you want print jobs to be processed. For example, make these decisions:

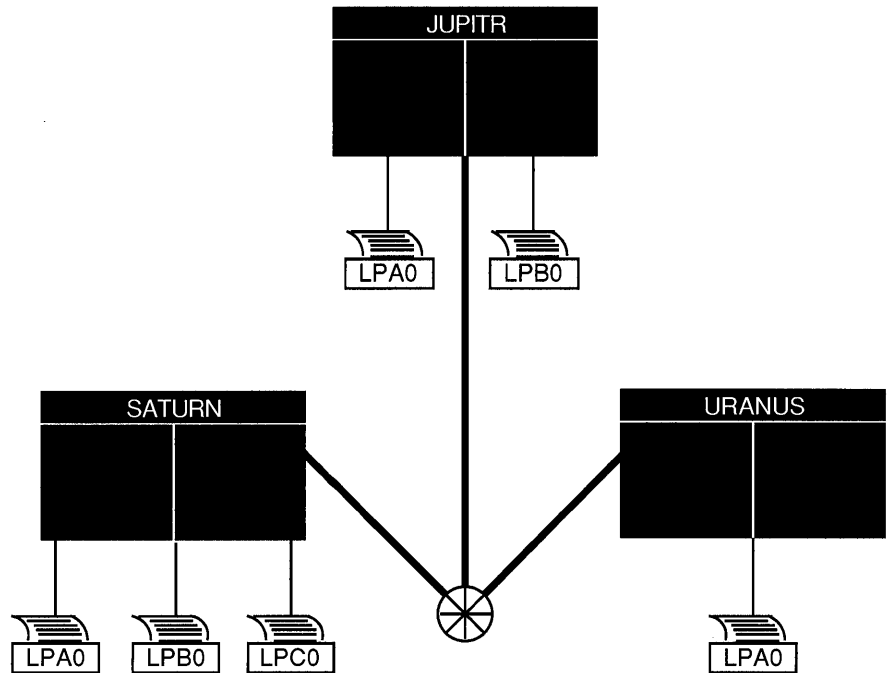
- Which printer queues you want to establish on each computer
- Whether to set up any clusterwide generic queues to distribute print job processing across the cluster

Once you determine the appropriate strategy for your cluster, you can create a common command procedure to set up queues (see Example 4-1). Figure 4-1 shows the printer configuration for a cluster consisting of the active computers JUPITR, SATURN, and URANUS. Section 4.2.1 and Section 4.2.2 describe various methods for establishing and naming the cluster printer queues shown in this configuration.

Setting Up and Managing Cluster Queues

4.2 Cluster Printer Queues

Figure 4-1 Sample Printer Configuration



ZK-1631-GE

4.2.1 Setting Up Printer Queues

You set up printer queues using the same procedures that you would use for a standalone computer (see the *Guide to Setting Up a VMS System*). However, in a VAXcluster system, you must provide a unique name for each queue you create.

In the appropriate startup file, you assign a unique name to a printer queue by specifying the DCL command `INITIALIZE/QUEUE` in the following format:

```
INITIALIZE/QUEUE/ON=node-name::device[/START] queue-name
```

The `/ON` qualifier specifies the computer and printer to which the queue is assigned. If you specify the `/START` qualifier, the queue is started.

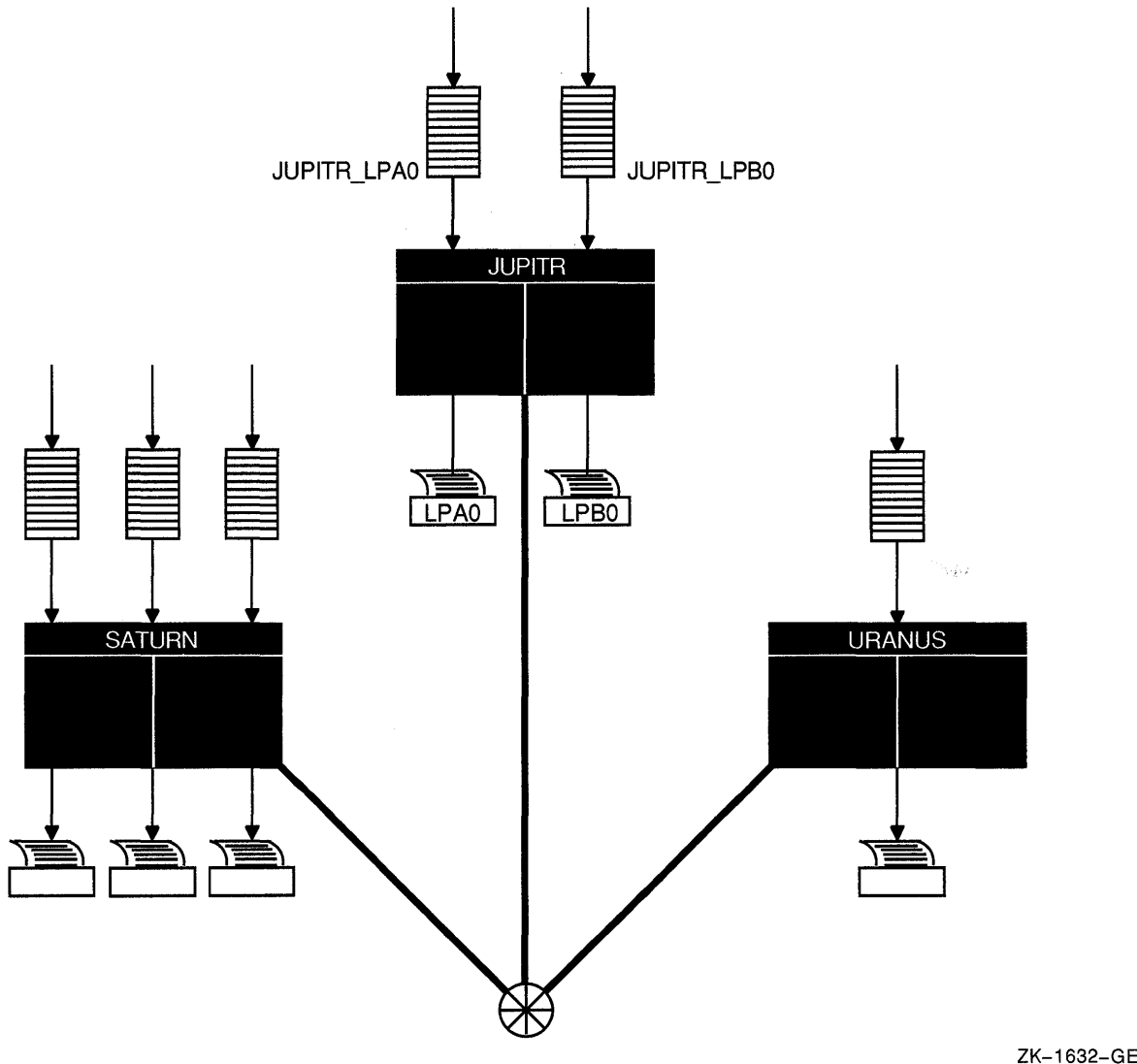
The following commands make the local printer queue assignments for JUPITR shown in Figure 4-2 and start the queues:

```
$ INITIALIZE/QUEUE/ON=JUPITR::LPA0/START JUPITR_LPA0  
$ INITIALIZE/QUEUE/ON=JUPITR::LPB0/START JUPITR_LPBO
```

Setting Up and Managing Cluster Queues

4.2 Cluster Printer Queues

Figure 4-2 Printer Queue Configuration



4.2.2 Setting Up Clusterwide Generic Printer Queues

The clusterwide job-controller queue file enables you to establish generic queues that function throughout the cluster. Jobs queued to clusterwide generic queues are placed in any assigned printer queue that is available, regardless of its location in the cluster. However, the file queued for printing must be accessible to the computer to which the printer is connected.

Figure 4-3 illustrates a clusterwide generic printer queue in which the queues for all LPA0 printers in the cluster are assigned to a clusterwide generic queue named SYS\$PRINT.

Setting Up and Managing Cluster Queues

4.2 Cluster Printer Queues

The following command initializes and starts the clusterwide generic queue SYS\$PRINT:

```
$ INITIALIZE/QUEUE/GENERIC=(JUPITR_LPA0,SATURN_LPA0,-  
URANUS_LPA0)/START SYS$PRINT
```

Jobs queued to SYS\$PRINT are placed in whichever assigned printer queue is available. Thus, in this example, a print job from JUPITR that is queued to SYS\$PRINT can be queued to JUPITR_LPA0, SATURN_LPA0, or URANUS_LPA0.

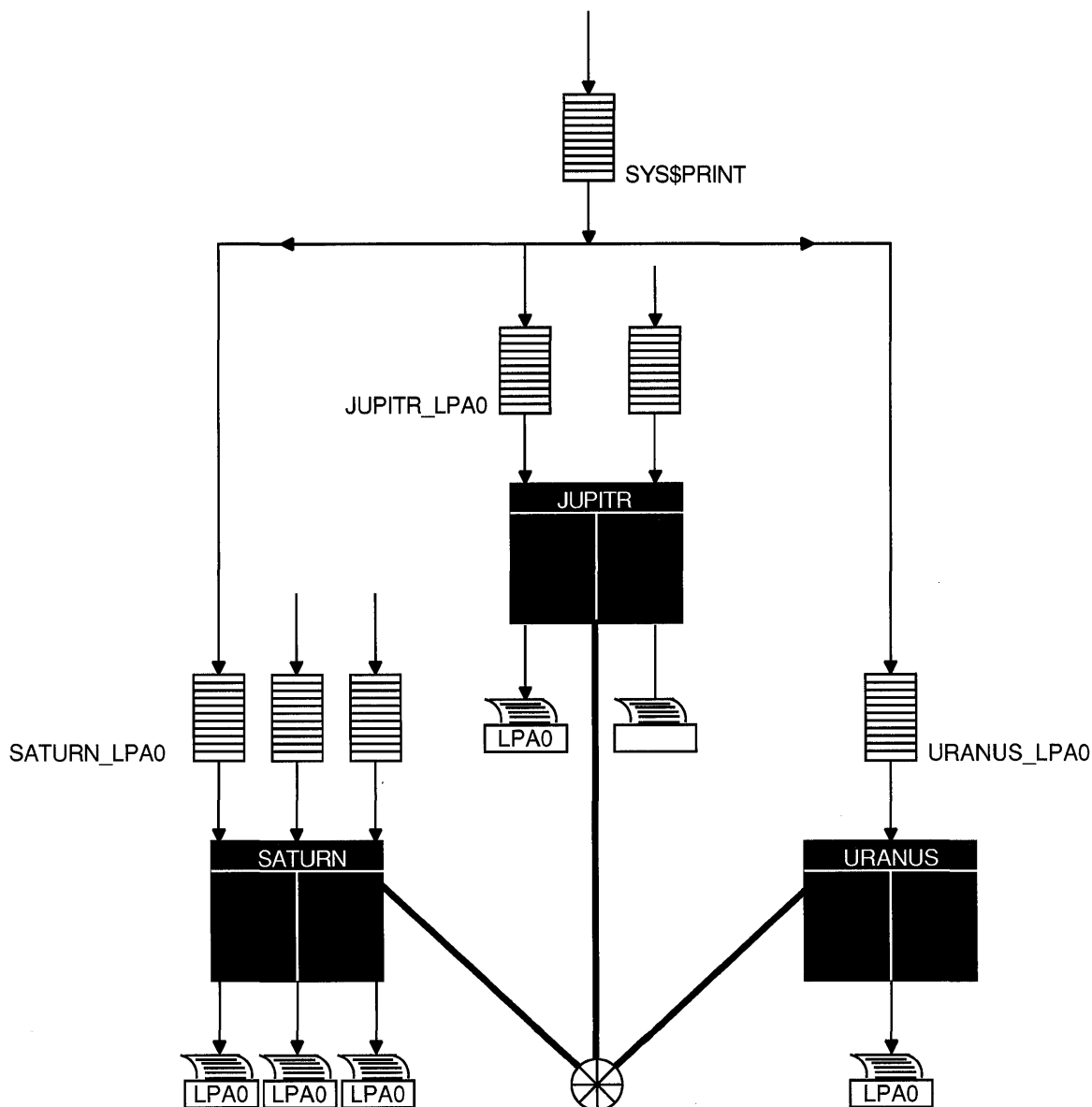
Because print jobs on each VAXcluster computer are queued to SYS\$PRINT by default, you should consider defining a logical name to establish this queue as a clusterwide generic printer queue that distributes print job processing throughout the cluster (see Example 4-1). Note, however, that you should do this only if you plan to set up a common-environment cluster.

A clusterwide generic printer queue needs to be initialized and started only once. The most efficient way to perform these operations is to create a common command procedure that is executed by each VAXcluster computer (see Example 4-1).

Setting Up and Managing Cluster Queues

4.2 Cluster Printer Queues

Figure 4-3 Clusterwide Generic Printer Queue Configuration



ZK-1634-GE

4.3 Cluster Batch Queues

Before you establish batch queues, you should decide on the type of queue configuration that best suits your cluster. As system manager, you are responsible for setting up batch queues to maintain efficient batch job processing on the cluster. For example, you should do the following:

- Determine what type of processing will be performed on each computer

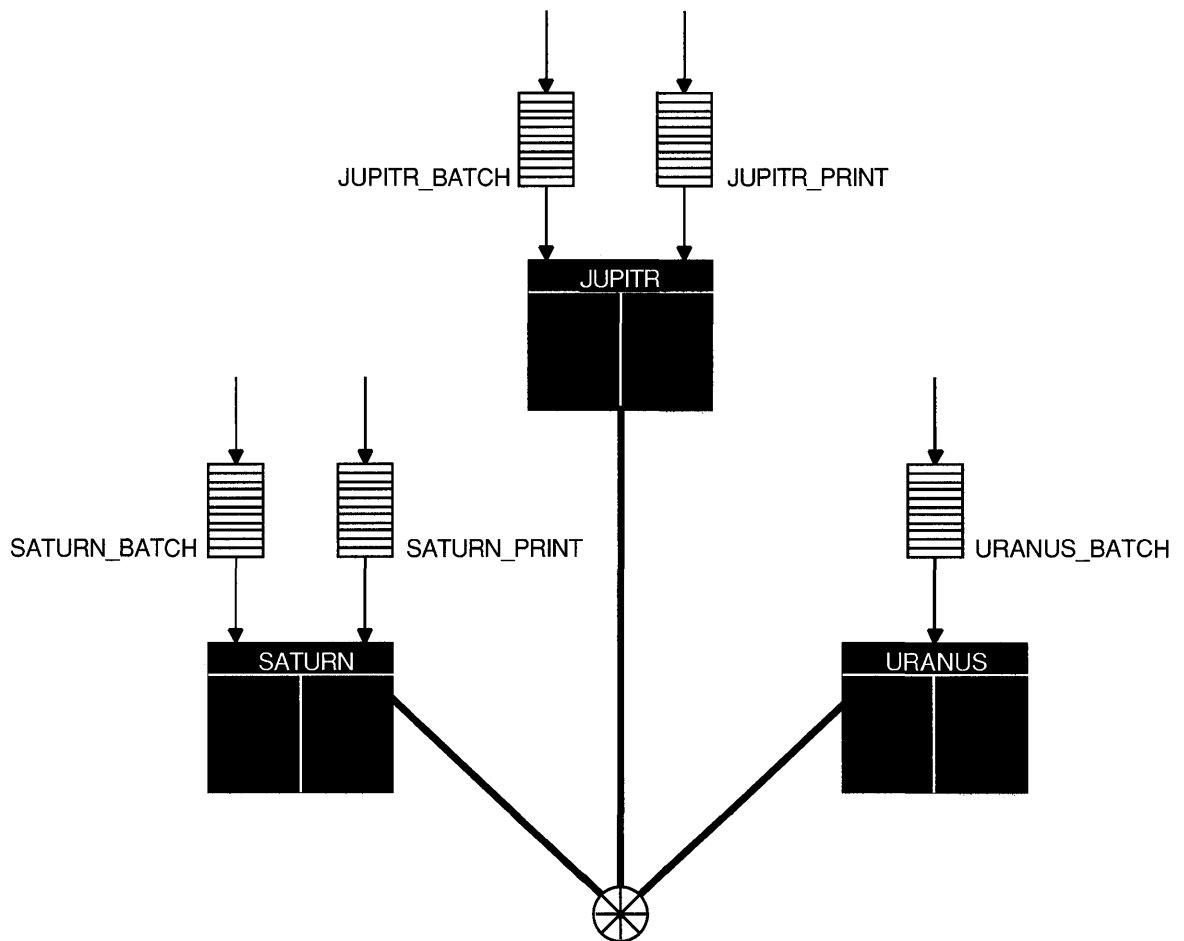
Setting Up and Managing Cluster Queues

4.3 Cluster Batch Queues

- Set up local batch queues that conform to these processing needs
- Decide whether to set up any clusterwide generic queues that will distribute batch job processing across the cluster

Once you determine the strategy that best suits your needs, you can create a command procedure to set up your queues (see Example 4-1). Figure 4-4 shows a batch queue configuration for a cluster consisting of computers JUPITR, SATURN, and URANUS.

Figure 4-4 Sample Batch Queue Configuration



ZK-1635-GE

4.3.1 Setting Up Executor Batch Queues

Generally, you set up executor batch queues on each VAXcluster computer using the same procedures you use for a standalone computer. For more detailed information on how this is done, see the *Guide to Setting Up a VMS System*.

Setting Up and Managing Cluster Queues

4.3 Cluster Batch Queues

In the appropriate startup file, you assign a unique name to a batch queue by specifying the DCL command `INITIALIZE/QUEUE` in the following format:

```
INITIALIZE/QUEUE/ON=node-name::[/START] queue-name
```

The `/ON` qualifier specifies the computer on which the batch queue runs. If you specify the `/START` qualifier, the queue is started.

The following commands make the local batch queue assignments for JUPITR, SATURN, and URANUS shown in Figure 4–4:

```
$ INITIALIZE/QUEUE/BATCH/ON=JUPITR:./START JUPITR_BATCH  
$ INITIALIZE/QUEUE/BATCH/ON=SATURN:./START SATURN_BATCH  
$ INITIALIZE/QUEUE/BATCH/ON=URANUS:./START URANUS_BATCH
```

Because batch jobs on each VAXcluster computer are queued to `SYS$BATCH` by default, you should consider defining a logical name to establish this queue as a clusterwide generic batch queue that distributes batch job processing throughout the cluster (see Example 4–1). Note, however, that you should do this only if you have a common-environment cluster. Guidelines for establishing clusterwide generic batch queues are presented in Section 4.3.2.

4.3.2 Setting Up Clusterwide Generic Batch Queues

In a VAXcluster system, you can distribute batch processing among computers to balance the use of processing resources. You can achieve this workload distribution by assigning local batch queues to one or more clusterwide generic batch queues. These generic batch queues control batch processing across the cluster by placing batch jobs in assigned batch queues that are available. You can create a clusterwide generic batch queue as shown in Example 4–1.

In Figure 4–5, batch queues from each VAXcluster computer are assigned to a clusterwide generic batch queue named `SYS$BATCH`. Users can submit a job to a specific queue (for example, `JUPITR_BATCH` or `SATURN_BATCH`) or, if they have no special preference, they can submit it by default to the clusterwide generic queue `SYS$BATCH`. The generic queue in turn places the job in an available assigned queue in the cluster.

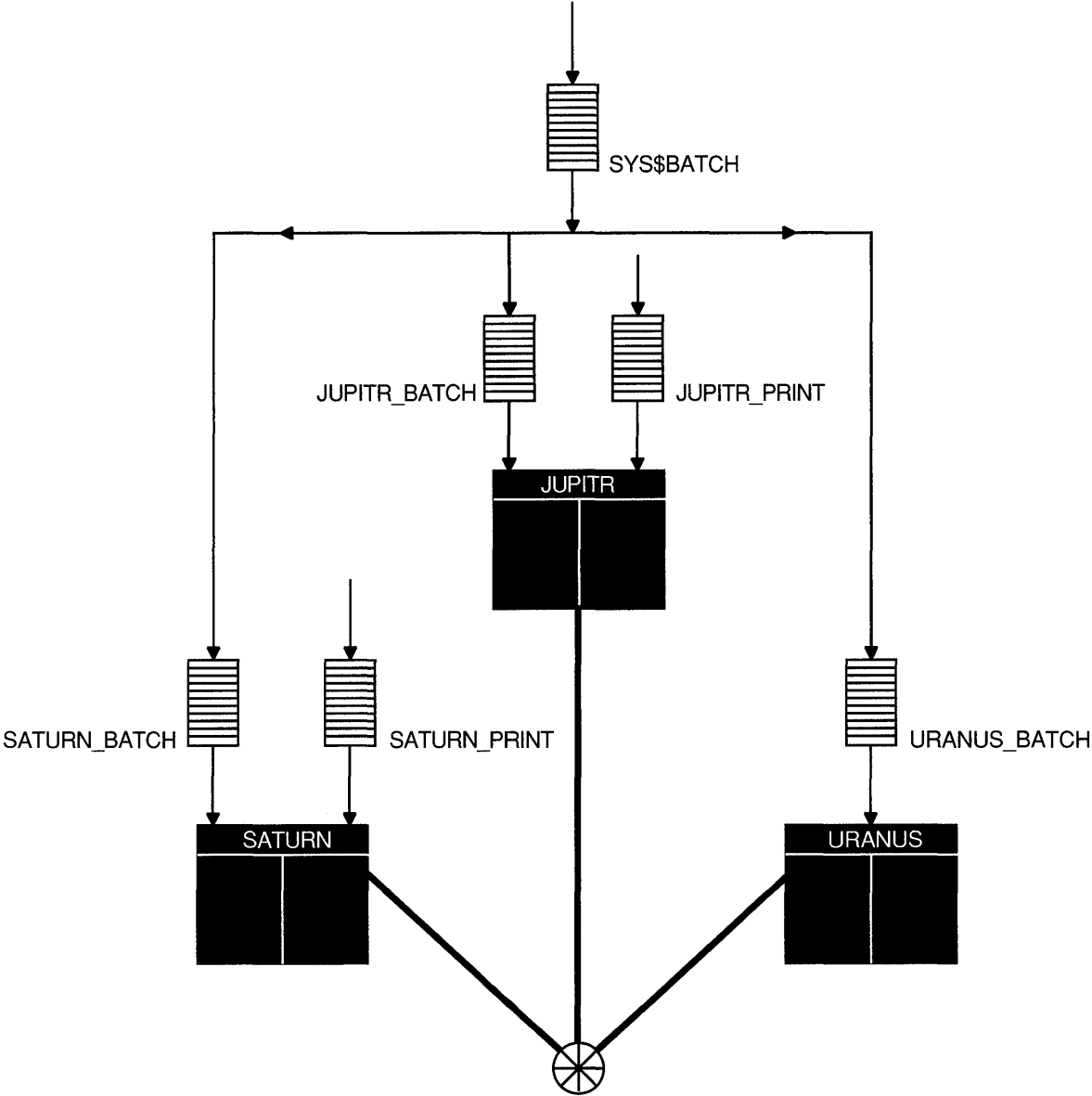
If more than one assigned queue is available, the VMS operating system selects the queue that minimizes the ratio (executing jobs/job limit) for all assigned queues.

A clusterwide generic batch queue needs to be initialized and started only once. The most efficient way to perform these operations is to create a common command procedure that is executed by each VAXcluster computer (see Example 4–1).

Setting Up and Managing Cluster Queues

4.3 Cluster Batch Queues

Figure 4-5 Clusterwide Generic Batch Queue Configuration



ZK-1636-GE

Setting Up and Managing Cluster Queues

4.4 Using a Common Command Procedure to Set Up Cluster Queues

4.4 Using a Common Command Procedure to Set Up Cluster Queues

To ensure that all queues are recognized by all computers in a VAXcluster system and by new computers joining the cluster, you must coordinate procedures that initialize and start queues. The first VAXcluster computer typically defines the entire batch and print system, initializing all queues and starting its own local queues. As other computers boot, they start their local queues.

You can include commands to establish queues in a common SYSTARTUP procedure or in a separate, common queue startup command file named, for example, STARTQ.COM, that is invoked by the common SYSTARTUP procedure. Digital suggests that you set up a common STARTQ.COM command procedure file on a shared disk and invoke this file from the common SYSTARTUP procedure. The common STARTQ.COM file can reside on the same disk as the job-controller queue file. With this method, each computer can share the same copy of the common STARTQ.COM procedure, and each computer invokes that procedure from the common version of SYSTARTUP.

Example 4–1 shows a common STARTQ.COM command procedure that is executed by all computers to initialize and start cluster batch and print queues. This procedure conditionally performs a full or partial startup of the clusterwide queue system. On a full startup (for example, after a cluster reboot), the procedure defines all printer forms and characteristics and initializes and starts all queues. On a partial startup (for example, when a computer boots into an active cluster), the procedure initializes and starts only that computer's queues. Thus, conditional startup results in substantial performance gains.

The sample procedure performs the following operations:

- Starts the system job queue manager.
- Specifies the location of the job-controller queue file.
- Defines logical names for clusterwide generic batch and print queues.
- Determines whether full or partial startup is needed. For full startup, the procedure does the following:
 - Defines all printer forms and characteristics.
 - Initializes all local queues and starts the executing computer's local queues.
 - Initializes and starts the clusterwide generic queues SYS\$BATCH and SYS\$PRINT.
 - Submits a batch job to start layered products.

For partial startup, the procedure initializes and starts only local queues.

Setting Up and Managing Cluster Queues

4.4 Using a Common Command Procedure to Set Up Cluster Queues

Example 4-1 Common Procedure to Set Up VAXcluster Queues

```
$!  
$! STARTQ.COM -- Common procedure to set up cluster queues  
$!  
$! Start the job controller. ①  
$!  
$ START/QUEUE/MANAGER/RESTART/BUFFER=500 WORK1:[QUEUE]JBCSYSQUEUE.DAT  
$!  
$! Define clusterwide logical names for generic batch and print queues. ②  
$!  
$ DEFINE/SYSTEM/EXEC SYS$BATCH CLUSTER_BATCH  
$ DEFINE/SYSTEM/EXEC SYS$PRINT CLUSTER_PRINT  
$!  
$! Determine the name of the computer executing this procedure.  
$!  
$ NODE = F$GETSYI("NODENAME")  
$!  
$! Check for full or partial startup. ③  
$!  
$ IF F$GETQUI("DISPLAY_QUEUE","QUEUE_STOPPED","CLUSTER_BATCH") -  
    .EQS. "FALSE" .AND. P1 .NES. "ALL" -  
    THEN GOTO PARTIAL_STARTUP  
$!  
$FULL_STARTUP: ④  
$!  
$! Do a full startup. Define printer forms and characteristics.  
$! Initialize and start queues for all computers and then  
$! initialize and start all generic queues.  
$!  
$ GOSUB DEFINE_PRINTERS  
$ GOSUB JUPITR_STARTUP  
$ GOSUB SATURN_STARTUP  
$ GOSUB URANUS_STARTUP  
$ GOSUB SATELLITE_STARTUP  
$ GOSUB INITIALIZE_GENERIC  
$ GOTO ENDING  
$!  
$DEFINE_PRINTERS: ⑤  
$!  
$! Define all printer forms and characteristics.  
$!  
$ DEFINE/FORM PRETTY 101  
$ DEFINE/CHARACTERISTIC 2ND_FLOOR 2  
.  
.  
.  
$ RETURN
```

(continued on next page)

Setting Up and Managing Cluster Queues

4.4 Using a Common Command Procedure to Set Up Cluster Queues

Example 4-1 (Cont.) Common Procedure to Set Up VAXcluster Queues

```
$!  
$! Intialize and start local printer and batch queues.  
$!  
$JUPITR_STARTUP: ⑥  
$!  
$! Startup Batch/Print for JUPITR::  
$!  
$ IF NODE .EQS. "JUPITR"  
$ THEN  
$   STARTER := /START  
$   SET PRINTER/PAGE=66 LPA0:  
$ ELSE  
$   STARTER := /NOSTART  
$ ENDIF  
$ INITIALIZE/QUEUE/ON=JUPITR::LPA0: JUPITR_PRINT 'STARTER'  
$ INITIALIZE/QUEUE/BATCH JUPITR_BATCH 'STARTER'  
$ RETURN  
$!  
$$SATURN_STARTUP:  
$!  
$! Startup Batch/Print for SATURN::  
.  
.  
.  
$ RETURN  
$!  
$URANUS_STARTUP:  
$!  
$! Startup Batch/Print for URANUS::  
.  
.  
.  
$ RETURN  
$!  
$$SATELLITE_STARTUP: ⑦  
$!  
$! Initialize and start a batch queue for satellites.  
$!  
$ INITIALIZE/QUEUE/BATCH/START 'NODE' _BATCH  
$ RETURN  
$!  
$INITIALIZE_GENERIC: ⑧  
$!  
$! Initialize generic queues only on full startup.  
$! Do CLUSTER_BATCH last. If procedure does not complete  
$! on this computer, another will assume the task.  
$!  
$ INITIALIZE/QUEUE/START CLUSTER_PRINT -  
  /GENERIC=(JUPITR_PRINT,SATURN_PRINT,URANUS_PRINT)  
$ INITIALIZE/QUEUE/BATCH/START CLUSTER_BATCH -  
  /GENERIC=(JUPITR_BATCH,SATURN_BATCH,URANUS_BATCH)  
$ RETURN
```

(continued on next page)

Setting Up and Managing Cluster Queues

4.4 Using a Common Command Procedure to Set Up Cluster Queues

Example 4-1 (Cont.) Common Procedure to Set Up VAXcluster Queues

```
$!  
$PARTIAL_STARTUP: ⑨  
$!  
$! Do a partial startup: initialize and start only local queues.  
$!  
$ IF NODE .NES. "JUPITR" .AND. NODE .NES. "SATURN" .AND. NODE .NES. "URANUS"  
$ THEN GOSUB SATELLITE_STARTUP  
$ ELSE GOSUB 'NODE'_STARTUP  
$ ENDIF  
$ GOTO ENDING  
$!  
$ENDING: ⑩  
$! Submit a batch job to start up layered products.  
$!  
$ SUBMIT/NOIDENT/NOLOG LAYERED_PRODUCT.COM /QUEUE='NODE'_BATCH  
$ EXIT
```

Following are brief descriptions of each phase of the common STARTQ.COM command procedure:

- ① Start the job controller. Specify an appropriate buffer size and the location of the queue file.
- ② The logical names point to all computer-specific SYS\$BATCH and SYS\$PRINT queues. Use of these logical names allows jobs to be submitted to any available batch and print queues in the cluster.
- ③ Because CLUSTER_BATCH is usually stopped only if the entire cluster is shut down, do a full startup if this queue is stopped. Also do a full startup if the parameter ALL is specified. System managers can use this parameter to implement clusterwide changes without rebooting the cluster.
- ④ A full startup assumes that the queue file is empty. All printer forms and characteristics must be defined, all local and clusterwide generic queues must be initialized, and all generic queues must be started. Local queues are started on each computer executing this procedure.
- ⑤ Define all printer forms and characteristics. Make all changes or additions here.
- ⑥ Each computer with specific batch and print duties executes its own subroutine. For example, when this procedure executes on JUPITR, the STARTER symbol is set so that JUPITR's local queues are started when initialized. However, when the procedure executes on another computer, the symbol is set so that JUPITR's queues are initialized but not started on that computer. Thus, each computer initializes all queues but starts only its own.
- ⑦ Initialize and start a local batch queue on any computer without specific batch or print duties. This queue is not included in the clusterwide generic queue scheme.

Setting Up and Managing Cluster Queues

4.4 Using a Common Command Procedure to Set Up Cluster Queues

- ⑧ Generic queues must be initialized and started only on a full startup. This operation is performed last, because starting `CLUSTER_BATCH` indicates to any computer booting into the cluster that the queue file has been set up. At that point, the booting computer needs to perform only a partial startup.
- ⑨ On a partial startup, it is assumed that the clusterwide batch and print system is operating. The computer executing this procedure must initialize and start only its own local queues.
- ⑩ A batch job to start up layered products can be submitted to a computer's local batch queue.

5

Building and Maintaining the Cluster

Before you attempt to build your cluster, be sure you have read the previous chapters and made the following preparations:

- Determined the VAXcluster configuration type (CI-only, local area, or mixed-interconnect)
- Determined whether you want a common-environment or multiple-environment VAXcluster system
- Determined how you will set up and distribute the startup and system files that define the operating environment
- Planned your disk and queue configurations
- Installed or upgraded the VMS operating system on the first VAXcluster computer and installed any required licenses
- Configured and started the DECnet-VAX network

Once you have made these preparations, you can use the information in this chapter to build and maintain your cluster. Topics include the following:

- CLUSTER_CONFIG.COM functions
- Determining locations and sizes for satellite page and swap files
- Specifying allocation classes in mixed-interconnect clusters
- Configuring the cluster
- Reconfiguring the cluster after a major change
- Maintaining the cluster

5.1

CLUSTER_CONFIG.COM Functions

When you invoke CLUSTER_CONFIG.COM, the procedure displays a menu of configuration options. By selecting the appropriate option, you can configure the cluster easily and reliably, without invoking VMS utilities directly. You use CLUSTER_CONFIG.COM to perform these functions:

- Add a computer to the cluster
- Remove a computer from the cluster
- Change a computer's characteristics
- Create a duplicate system disk

Building and Maintaining the Cluster

5.1 CLUSTER_CONFIG.COM Functions

Table 5–1 summarizes the operations that CLUSTER_CONFIG.COM performs for each configuration function.

Table 5–1 Summary of CLUSTER_CONFIG.COM Functions

Function	Operations Performed
ADD	<p>Establish the new computer's root directory on a cluster common system disk and generate the computer's system parameter files (VAXVMSSYS.PAR and MODPARAMS.DAT) in its SYS\$SPECIFIC:[SYSEXE] directory.</p> <p>Update the permanent and volatile remote node network databases for the computer on which CLUSTER_CONFIG.COM is executed to add the new computer. If the new computer is a satellite, update SYS\$MANAGER:NETNODE_UPDATE.COM on the local computer (see Section 5.5.1.2).</p> <p>Generate the new computer's page and swap files (PAGEFILE.SYS and SWAPFILE.SYS).</p> <p>Optionally set up a cluster quorum disk.</p> <p>Set allocation class (ALLOCLASS) value for the new computer, if the computer is being added as a disk server.</p> <p>Generate an initial (temporary) startup procedure for the new computer. This initial procedure runs NETCONFIG.COM to configure the network, runs AUTOGEN to set appropriate SYSGEN parameter values for the computer, and reboots the computer with normal startup procedures.</p>
REMOVE	<p>Delete another computer's root directory and its contents from the local computer's system disk. If the computer being removed is a satellite, update SYS\$MANAGER:NETNODE_UPDATE.COM on the local computer.</p> <p>Update the permanent and volatile remote node network databases on the local computer.</p>
CHANGE	<p>Enable or disable the local computer as a disk server; enable or disable the local computer as a boot server; enable or disable the Ethernet for cluster communications on the local computer; enable or disable a quorum disk on the local computer; change the local computer's ALLOCLASS value; change a satellite's Ethernet hardware address. Procedure displays CHANGE menu and prompts for appropriate information.</p>
CREATE	<p>Duplicate the local computer's system disk and remove all system roots from the new disk.</p>

If you intend to set up a local area or mixed-interconnect cluster, you must, before executing CLUSTER_CONFIG.COM, do the following:

- Determine locations and sizes for satellite page and swap files
- Select cluster boot servers and disk servers
- Determine allocation classes for computers and disks (also applicable for CI-based configurations)

Guidelines are provided in Section 5.2, Section 5.3, and Section 5.4, respectively.

Building and Maintaining the Cluster

5.1 CLUSTER_CONFIG.COM Functions

Note that some configuration functions, such as adding or removing a **voting member** (a computer with a nonzero value for the SYSGEN parameter VOTES), or enabling or disabling a quorum disk, require one or more additional operations. Refer to Section 5.6 for instructions.

5.2 Determining Locations and Sizes for Satellite Page and Swap Files

When you add a computer to the cluster, CLUSTER_CONFIG.COM prompts for the sizes and location of the computer's page and swap files. (The default sizes supplied by the procedure are minimums.) Depending on the configuration of your VAXcluster system disk and your network, you may realize a performance improvement in local area and mixed-interconnect configurations by locating page and swap files for satellites on a satellite's local disk, if such a disk is available.

To set up page and swap files on a satellite's local disk, CLUSTER_CONFIG.COM creates (in the satellite's [SYSx.SYSEX] directory on the boot server's system disk) the command procedure SATELLITE_PAGE.COM. This procedure executes when AUTOGEN reboots the satellite at the end of CLUSTER_CONFIG.COM. SATELLITE_PAGE.COM performs the following functions:

- Mounts the satellite's local disk with a volume label in the format *node-name_SCSSYSTEMID*
- Installs the page and swap files on the satellite's local disk

If you want to alter the volume label, follow these steps after the satellite has been added to the cluster:

- 1 Log in as system manager and enter a DCL command in the following format:

```
SET VOLUME/LABEL=volume-label device-spec[:]
```

Note that the SET VOLUME command requires write access (W) to the index file on the volume. If you are not the volume's owner, you must have either a system UIC or the SYSPRV privilege.

- 2 Update SATELLITE_PAGE.COM to reflect the new label.

To relocate the satellite's page and swap files (for example, from the satellite's local disk to the boot server's system disk, or the reverse), or to change file sizes, the easiest way is to remove the satellite from the cluster and then add it again, using CLUSTER_CONFIG.COM.

5.3 Selecting Boot and Disk Servers

While every local area and mixed-interconnect cluster must include at least one boot server, multiple boot and disk servers offer the following advantages:

- Higher availability—Satellites can access served disks and boot, even if one of the servers is temporarily unavailable.

Building and Maintaining the Cluster

5.3 Selecting Boot and Disk Servers

- Better workload balancing—The task of serving disks to satellites can place a significant load on a server. With multiple servers, this workload is distributed across more computers and Ethernet adapters.

As a general rule, use as boot and disk servers the most powerful computers in the cluster. Low-powered computers can become overloaded when serving many busy satellites, or when many satellites boot simultaneously. Note, however, that two or more moderately powered servers can provide better performance than a single high-powered server. Multiple servers give better availability, and they distribute the workload across more Ethernet adapters. If you have several computers of roughly comparable power, it is reasonable to use them all as boot servers. This arrangement gives optimal load balancing. In addition, if one computer fails or is shut down, others remain available to serve satellites.

After compute power, the most important factor in selecting a server is the speed of its Ethernet adapter. Servers should be equipped with the highest-bandwidth Ethernet adapters in the cluster.

5.4 Determining Allocation Class Values in Mixed-Interconnect Clusters

Before setting up any mixed-interconnect cluster, you must determine allocation class values for boot servers and HSC subsystems. It is easiest to use the same value for all servers and HSC subsystems; you can arbitrarily choose a number between 1 and 255. Note, however, that to change the allocation class value on any CI-connected computer or HSC subsystem, you must shut down and reboot the entire cluster. (See Section 5.6.)

As explained in Section 3.2, every device allocation class name (in the form `1ddcu`) must be the same for all servers and HSC subsystems that share the devices. For RA-series disks, make sure that all the removable unit plugs on all disks of that allocation class are unique. As long as you have no more than 256 such disks, this is easy to accomplish.

Assume, for instance, that 10 disks are dual pathed between the HSC subsystems VOYGR1 and VOYGR2, and assume that 10 others are dual pathed between the HSC subsystems VIKNG1 and VIKNG2. Provided that all 20 disks have unique unit numbers, you can assign the same allocation class value to all four HSC subsystems.

However, if you have more than 256 HSC-connected disks, you must define unique disk names by using two or more allocation class values for the HSC subsystems. You must also configure one or more computers to serve HSC disks and assign allocation class values accordingly. To perform those operations, you can execute the `CLUSTER_CONFIG.COM CHANGE` function, which is described in Section 5.5.3.

Additionally, you must ensure that all locally connected disks have unique device names. For example, if SATURN and URANUS each have a single-pathed RA81 disk connected to a local BDA controller with unit plug 0, and if both computers have an allocation class value of 1, both RA81 disks receive the same device name (`1DUA0`). Because both disks have the same device name, they appear to VMS software to be the same disk. This

Building and Maintaining the Cluster

5.4 Determining Allocation Class Values in Mixed-Interconnect Clusters

condition can endanger data integrity. You can avoid potential problems by selecting a different unit number for one of the disks.

Note that because fewer unit numbers are available for MASSBUS or UNIBUS disks, fewer unique device names can be defined. To ensure that device names remain unique in your cluster, you may have to relocate such disks or disqualify a computer as a disk server.

5.5 Configuring the Cluster

To perform configuration functions, you execute `CLUSTER_CONFIG.COM`. Before invoking the procedure, be sure to verify the following:

- You are logged in to the system manager's account on an appropriate computer. If you are building a new local area or mixed-interconnect cluster, you must be logged in on a computer that you want to set up as a boot server. If you are adding a satellite, you must be logged in on a boot server. Note that the process privileges `SYSPRV`, `OPER`, `CMKRNL`, `BYPASS`, and `NETMBX` are required, because the procedure performs sensitive system operations.
- The DECnet-VAX network is up and running and all computers are connected to the Ethernet.
- You have at hand the data listed in Table 5-2. Note that some items are configuration specific.
- If your configuration has two or more system disks, you have coordinated cluster common files, as described in Section 2.5.4.

Sections 5.5.1 through 5.5.6 provide examples of typical interactive `CLUSTER_CONFIG.COM` sessions. Section 5.6 describes tasks you must perform after executing `CLUSTER_CONFIG.COM` to make major configuration changes.

Caution: You may not initiate concurrent `CLUSTER_CONFIG.COM` sessions.

Table 5-2 Data Requested by `CLUSTER_CONFIG.COM`

Item	How to Specify or Obtain
Device name of cluster system disk on which root directories will be created.	System manager specifies. Default is logical volume name of <code>SYS\$SYSDEVICE:</code> (for example, <code>DISK\$VAXVMSRL5:</code>).
Computer's root directory name on cluster system disk.	System manager specifies. Name must be of the form <code>SYSx</code> . For CI-connected computers, x is a hexadecimal digit in the range 1 through 9 or A through D (for example, <code>SYS1</code> or <code>SYSA</code>). For satellites, x must be in the range from 10 through <code>FFFF</code> . Procedure supplies valid default.
Computer's DECnet node name.	Network manager supplies. Name must be from 1 to 6 alphanumeric characters and <i>may not</i> include dollar signs (\$) or underscores (_).
Computer's DECnet node address.	Network manager supplies.

(continued on next page)

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Table 5–2 (Cont.) Data Requested by CLUSTER_CONFIG.COM

Item	How to Specify or Obtain
Cluster group number and password if CHANGE is run to enable cluster communications over the Ethernet.	System manager specifies.
If computer is a satellite, satellite's Ethernet hardware address. Address has the form xx-xx-xx-xx-xx. Note that you must include the dashes when you specify a hardware address.	When DECnet-VAX network is running on boot server, proceed as follows: <ul style="list-style-type: none">• For MicroVAX II and VAXstation II satellites, enter the following commands at satellite's console:<pre>>>> B/100 XQ Bootfile: READ_ADDR</pre>• For MicroVAX 2000 and VAXstation 2000 satellites, enter the following commands at successive console-mode prompts:<pre>>>> T 53 2 ?>>> 3 >>> B/100 ES Bootfile: READ_ADDR</pre>If the second prompt appears as 3 ?>>>, press RETURN.• For MicroVAX 3xxx series satellites, enter the following command at satellite's console:<pre>>>> SHOW ETHERNET</pre>
Workstation windowing system.	System manager specifies. Workstation software must be installed before workstation satellites are added. If it is not, the procedure indicates that fact.
Location and sizes of page and swap files.	System manager specifies.
Value for local computer's allocation class (ALLOCLASS) parameter.	System manager specifies.
Physical device name of quorum disk.	System manager specifies.

5.5.1 Adding a Computer to the Cluster

Once you have made the necessary preparations, you can execute CLUSTER_CONFIG.COM to add a new computer to the cluster.

- If you are setting up a CI-only cluster, invoke CLUSTER_CONFIG.COM on an active VAXcluster computer and select the ADD function.
- If you are setting up a new local area or mixed-interconnect cluster, follow these steps:
 - 1 Invoke CLUSTER_CONFIG.COM and execute the CHANGE function described in Section 5.5.3 to enable the local computer as a boot server.

Building and Maintaining the Cluster

5.5 Configuring the Cluster

- 2 After the CHANGE function completes, execute the ADD function to add either CI-connected computers or satellites to the cluster. To add satellites, you must be logged in on a cluster boot server.

While adding computers, you may want to disable broadcast messages to your terminal—the ADD function generates many such messages. To disable the messages, you can enter the DCL command `REPLY /DISABLE=(NETWORK, CLUSTER)`.

Whenever you add a voting member to the cluster, you must, after the ADD function completes, reconfigure the cluster, following instructions in Section 5.6. In addition, if you add a CI-connected computer that boots from a cluster common system disk, you must create a new default bootstrap command procedure for the computer before booting it into the cluster. For instructions, refer to your computer-specific installation and operations guide.

Example 5–1 and Example 5–2 illustrate the use of `CLUSTER_CONFIG.COM` on JUPITR to add, respectively, CI-connected computer SATURN and satellite computer EUROPA to the cluster.

Caution: If either the local or the new computer fails before the ADD function completes, you must, after normal conditions are restored, perform the REMOVE function to erase any invalid data and then restart the ADD function.

Example 5–1 Sample Interactive `CLUSTER_CONFIG.COM` Session to Add a CI-Connected Computer as a Boot Server

```
$ @CLUSTER_CONFIG.COM
```

Cluster Configuration Procedure

```
Use CLUSTER_CONFIG.COM to set up or change a VAXcluster configuration. To ensure that you have the required privileges, invoke this procedure from the system manager's account.
```

```
Enter ? for help at any prompt.
```

1. ADD a node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster node's characteristics.
4. CREATE a second system disk for JUPITR.

```
Enter choice [1]: 
```

```
The ADD function adds a new node to the cluster.
```

```
If the node being added is a voting member, EXPECTED_VOTES in all other cluster members' MODPARAMS.DAT must be adjusted, and the cluster must be rebooted.
```

```
If the new node is a satellite, the network databases on JUPITR are updated. The network databases on all other cluster members must be updated.
```

```
For instructions, see the VMS VAXcluster Manual.
```

(continued on next page)

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Example 5-1 (Cont.) Sample Interactive CLUSTER_CONFIG.COM Session to Add a Clustered Computer as a Boot Server

```
What is the node's DECnet node name? SATURN
What is the node's DECnet address? 2.3
Will SATURN be a satellite [Y]? N
Will SATURN be a boot server [Y]?  RET

    This procedure will now ask you for the device name of SATURN's system root.
    The default device name (DISK$VAXVMSRL5:) is the logical volume name of
    SYS$SYSDEVICE:.

What is the device name for SATURN's system root [DISK$VAXVMSRL5:]?  RET
What is the name of the new system root [SYSA]?  RET
Creating directory tree SYSA...
%CREATE-I-CREATED, $1$DJA11:<SYSA> created
%CREATE-I-CREATED, $1$DJA11:<SYSA.SYSEXE> created
.
.
.
System root SYSA created.
Enter a value for SATURN's ALLOCLASS parameter: 1
Does this cluster contain a quorum disk [N]? Y
What is the device name of the quorum disk? $1$DJA12
Updating network database...
Size of page file for SATURN [10000 blocks]? 50000
Size of swap file for SATURN [8000 blocks]? 20000
Will a local (non-HSC) disk on SATURN be used for paging and swapping? N

    If you specify a device other than DISK$VAXVMSRL5: for SATURN's
    page and swap files, this procedure will create PAGEFILE_SATURN.SYS
    and SWAPFILE_SATURN.SYS in the <SYSEXE> directory on the device you
    specify.

What is the device name for the page and swap files [DISK$VAXVMSRL5:]?  RET
%SYSGEN-I-CREATED, $1$DJA11:<SYSA.SYSEXE>PAGEFILE.SYS;1 created
%SYSGEN-I-CREATED, $1$DJA11:<SYSA.SYSEXE>SWAPFILE.SYS;1 created
The configuration procedure has completed successfully.
    SATURN has been configured to join the cluster.

    Before booting SATURN, you must create a new default bootstrap
    command procedure for SATURN. See your processor-specific
    installation and operations guide for instructions.

    The first time SATURN boots, NETCONFIG.COM and
    AUTOGEN.COM will run automatically.

    The following parameters have been set for SATURN:

        VOTES = 1
        QDSKVOTES = 1

    After SATURN has booted into the cluster, you must increment
    the value for EXPECTED_VOTES in every cluster member's
    MODPARAMS.DAT. You must then reconfigure the cluster, using the
    procedure described in the VMS VAXcluster Manual.
```

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Example 5-2 Sample Interactive CLUSTER_CONFIG.COM Session to Add a Satellite with Local Page and Swap Files

\$ @CLUSTER_CONFIG.COM

Cluster Configuration Procedure

Use CLUSTER_CONFIG.COM to set up or change a VAXcluster configuration. To ensure that you have the required privileges, invoke this procedure from the system manager's account.

Enter ? for help at any prompt.

1. ADD a node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster node's characteristics.
4. CREATE a second system disk for JUPITR.

Enter choice [1]:

The ADD function adds a new node to the cluster.

If the node being added is a voting member, EXPECTED_VOTES in all other cluster members' MODPARAMS.DAT must be adjusted, and the cluster must be rebooted.

If the new node is a satellite, the network databases on JUPITR are updated. The network databases on all other cluster members must be updated.

For instructions, see the VMS VAXcluster Manual.

What is the node's DECnet node name? EUROPA

What is the node's DECnet address? 2.21

Will EUROPA be a satellite [Y]?

Verifying circuits in network database...

This procedure will now ask you for the device name of EUROPA's system root. The default device name (DISK\$VAXVMSRL5:) is the logical volume name of SYSSYSDEVICE:.

What is the device name for EUROPA'S system root [DISK\$VAXVMSRL5:]?

What is the name of the new system root [SYS10]?

Allow conversational bootstraps on EUROPA [NO]?

The following workstation windowing options are available:

1. No workstation software
2. VWS Workstation Software
3. DECwindows Workstation Software

Enter choice [1]: 3

(continued on next page)

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Example 5-2 (Cont.) Sample Interactive CLUSTER_CONFIG.COM Session to Add a Satellite with Local Page and Swap Files

```
Creating directory tree SYS10...
%CREATE-I-CREATED, $1$DJAl1:<SYS10> created
%CREATE-I-CREATED, $1$DJAl1:<SYS10.SYSEXE> created
.
.
System root SYS10 created.
Will EUROPA be a disk server [N]? 
What is EUROPA's Ethernet hardware address? 08-00-2B-03-51-75
Updating network database...
Size of pagefile for EUROPA [10000 blocks]? 20000
Size of swap file for EUROPA [8000 blocks]? 12000
Will a local disk on EUROPA be used for paging and swapping? YES
Creating temporary page file in order to boot EUROPA for the first time...
%SYSGEN-I-CREATED, $1$DJAl1:<SYS10.SYSEXE>PAGEFILE.SYS;1 created

    This procedure will now wait until EUROPA joins the cluster.

    Once EUROPA joins the cluster, this procedure will ask you
    to specify a local disk on EUROPA for paging and swapping.

    Please boot EUROPA now.

Waiting for EUROPA to boot...
.
.
.
(User enters boot command at satellite's console-mode prompt (>>>)).
For MicroVAX II, VAXstation II, and MicroVAX 3xxx series satellites, user enters B XQ.
For MicroVAX 2000 and VAXstation 2000 satellites, user enters B ES.)
.
.
.
The local disks on EUROPA are:

Device          Device      Error      Volume      Free      Trans      Mnt
Name            Status      Count      Label       Blocks   Count     Cnt
EUROPA$DUA0:    Online      0          Label       0         0         0
EUROPA$DUA1:    Online      0          Label       0         0         0

Which disk can be used for paging and swapping? EUROPA$ DUA0:
May this procedure INITIALIZE EUROPA$DUA0: [YES]? NO
Mounting EUROPA$DUA0:...
PAGEFILE.SYS already exists on EUROPA$DUA0:
*****
Directory EUROPA$DUA0:[SYS0.SYSEXE]
PAGEFILE.SYS;1      23600/23600
Total of 1 file, 23600/23600 blocks.
*****
What is the file specification for the page file on
EUROPA$DUA0: [ <SYS0.SYSEXE>PAGEFILE.SYS ]? 
```

(continued on next page)

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Example 5-2 (Cont.) Sample Interactive CLUSTER_CONFIG.COM Session to Add a Satellite with Local Page and Swap Files

```
%CREATE-I-EXISTS, EUROPA$DUA0:<SYS0.SYSEXE> already exists
This procedure will use the existing pagefile,
EUROPA$DUA0:<SYS0.SYSEXE>PAGEFILE.SYS;.

SWAPFILE.SYS already exists on EUROPA$DUA0:
*****
Directory EUROPA$DUA0:[SYS0.SYSEXE]
SWAPFILE.SYS;1      12000/12000
Total of 1 file, 12000/12000 blocks.
*****

What is the file specification for the swap file on
EUROPA$DUA0: [ <SYS0.SYSEXE>SWAPFILE.SYS ]? 
This procedure will use the existing swapfile,
EUROPA$DUA0:<SYS0.SYSEXE>SWAPFILE.SYS;.

    AUTOGEN will now reconfigure and reboot EUROPA automatically.
    These operations will complete in a few minutes, and a
    completion message will be displayed at your terminal.

The configuration procedure has completed successfully.
```

5.5.1.1 Updating Network Data After Adding a Satellite

Whenever you add a satellite, CLUSTER_CONFIG.COM updates both the permanent and volatile remote node network databases on the boot server. However, the volatile databases on other cluster members are not automatically updated. To share the new data throughout the cluster, you must update the volatile databases on all other cluster members. Log in as system manager, invoke the SYSMAN Utility, and enter the following commands at the SYSMAN> prompt:

```
SYSMAN> SET ENVIRONMENT/CLUSTER
%SYSMAN-I-ENV, current command environment:
    Clusterwide on local cluster
    Username LAZARUS      will be used on nonlocal nodes
SYSMAN> SET PROFILE/PRIVILEGES=(OPER,SYSPRV)
SYSMAN> DO MCR NCP SET KNOWN NODES ALL
%SYSMAN-I-OUTPUT, command execution on node X...
.
.
.
SYSMAN> EXIT
$
```

Note that the file NETNODE_REMOTE.DAT must be located in the directory SYS\$COMMON:[SYSEXE].

Building and Maintaining the Cluster

5.5 Configuring the Cluster

5.5.1.2 Restoring a Satellite's Network Data

The first time you execute `CLUSTER_CONFIG.COM` to add a satellite, the procedure creates the file `NETNODE_UPDATE.COM` in the boot server's `SYS$SPECIFIC:[SYSMGR]` directory. (For a common-environment cluster, you must rename this file to the `SYS$COMMON:[SYSMGR]` directory, as described in Section 2.5.4.) This file, which is updated each time you add or remove a satellite, or change its Ethernet hardware address, contains all essential network configuration data for the satellite. If an unexpected condition at your site should cause configuration data to be lost, you can use `NETNODE_UPDATE.COM` to restore it. You can also read the file when you need to obtain data about individual satellites. Note that you may want to edit the file occasionally to remove obsolete entries.

Example 5-3 shows the contents of the file after satellites `EUROPA` and `GANYMD` have been added to the cluster.

Example 5-3 Sample `NETNODE_UPDATE.COM` File

```
$ run sys$system:ncp
  define node EUROPA address 2.21
  define node EUROPA hardware address 08-00-2B-03-51-75
  define node EUROPA load assist agent sys$share:niscs_laa.exe
  define node EUROPA load assist parameter $1$DJAI1:<SYS10.>
  define node EUROPA tertiary loader sys$system:tertiary_vmb.exe
  define node GANYMD address 2.22
  define node GANYMD hardware address 08-00-2B-03-58-14
  define node GANYMD load assist agent sys$share:niscs_laa.exe
  define node GANYMD load assist parameter $1$DJAI1:<SYS11.>
  define node GANYMD tertiary loader sys$system:tertiary_vmb.exe
```

5.5.1.3 Controlling Clusterwide Broadcast Messages on Satellites and Boot Servers

When a satellite joins the cluster, broadcasts for all message classes are initially enabled for the satellite by default. Users can disable such broadcasts selectively by including a form of the DCL command `SET BROADCAST` in their `LOGIN.COM` files. For example, the following command would disable `OPCOM` and shutdown messages:

```
$ SET BROADCAST=(NOOPCOM,NOSHUTDOWN)
```

Note that broadcasts to the operator console terminal (`OPA0:`) on satellite workstation computers are disabled by default and should remain disabled at all times. Users who want to receive broadcast messages can create a terminal window and then enter the DCL command `REPLY/ENABLE`. (This command requires `OPER` privilege.) For more detailed information on workstation operations, refer to the documentation supplied with the workstation software.

In large clusters, state transitions (computers joining or leaving the cluster) generate many multiline `OPCOM` messages on a boot server's console device. You can abbreviate such messages by including the DCL command `REPLY/DISABLE=CLUSTER` in the appropriate site-specific startup command file or by entering the command interactively from the system manager's account.

5.5.2 Removing a Computer from the Cluster

You must shut down a computer before removing it from the cluster. If possible, use the command procedure SYS\$SYSTEM:SHUTDOWN.COM to perform an orderly shutdown. Otherwise, halt the computer.

Note that because the REMOVE function deletes the computer's entire root directory tree, it generates VMS RMS error messages while deleting directory files. You can ignore these messages.

Whenever you remove a voting member, you must, after the REMOVE function completes, reconfigure the cluster, following instructions in Section 5.6.

Example 5-4 illustrates the use of CLUSTER_CONFIG.COM on JUPITR to remove satellite EUROPA from the cluster.

Note: If the page and swap files for the computer being removed do not reside on the same disk as the computer's root directory tree, the REMOVE function does not delete these files. It displays a message warning that the files will not be deleted, as in Example 5-4. If you want to delete the files, you must do so after the REMOVE function completes.

Example 5-4 Sample Interactive CLUSTER_CONFIG.COM Session to Remove a Satellite with Local Page and Swap Files

```

$ @CLUSTER_CONFIG.COM

                Cluster Configuration Procedure

Use CLUSTER_CONFIG.COM to set up or change a VAXcluster configuration.
To ensure that you have the required privileges, invoke this procedure
from the system manager's account.

Enter ? for help at any prompt.

    1. ADD a node to the cluster.
    2. REMOVE a node from the cluster.
    3. CHANGE a cluster node's characteristics.
    4. CREATE a second system disk for JUPITR.

Enter choice [1]: 2

The REMOVE function disables a node as a cluster member.

    o It deletes the node's root directory tree.
    o It removes the node's network information
      from the network database.

If the node being removed is a voting member, you must adjust
EXPECTED_VOTES in each remaining cluster member's MODPARAMS.DAT.
You must then reconfigure the cluster, using the procedure described
in the VMS VAXcluster Manual.

What is the node's DECnet node name?  EUROPA
Verifying network database...
Verifying that SYS10 is EUROPA's root...

```

(continued on next page)

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Example 5-4 (Cont.) Sample Interactive CLUSTER_CONFIG.COM Session to Remove a Satellite with Local Page and Swap Files

```
WARNING - EUROPA's page and swap files will not be deleted.
          They do not reside on $1$DJA11:..

Deleting directory tree SYS10...
%DELETE-I-FILDEL, $1$DJA11:<SYS10>SYSCBI.DIR;1 deleted (1 block)
%DELETE-I-FILDEL, $1$DJA11:<SYS10>SYSERR.DIR;1 deleted (1 block)
.
.
.
System root SYS10 deleted.
Updating network database...
The configuration procedure has completed successfully.
```

5.5.3 Changing a Computer's Characteristics

You select the CHANGE function when you want to accomplish any of the operations described in Table 5-3. When you select this function, CLUSTER_CONFIG.COM displays a menu of CHANGE options. Note that all operations except changing a satellite's Ethernet hardware address must be executed on the computer whose characteristics you want to change.

Before adding computers in a new local area or mixed-interconnect cluster, you must execute the CHANGE function to enable the first installed computer as a boot server (see Example 5-7).

Caution: Whenever you enable or disable disk-serving functions, you must run AUTOGEN with the REBOOT option to reboot the local computer. For all other CHANGE operations (except changing a satellite's hardware address), you must reconfigure the cluster, following instructions in Section 5.6.

Table 5-3 CLUSTER_CONFIG.COM CHANGE Options

Option	Operation Performed
Enable the local computer as a disk server.	Load the MSCP server by setting, in MODPARAMS.DAT, the value of the MSCP_LOAD parameter to 1, and setting an appropriate value for the MSCP_SERVE_ALL parameter.
Disable the local computer as a disk server.	Set MSCP_LOAD to 0.

(continued on next page)

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Table 5–3 (Cont.) CLUSTER_CONFIG.COM CHANGE Options

Option	Operation Performed
Enable the local computer as a boot server.	If you are setting up a local area or mixed-interconnect cluster, you must execute this operation once before you attempt to add computers to the cluster. You thereby enable DECnet MOP service for the Ethernet adapter circuit that the computer uses to service operating system load requests from satellites. When you enable the computer as a boot server, it automatically becomes a disk server (if it is not one already), because it must serve its system disk to satellites.
Disable the local computer as a boot server.	Disable DECnet MOP service for the computer's Ethernet adapter circuit.
Enable the Ethernet for cluster communications on the local computer.	Load the VAXport driver PEDRIVER by setting the value of the NISCS_LOAD_PEA0 parameter to 1 in MODPARAMS.DAT. Create the cluster security database file, SYS\$SYSTEM:[SYSEXE]CLUSTER_AUTHORIZE.DAT, on the local computer's system disk.
Disable the Ethernet for cluster communications on the local computer.	Set NISCS_LOAD_PEA0 to 0.
Enable a quorum disk on the local computer.	In MODPARAMS.DAT, set an appropriate value for the SYSGEN parameter DISK_QUORUM; set the value of QDSKVOTES to 1 (default value).
Disable a quorum disk on the local computer.	In MODPARAMS.DAT, set a blank value for the SYSGEN parameter DISK_QUORUM; set the value of QDSKVOTES to 1.
Change the local computer's allocation class value.	Set a value for the computer's ALLOCLASS parameter in MODPARAMS.DAT.
Change a satellite's Ethernet hardware address.	Change a satellite's hardware address if its Ethernet device needs replacement. Both the permanent and volatile network databases, and NETNODE_UPDATE.COM, are updated on the local computer. <i>You must execute this operation on a computer enabled as a boot server for the satellite.</i>

Note: When CLUSTER_CONFIG.COM sets or changes values in MODPARAMS.DAT, the new values are always appended at the end of the file, so that they override earlier values. You may want to edit the file occasionally and delete lines that specify earlier values.

Example 5–5, Example 5–6, Example 5–7, and Example 5–8, respectively, show the use of CLUSTER_CONFIG.COM to perform the following operations:

- Enable node URANUS as a disk server
- Change node URANUS's ALLOCLASS value
- Enable node URANUS as a boot server

Building and Maintaining the Cluster

5.5 Configuring the Cluster

- Specify a new hardware address for satellite node ARIEL, which boots from URANUS's system disk.

Example 5-5 Sample Interactive CLUSTER_CONFIG.COM Session to Enable the Local Computer as a Disk Server

```
$ @CLUSTER_CONFIG.COM

                Cluster Configuration Procedure

Use CLUSTER_CONFIG.COM to set up or change a VAXcluster configuration.
To ensure that you have the required privileges, invoke this procedure
from the system manager's account.

Enter ? for help at any prompt.

    1. ADD a node to the cluster.
    2. REMOVE a node from the cluster.
    3. CHANGE a cluster node's characteristics.
    4. CREATE a second system disk for URANUS.

Enter choice [1]: 3

CHANGE Menu

    1. Enable URANUS as a disk server.
    2. Disable URANUS as a disk server.
    3. Enable URANUS as a boot server.
    4. Disable URANUS as a boot server.
    5. Enable Ethernet for cluster communications on URANUS.
    6. Disable Ethernet for cluster communications on URANUS.
    7. Enable a quorum disk on URANUS.
    8. Disable a quorum disk on URANUS.
    9. Change URANUS's ALLOCLASS value.
   10. Change a satellite's hardware address.

Enter choice [1]: 

Will URANUS serve HSC disks [Y]? 
Enter a value for URANUS's ALLOCLASS parameter: 2
The configuration procedure has completed successfully.

URANUS has been enabled as a disk server. MSCP_LOAD has been
set to 1 in MODPARAMS.DAT. Please run AUTOGEN to reboot URANUS:

    $ @SYS$UPDATE:AUTOGEN GETDATA REBOOT

If you have changed URANUS's ALLOCLASS value, you must reconfigure the
cluster, using the procedure described in the VMS VAXcluster Manual.
```

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Example 5-6 Sample Interactive CLUSTER_CONFIG.COM Session to Change the Local Computer's ALLOCLASS Value

```
$ @CLUSTER_CONFIG.COM

      Cluster Configuration Procedure

Use CLUSTER_CONFIG.COM to set up or change a VAXcluster configuration.
To ensure that you have the required privileges, invoke this procedure
from the system manager's account.

Enter ? for help at any prompt.

      1. ADD a node to the cluster.
      2. REMOVE a node from the cluster.
      3. CHANGE a cluster node's characteristics.
      4. CREATE a second system disk for URANUS.

Enter choice [1]: 3

CHANGE Menu

      1. Enable URANUS as a disk server.
      2. Disable URANUS as a disk server.
      3. Enable URANUS as a boot server.
      4. Disable URANUS as a boot server.
      5. Enable Ethernet for cluster communications on URANUS.
      6. Disable Ethernet for cluster communications on URANUS.
      7. Enable a quorum disk on URANUS.
      8. Disable a quorum disk on URANUS.
      9. Change URANUS's ALLOCLASS value.
     10. Change a satellite's hardware address.

Enter choice [1]: 9

Enter a value for URANUS's ALLOCLASS parameter [2]: 1
The configuration procedure has completed successfully

If you have changed URANUS's ALLOCLASS value, you must reconfigure the
cluster, using the procedure described in the VMS VAXcluster Manual.
```

Example 5-7 Sample Interactive CLUSTER_CONFIG.COM Session to Enable the Local Computer as a Boot Server

```
$ @CLUSTER_CONFIG.COM

      Cluster Configuration Procedure

Use CLUSTER_CONFIG.COM to set up or change a VAXcluster configuration.
To ensure that you have the required privileges, invoke this procedure
from the system manager's account.

Enter ? for help at any prompt.

      1. ADD a node to the cluster.
      2. REMOVE a node from the cluster.
      3. CHANGE a cluster node's characteristics.
      4. CREATE a second system disk for URANUS.

Enter choice [1]: 3
```

(continued on next page)

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Example 5-7 (Cont.) Sample Interactive CLUSTER_CONFIG.COM Session to Enable the Local Computer as a Boot Server

CHANGE Menu

1. Enable URANUS as a disk server.
2. Disable URANUS as a disk server.
3. Enable URANUS as a boot server.
4. Disable URANUS as a boot server.
5. Enable Ethernet for cluster communications on URANUS.
6. Disable Ethernet for cluster communications on URANUS.
7. Enable a quorum disk on URANUS.
8. Disable a quorum disk on URANUS.
9. Change URANUS's ALLOCLASS value.
10. Change a satellite's hardware address.

Enter choice [1]: 3

Verifying circuits in network database...

Updating permanent network database...

In order to enable or disable DECnet MOP service in the volatile network database, DECnet traffic must be interrupted temporarily.

Do you want to proceed [Y]? RET

Enter a value for URANUS's ALLOCLASS parameter [1]: RET

The configuration procedure has completed successfully.

URANUS has been enabled as a boot server. Disk serving and Ethernet capabilities are enabled automatically. If URANUS was not previously set up as a disk server, please run AUTOGEN to reboot URANUS:

```
$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

If you have changed URANUS's ALLOCLASS value, you must reconfigure the cluster, using the procedure described in the VMS VAXcluster Manual.

Example 5-8 Sample Interactive CLUSTER_CONFIG.COM Session to Change a Satellite's Hardware Address

```
$ @CLUSTER_CONFIG.COM
```

Cluster Configuration Procedure

Use CLUSTER_CONFIG.COM to set up or change a VAXcluster configuration. To ensure that you have the required privileges, invoke this procedure from the system manager's account.

Enter ? for help at any prompt.

1. ADD a node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster node's characteristics.
4. CREATE a second system disk for URANUS.

Enter choice [1]: 3

(continued on next page)

Example 5–8 (Cont.) Sample Interactive CLUSTER_CONFIG.COM Session to Change a Satellite's Hardware Address

```
CHANGE Menu
  1. Enable URANUS as a disk server.
  2. Disable URANUS as a disk server.
  3. Enable URANUS as a boot server.
  4. Disable URANUS as a boot server.
  5. Enable Ethernet for cluster communications on URANUS.
  6. Disable Ethernet for cluster communications on URANUS.
  7. Enable a quorum disk on URANUS.
  8. Disable a quorum disk on URANUS.
  9. Change URANUS's ALLOCLASS value.
 10. Change a satellite's hardware address.

Enter choice [1]: 10

What is the node's DECnet node name?  ARIEL
What is the new hardware address [XX-XX-XX-XX-XX-XX]? 08-00-3B-05-37-78
Updating network database...
The configuration procedure has completed successfully.
```

5.5.4 Changing the Cluster Configuration Type

As your processing needs change, you may want to add satellites to an existing CI-based cluster, or you may want to add CI-connected computers or HSC subsystems to an existing local area cluster. In either case, you can use CLUSTER_CONFIG.COM to convert your existing cluster to a mixed-interconnect configuration.

5.5.4.1 Changing an Existing CI-Only Cluster to a Mixed-Interconnect Configuration

If you want to convert an existing CI-only cluster to a mixed-interconnect configuration, you must enable cluster communications over the Ethernet on all computers, and you must enable one or more computers as boot servers. Proceed as follows:

- 1 Log in as system manager on each computer, invoke CLUSTER_CONFIG.COM, and execute the CHANGE function to enable the Ethernet for cluster communications. *You must perform this operation on all computers.*
- 2 Execute the CHANGE function to enable one or more computers as boot servers.
- 3 Shut down and reboot the cluster, following instructions in Section 5.6.

Building and Maintaining the Cluster

5.5 Configuring the Cluster

5.5.4.2 Changing an Existing Local Area Cluster to a Mixed-Interconnect Configuration

Before performing the operations described in this section, be sure that the computers and HSC subsystems you intend to include in your new mixed-interconnect configuration are correctly installed and checked for proper operation.

The method you use to convert an existing local area cluster to a mixed-interconnect configuration depends on whether your current boot server is a CI-capable computer. Note that the following procedures assume that the system disk containing satellite roots will reside on an HSC disk.

If the boot server is a CI-capable computer, proceed as follows:

- 1 Log in as system manager on the boot server and perform an image backup operation to back up the current system disk to a disk on an HSC subsystem. (For complete information on backup operations, refer to the *VMS Backup Utility Manual*.)
- 2 Modify the computer's default bootstrap command procedure to boot the computer from the HSC disk, following instructions in the appropriate computer-specific installation and operations guide.
- 3 Shut down the cluster. Shut down the satellites first, and then shut down the boot server.
- 4 Boot the boot server from the newly created system disk on the HSC subsystem.
- 5 Reboot the satellites.

If your current boot server is not a CI-capable computer, proceed as follows:

- 1 Shut down the old local area cluster. Shut down the satellites first, and then shut down the boot server.
- 2 Install the VMS operating system on the new CI-connected computer's HSC system disk. When the installation procedure asks if you want to enable the Ethernet for cluster communications, answer YES.
- 3 When the installation completes, log in as system manager and configure and start the DECnet-VAX network, as described in Chapter 2.
- 4 Execute the CLUSTER_CONFIG.COM CHANGE function to enable the computer as a boot server.
- 5 Log in as system manager on the newly added CI-connected computer and execute CLUSTER_CONFIG.COM's ADD function to add the former local area cluster members (including the former boot server) as satellites on the new HSC system disk.

5.5.5 **Converting a Standalone Computer to a VAXcluster Computer**

You execute CLUSTER_CONFIG.COM on a standalone computer to perform the following operations:

- Add the standalone computer to an existing cluster.
- Set up the standalone computer to form a new cluster, if the computer was not set up as a cluster computer during installation of the VMS operating system.

Example 5-9 illustrates the use of CLUSTER_CONFIG.COM on standalone computer PLUTO to convert PLUTO to a cluster boot server.

Example 5-9 Sample Interactive CLUSTER_CONFIG.COM Session to Convert a Standalone Computer to a Cluster Boot Server

```
$ @CLUSTER_CONFIG.COM

                Cluster Configuration Procedure

This procedure sets up this standalone node to join an existing
cluster or to form a new cluster.

What is the node's DECnet node name? PLUTO
What is the node's DECnet address? 2.5
Will the Ethernet be used for cluster communications (Y/N)? Y
Enter this cluster's group number: 3378
Enter this cluster's password:
Re-enter this cluster's password for verification:
Will PLUTO be a boot server [Y]? 
Verifying circuits in network database...
Enter a value for PLUTO's ALLOCLASS parameter: 1
Does this cluster contain a quorum disk [N]? 

AUTOGEN computes the SYSGEN parameters for your configuration
and then reboots the system with the new parameters.
```

5.5.6 **Creating a Duplicate System Disk**

As you continue to add computers to a VAXcluster system with a single common system disk, you may eventually reach the disk's storage or I/O capacity. In that case, you may want to add one or more common system disks to handle the increased load. You can use CLUSTER_CONFIG.COM to set up these disks. Proceed as follows, *after* you have coordinated cluster common files, as described in Section 2.5.4.

- 1 Log in as system manager.
- 2 Place a blank disk in an appropriate drive and spin up the disk.
- 3 Invoke CLUSTER_CONFIG.COM and select the CREATE function. The procedure prompts for the device names of the current and new system disks, as shown in Example 5-10. It then backs up the current system disk to the new one, deletes all directory roots from the new disk, and mounts that disk clusterwide. Note that VMS RMS error messages are displayed while the procedure deletes directory files. You can ignore these messages.

Building and Maintaining the Cluster

5.5 Configuring the Cluster

Example 5-10 Sample Interactive CLUSTER_CONFIG.COM CREATE Session

```
$ @CLUSTER_CONFIG.COM

                Cluster Configuration Procedure

Use CLUSTER_CONFIG.COM to set up or change a VAXcluster configuration.
To ensure that you have the required privileges, invoke this procedure
from the system manager's account.

Enter ? for help at any prompt.

    1. ADD a node to the cluster.
    2. REMOVE a node from the cluster.
    3. CHANGE a cluster node's characteristics.
    4. CREATE a second system disk for JUPITR.

Enter choice [1]: 4

The CREATE function generates a duplicate system disk.

    o It backs up the current system disk to the new system disk.
    o It then removes from the new system disk all system roots.

WARNING - Do not proceed unless you have defined appropriate
          logical names for cluster common files in your
          site-specific startup procedures. For instructions,
          see the VMS VAXcluster Manual.

          Do you want to continue [N]? YES

This procedure will now ask you for the device name of JUPITR's system root.
The default device name (DISK$VAXVMSRL5:) is the logical volume name of
SYS$SYSDEVICE:.

What is the device name of the current system disk [DISK$VAXVMSRL5:]? 
What is the device name for the new system disk? $1$DJA16:
%DCL-I-ALLOC, _$1$DJA16: allocated
%MOUNT-I-MOUNTED, SCRATCH mounted on _$1$DJA16:
What is the unique label for the new system disk [JUPITR_SYS2]? 

Backing up the current system disk to the new system disk...

Deleting all system roots...

                Deleting directory tree SYS1...

%DELETE-I-FILDEL, $1$DJA16:<SYS0>DECNET.DIR;1 deleted (2 blocks)
.
.
.
System root SYS1 deleted.

                Deleting directory tree SYS2...

%DELETE-I-FILDEL, $1$DJA16:<SYS1>DECNET.DIR;1 deleted (2 blocks)
.
.
.
System root SYS2 deleted.

All the roots have been deleted.
%MOUNT-I-MOUNTED, JUPITR_SYS2 mounted on _$1$DJA16:

The second system disk has been created and mounted clusterwide.
Satellites can now be added.
```

Building and Maintaining the Cluster

5.6 Reconfiguring the Cluster After a Major Change

5.6 Reconfiguring the Cluster After a Major Change

Because the following operations affect the integrity of the entire cluster, you must reconfigure the cluster after executing any of them:

- Adding or removing a voting member
- Enabling or disabling the Ethernet for cluster communications
- Enabling or disabling a quorum disk
- Changing allocation class values
- Changing the cluster group number or password (see Section 5.7.7)

In all cases, you must shut down and reboot the entire cluster. Note that if you add or remove a voting member, or if you enable or disable a quorum disk, you must update MODPARAMS.DAT files before shutting down the cluster. To perform reconfiguration tasks, follow instructions in Section 5.6.1 Section 5.6.2, Section 5.6.3, and Section 5.6.4.

5.6.1 Updating MODPARAMS.DAT Files to Adjust Cluster Quorum

Whenever you add or remove a voting cluster member, or whenever you enable or disable a quorum disk, you must edit MODPARAMS.DAT in all other cluster members' [SYSx.SYSEXE] directories and adjust the value for the SYSGEN parameter EXPECTED_VOTES appropriately. For example, if you add a voting member, or if you enable a quorum disk, you must increment the value by the number of votes assigned to the new member (usually 1). If you add a voting member with 1 vote and enable a quorum disk with 1 vote on that computer, you must increment the value by 2.

You must then prepare to shut down and reboot the entire cluster. To ensure that the new values take effect when you reboot, log in on each computer as system manager and run AUTOGEN to propagate the values to the computer's VAXVMSSYS.PAR file. Enter the following command:

```
$ @SYSSUPDATE:AUTOGEN GETDATA SETPARAMS
```

Be sure *not* to specify the SHUTDOWN or REBOOT options.

Caution: Do not perform this operation until you are ready to shut down and reboot the entire cluster. If a computer fails and then reboots with the new parameters, normal cluster operations can be seriously compromised.

5.6.2 Shutting Down the Cluster

After you have run AUTOGEN to set parameter values correctly, you must shut down the entire cluster. Shut down nonvoting members (such as satellites) before shutting down voting members. Log in as system manager on each computer *locally* and enter the following command to perform an orderly shutdown:

```
$ @SYSSSYSTEM:SHUTDOWN
```

Building and Maintaining the Cluster

5.6 Reconfiguring the Cluster After a Major Change

When you are prompted for shutdown options, specify `CLUSTER_SHUTDOWN`. Note that you must run the shutdown procedure and specify this option on each computer. When all computers have reached a point in the procedure where activity is suspended, you must halt each computer at its console. You cannot shut down the entire cluster from one computer. (For more information on the `CLUSTER_SHUTDOWN` option see Section 5.7.5.2.)

5.6.3 Changing Allocation Class Values on HSC Subsystems

If it is necessary to change allocation class values on any HSC subsystem, you must do so while the entire cluster is shut down. For example, to change the allocation class value to 1, Enter a command sequence like the following at the appropriate HSC consoles:

```
CTRL/C
HSC> RUN SETSHO
SETSHO> SET ALLOCATE DISK 1
SETSHO> EXIT
SETSHO-Q Rebooting HSC; Y to continue, CTRL/Y to abort:? Y
```

5.6.4 Rebooting the Cluster

After all HSC subsystems have been set and rebooted, reboot each computer. Watch the console listings for unusual messages or warnings.

Caution: In local area and mixed-interconnect clusters, you must reboot boot servers before rebooting satellites.

Note that several new messages may appear. For example, if you have used the `CLUSTER_CONFIG.COM CHANGE` function to enable cluster communications over the Ethernet, one message reports that the local area VAXcluster security database is being loaded. Then, for every disk-serving computer, another message reports that the MSCP server is being loaded. This message is followed by a list of all the disks being served by that computer. You should verify that all disks are being served in the manner that you specified when you designed the configuration.

5.7 Maintaining the Cluster

Once your cluster is up and running, you can implement routine site-specific maintenance operations—for example, backing up disks or adding user accounts. You should plan to run `AUTOGEN` with the `FEEDBACK` option on a regular basis, as described in Section 5.7.1.

You should also maintain records of current configuration data, especially any changes to hardware or software components. Section 5.7.2 lists items that should be included in your records.

If you are managing a local area or mixed-interconnect cluster, it is important to monitor Ethernet activity. Section 5.7.3 provides information to help you set up a monitoring procedure.

Building and Maintaining the Cluster

5.7 Maintaining the Cluster

From time to time conditions may occur that require the following special maintenance operations:

- Restoring cluster quorum After an unexpected computer failure
- Executing conditional shutdown operations
- Performing security functions in local area and mixed-interconnect clusters

These operations are discussed in Section 5.7.4, Section 5.7.5, and Section 5.7.7, respectively.

5.7.1 Running AUTOGEN with the FEEDBACK Option

AUTOGEN includes a mechanism called **feedback**. This mechanism examines data collected during normal system operations, and it adjusts system parameters on the basis of the collected data whenever you run AUTOGEN with the FEEDBACK option. For example, the system records each instance of a disk server waiting for buffer space to process a disk request. Based on this information, AUTOGEN can automatically size the disk server's buffer pool to ensure that sufficient space is allocated.

DIGITAL strongly recommends that you use the FEEDBACK option. Without FEEDBACK, it is difficult for AUTOGEN to anticipate patterns of resource usage, particularly in complex configurations. Factors such as the number of computers and disks in the cluster, and the types of applications being run, require adjustment of system parameters for optimal performance.

You should therefore run AUTOGEN with FEEDBACK frequently. As a cluster grows, settings for many parameters must be adjusted. The settings AUTOGEN chooses for a cluster with three CI-connected computers and five satellites will no longer be appropriate when you add more computers or satellites. In summary, you should rerun AUTOGEN on a regular basis to compensate for changes in user workloads and whenever you make significant changes in your configuration. For detailed information on AUTOGEN, refer to the *Guide to Setting Up a VMS System*.

5.7.2 Recording Configuration Data

To maintain a VAXcluster system effectively, you must keep accurate records on the current status of all hardware and software components and on any changes made to those components. Changes to cluster components can have a significant effect on the operation of the entire cluster. If a failure occurs, you will need to consult your records when diagnosing problems.

At a minimum, your configuration records should include the following:

- SCSNODE and SYSSYSTEMID parameter values for all computers.
- DECnet names and addresses for all computers.

Building and Maintaining the Cluster

5.7 Maintaining the Cluster

- Current values for cluster-related SYSGEN parameters, especially ALLOCLASS values for HSC subsystems and computers. (Cluster SYSGEN parameters are described in Appendix A.)
- Names and locations of default bootstrap command procedures for all CI-connected computers.
- Names of Ethernet adapter circuits.
- Names of cluster disk and tape devices.
- In local area and mixed-interconnect clusters, Ethernet hardware addresses for satellites.
- Serial numbers of all hardware components.
- Changes to any hardware or software components (including site-specific command procedures) along with dates and times when changes were made.

Maintaining current records for your configuration is necessary both for routine operations and for eventual troubleshooting activities.

5.7.3 **Monitoring Ethernet Activity in Local Area and Mixed-Interconnect Clusters**

In local area and mixed-interconnect clusters, it is important that you monitor Ethernet activity on a regular basis. Using NCP commands like those shown in the following example (where BNA-0 is the line-id of the Ethernet line), you can set up a convenient monitoring procedure to report activity for each 12-hour period. Note that DECnet event logging for event 0.2 (automatic line counters) must be enabled. (For detailed information on DECnet-VAX event logging, refer to the *VMS Network Control Program Manual*.)

```
NCP> DEFINE LINE BNA-0 COUNTER TIMER 43200
NCP> SET LINE BNA-0 COUNTER TIMER 43200
```

Every timer interval (in this case 12 hours) DECnet will create an event that sends counter data to the DECnet event log. If you experience a performance degradation in your cluster, check the event log for increases in counter values that exceed normal variations for your cluster. If all computers show the same increase, there may be a general problem with your Ethernet configuration. If, on the other hand, only one computer shows a deviation from usual values, there is probably a problem with that computer or its Ethernet interface device.

5.7.4 **Restoring Cluster Quorum After an Unexpected Computer Failure**

During the life of a VAXcluster system, computers join and leave the cluster. For example, you may need to add more computers to the cluster to extend the cluster's processing capabilities, or a computer may shut down unexpectedly because of a hardware or fatal software error. The connection management software coordinates these cluster transitions and controls cluster operation.

Building and Maintaining the Cluster

5.7 Maintaining the Cluster

When a computer shuts down unexpectedly, the remaining computers, with the help of the connection manager, reconfigure the cluster, excluding the computer that shut down. The cluster can survive the failure of the computer and continue process operations, as long as the cluster votes total is greater than the cluster quorum value. If the cluster votes total falls below the cluster quorum value, the cluster suspends the execution of all processes.

For process execution to resume, the cluster votes total must be restored to a value greater than or equal to the cluster quorum value. Often, the required votes are added as computers join or rejoin the cluster. However, waiting for a computer to join the cluster and raising the votes value is not always a simple or convenient remedy. An alternative solution, for example, might be to shut down and reboot all the computers with a lower quorum value.

Following the failure of a computer, you may want to run the Show Cluster Utility and examine values for the VOTES, EXPECTED_VOTES, CL_VOTES, and CL_QUORUM fields. (See the *VMS Show Cluster Utility Manual* for a complete description of these fields.) The VOTES and EXPECTED_VOTES fields show the settings for each cluster member; the CL_VOTES and CL_QUORUM fields show the cluster votes total and the current cluster quorum value.

To examine these values, enter the following commands:

```
$ SHOW CLUSTER/CONTINUOUS
COMMAND> ADD VOTES,EXPECTED_VOTES,CL_VOTES,CL_QUORUM
```

Note: If you want to enter SHOW CLUSTER commands interactively, you must specify the /CONTINUOUS qualifier as part of the SHOW CLUSTER command string. If you do not specify this qualifier, SHOW CLUSTER displays cluster status information returned by the DCL command SHOW CLUSTER and returns you to the DCL command level.

If the display from the Show Cluster Utility shows the CL_VOTES value equal to the CL_QUORUM value, the cluster cannot survive the failure of any remaining voting member. If one of these computers shuts down, all process activity in the cluster stops.

To prevent the disruption of cluster process activity, you can lower the cluster quorum value. You can use the DCL command SET CLUSTER /EXPECTED_VOTES to adjust the cluster quorum to a value you specify. If you do not specify a value, the operating system calculates an appropriate value for you. You need enter the command on only one computer to propagate the new value throughout the cluster. When you enter the command, the operating system reports the new value.

Normally, you use the SET CLUSTER/EXPECTED_VOTES command only when a computer is leaving the cluster for an extended period. (For more information on this command, see the *VMS DCL Dictionary*.)

If, for example, you want to change expected votes to set the cluster quorum to 2, enter the following command:

```
$ SET CLUSTER/EXPECTED_VOTES=3
```

Building and Maintaining the Cluster

5.7 Maintaining the Cluster

The resulting value is $(3 + 2) / 2 = 2$.

Note that no matter what value you specify for the SET CLUSTER /EXPECTED_VOTES command, you cannot increase quorum to a value that is greater than the number of the votes present, nor can you reduce quorum to a value that is half or fewer of the votes present.

To make the new value active clusterwide, you must adjust the SYSGEN parameter EXPECTED_VOTES in MODPARAMS.DAT files on each VAXcluster computer, and then reconfigure the cluster, following instructions in Section 5.6.

When a computer that was previously a cluster member is ready to rejoin, you must reset the SYSGEN parameter EXPECTED_VOTES to its original value in MODPARAMS.DAT on all computers and then reconfigure the cluster, following instructions in Section 5.6. You do not need to use the SET CLUSTER/EXPECTED_VOTES command to increase cluster quorum, because the quorum value is increased automatically when the computer rejoins the cluster.

You can also reduce cluster quorum by selecting one of the cluster-related shutdown options described in Section 5.7.5.

5.7.5 Selecting Cluster Shutdown Options

In addition to the default shutdown option NONE, the VMS operating system provides four options for shutting down VAXcluster computers:

- REMOVE_NODE
- CLUSTER_SHUTDOWN
- REBOOT_CHECK
- SAVE_FEEDBACK

These options are described in Section 5.7.5.1, Section 5.7.5.2, Section 5.7.5.3, and Section 5.7.5.4, respectively.

If you do not select any of these options (if you select the default SHUTDOWN option NONE), the SHUTDOWN procedure performs the normal operations for shutting down a standalone computer. If you want to shut down a computer that you expect to rejoin the cluster shortly, you can specify the default option NONE. In that case, cluster quorum is not adjusted because the VMS operating system assumes that the computer will soon rejoin the cluster.

5.7.5.1 The REMOVE_NODE Option

If you want to shut down a computer that you expect will not rejoin the cluster for an extended period, select the REMOVE_NODE option. For example, a computer may be waiting for new hardware, or you may decide that you want to use a computer for standalone operation indefinitely.

When you use the REMOVE_NODE option, the active quorum in the remainder of the cluster is adjusted downward to reflect the fact that the removed computer's votes no longer contribute to the quorum value. The SHUTDOWN procedure readjusts the quorum by issuing the SET

Building and Maintaining the Cluster

5.7 Maintaining the Cluster

CLUSTER/EXPECTED_VOTES command, which is subject to the usual constraints described in Section 5.7.4.

Note that it is still the responsibility of the system manager to change the SYSGEN parameter EXPECTED_VOTES on the remaining VAXcluster computers to reflect the new configuration.

5.7.5.2 The CLUSTER_SHUTDOWN Option

You must run the SHUTDOWN procedure and specify this option on every VAXcluster computer. Each computer suspends activity, just short of shutting down completely, until all other computers in the cluster have reached the same point in the SHUTDOWN procedure. You must then shut down every computer individually by halting each computer at its console. If any one computer is not completely shut down, clusterwide shutdown cannot occur. Instead, operations on all other computers are suspended. Be sure to shut down nonvoting members (such as satellites) before shutting down other computers.

5.7.5.3 The REBOOT_CHECK Option

When you select the REBOOT_CHECK option, the SHUTDOWN procedure checks for the existence of basic system files that are needed to reboot the computer successfully and notifies you if any files are missing. You should replace such files before proceeding. If all files are present, the following informational message appears:

```
%SHUTDOWN-I-CHECKOK, Basic reboot consistency check completed.
```

Note that you can select the REBOOT_CHECK option separately or in conjunction with either the REMOVE_NODE or CLUSTER_SHUTDOWN option. If you select REBOOT_CHECK with one of the other options, you must specify options as a common list.

5.7.5.4 The SAVE_FEEDBACK Option

You select the SAVE_FEEDBACK option to enable AUTOGEN feedback operation. Note that you should select this option only when a computer has been running long enough to reflect your typical workload. For detailed information on AUTOGEN feedback, see the *Guide to Setting Up a VMS System*.

5.7.6 Rebooting a Satellite with an Operating System on a Local Disk

In some circumstances, cluster software reboots satellites automatically. Before booting a satellite, the boot procedures check for the presence of an operating system on the satellite's local disk. If an operating system is found, that "local" operating system—not the VAXcluster operating system—is booted.

If an operating system is installed on a satellite's local disk, you should take one of the following measures before performing any operation that causes an automatic reboot—for example, executing SYSSYSTEM:SHUTDOWN.COM with the REBOOT option or using CLUSTER_CONFIG.COM to add that satellite to the cluster:

- Rename the directory file ddcu:[000000]SYS0.DIR on the local disk to ddcu:[000000]SYSx.DIR (where SYSx is a root other than SYS0,

Building and Maintaining the Cluster

5.7 Maintaining the Cluster

SYSE, or SYSF). Then enter the DCL command SET FILE/REMOVE to remove the old directory entry for the boot image SYSBOOT.EXE:

```
$ RENAME DUA0:[000000]SYS0.DIR DUA0:[000000]SYS1.DIR
$ SET FILE/REMOVE DUA0:[SYSEXE]SYSBOOT.EXE
```

For subsequent reboots of the computer from the local disk, enter a command in the format B/x000000 at the console-mode prompt (>>>). For example:

```
>>> B/10000000
```

- Disable the local disk. For instructions, refer to your computer-specific installation and operations guide. Note that this option is not available if the satellite's local disk is being used for paging and swapping.

5.7.7 Performing Security Functions in Local Area and Mixed-Interconnect Clusters

Because multiple local area and mixed-interconnect clusters coexist on a single extended Ethernet LAN, the VMS operating system provides mechanisms to ensure the integrity of individual clusters and to prevent access to a cluster (accidental or deliberate) by an unauthorized computer.

Cluster security mechanisms prevent problems that could otherwise occur under circumstances like the following:

- When setting up a new cluster, the system manager specifies a group number identical to that of an existing cluster on the same Ethernet. (This condition is not as unlikely as it may at first appear, because system managers probably do not assign group numbers randomly.) However, if each cluster's password is unique, the new cluster can form independently.
- A satellite user with access to a local system disk tries to join a cluster by executing a conversational SYSBOOT operation at the satellite's console.

The following mechanisms are designed to help system managers perform security functions:

- A cluster authorization file (SYS\$COMMON:[SYSEXE]CLUSTER_AUTHORIZE.DAT), initialized during installation of the VMS operating system or during execution of the CLUSTER_CONFIG.COM CHANGE function. The file is maintained with the SYSMAN Utility.
- Control of conversational bootstrap operations on satellites.

These mechanisms are discussed in Section 5.7.7.1 and Section 5.7.7.2, respectively.

Building and Maintaining the Cluster

5.7 Maintaining the Cluster

5.7.7.1 Maintaining Cluster Security Data

Security data is maintained in the cluster authorization file, `SYS$COMMON:[SYSEXE]CLUSTER_AUTHORIZE.DAT`, which contains the cluster group number and (in encrypted form) the cluster password. The file is accessible only to users with the `SYSPRV` privilege.

Under normal conditions, you need not alter records in the `CLUSTER_AUTHORIZE.DAT` file interactively. However, if you suspect a security breach, you may want to change the cluster password. In that case, you use the `SYSMAN` Utility to make the change.

Note that if your configuration has multiple system disks, each disk must have a copy of `CLUSTER_AUTHORIZE.DAT`. You must run the utility to update all copies.

Caution: If you change either the group number or password, you must reboot the entire cluster. For instructions, see Section 5.6.

To invoke the `SYSMAN` Utility, log in as system manager on a boot server and enter the following command:

```
$ RUN SYS$SYSTEM:SYSMAN
```

When the utility responds with the `SYSMAN>` prompt, you can enter any of the `CONFIGURATION` commands listed in Table 5–4.

Table 5–4 Summary of SYSMAN CONFIGURATION Commands for Cluster Authorization

Command	Qualifiers	Function
HELP CONFIGURATION SET - CLUSTER_AUTHORIZATION	None	Explains the command's functions.
CONFIGURATION SET - CLUSTER_AUTHORIZATION		Updates the cluster authorization file, <code>CLUSTER_AUTHORIZE.DAT</code> , in the directory <code>SYS\$COMMON:[SYSEXE]</code> . (The SET command creates this file if it does not already exist.)
	/GROUP_NUMBER	Specifies a cluster group number. Group number must be in the range from 1 to 4095 or 61440 to 65535.
	/PASSWORD	Specifies a cluster password. Password may be from 1 to 31 characters in length and may include alphanumeric characters, dollar signs (\$), and underscores (_).
CONFIGURATION SHOW - CLUSTER_AUTHORIZATION	None	Displays the cluster group number.

Example 5–11 illustrates the use of the `SYSMAN` Utility to change the cluster password.

Building and Maintaining the Cluster

5.7 Maintaining the Cluster

Example 5–11 Sample SYSMAN Session to Change the Cluster Password

```
$ RUN SYSSYSTEM:SYSMAN
SYSMAN> SET ENVIRONMENT/CLUSTER
%SYSMAN-I-ENV, current command environment:
    Clusterwide on local cluster
    Username LAZARUS          will be used on nonlocal nodes
SYSMAN> SET PROFILE/PRIVILEGES=SYSPRV
SYSMAN> CONFIGURATION SET CLUSTER_AUTHORIZATION/PASSWORD=newpassword
%SYSMAN-I-CAFOLDGROUP, existing group will not be changed
%SYSMAN-I-CAFREBOOT, cluster authorization file updated
    The entire cluster should be rebooted.
SYSMAN> EXIT
$
```

5.7.7.2 Controlling Conversational Bootstrap Operations for Satellites

When you add a satellite to the cluster using `CLUSTER_CONFIG.COM`, the procedure asks whether you want to allow conversational bootstrap operations for the satellite (default is NO). If you press RETURN, `SYSGEN` parameter `NISCS_CONV_BOOT` in the satellite's `SYSGEN` parameter file remains set to 0 to disable such operations. The parameter file, `VAXVMSSYS.PAR`, resides in the satellite's root directory on a boot server's system disk (*device:[SYSx.SYSEX]*). You can later enable conversational bootstrap operations for a given satellite at any time by setting this parameter to 1.

For example, to enable such operations for a satellite booted from root 10 on device `1DJA11`, you would proceed as follows:

- 1 Log in as system manager on the boot server.
- 2 Invoke the System Generation Utility (`SYSGEN`) and enter the following commands:

```
$ RUN SYSSYSTEM:SYSGEN
SYSGEN> USE $1$DJA11:[SYS10.SYSEX]VAXVMSSYS.PAR
SYSGEN> SET NISCS_CONV_BOOT 1
SYSGEN> WRITE $1$DJA11:[SYS10.SYSEX]VAXVMSSYS.PAR
SYSGEN> EXIT
$
```

The change remains in effect until the next time `AUTOGEN` runs on the satellite or until you reset `NISCS_CONV_BOOT` back to 0.

5.8 Guidelines for Configuring Large Clusters

This section provides guidelines for configuring VAXcluster systems that include many computers—approximately 20 or more—and describes procedures that you might find helpful. Typically, such VAXcluster systems are local area or mixed-interconnect configurations with a large number of satellites. Topics include the following:

- Configuring disk server Ethernet adapters and memory
- Configuring system disks
- Adding computers to an existing cluster

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

- Setting up a new large cluster
- Defining the VAXcluster alias

Note that the recommendations in Section 5.8.1 and Section 5.8.2 can prove beneficial in some clusters with fewer than 20 computers.

5.8.1 **Configuring Disk Server Ethernet Adapters and Memory**

Because disk-serving activity in a large local area or mixed-interconnect VAXcluster system can generate a substantial amount of I/O traffic on the Ethernet, boot and disk servers should use the highest-bandwidth Ethernet adapters in the cluster. In addition, a large local area or mixed-interconnect cluster should include multiple boot and disk servers to enhance availability and to distribute I/O traffic over several Ethernet adapters.

Relatively little memory is required to serve disks. Even busy boot and disk servers probably require no more than one-quarter to one-half megabyte of physical memory for disk serving activity. However, if boot and disk servers must also support timesharing users or run batch queues for the cluster, the servers should be configured with memory appropriate for those additional tasks.

5.8.2 **Configuring System Disks**

Depending on the number of computers to be included in a large cluster, you must evaluate the tradeoffs involved in configuring a single system disk or multiple system disks.

While a single system disk is easier to manage, a large cluster might require more system disk I/O capacity than a single system disk can provide. (Consider using the optional VMS Volume Shadowing Software product to increase disk I/O capacity. For more information on VMS Volume Shadowing Software, see the *VMS Volume Shadowing Manual*.) To achieve satisfactory performance, multiple system disks might be needed. However, you should recognize the increased system management efforts involved in maintaining multiple system disks.

5.8.2.1 **Concurrent User Activity**

In clusters with many workstation satellites, the amount and type of user activity on those satellites (for example, any active batch job or other task created on the workstation by or for the user) influence system disk load and therefore the number of satellites that can be supported by a single system disk. For example, if many users are active or run multiple applications simultaneously, the load on the system disk can be significant. Conversely, in an environment where few users are simultaneously active, or where most users run a single application for extended periods, a single system disk might support a large number of satellites. Note, however, that in these environments significant numbers of I/O requests can be directed to application data disks.

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

This situation is similar to the traditional timesharing model, because the probability is low that most users are simultaneously active at any given time. Thus, while a VAXcluster system can be configured assuming that all users are constantly active, a smaller and less expensive one can be configured for more typical working conditions. The tradeoff is between a more expensive VAXcluster system that handles rare peak loads without performance degradation and a less expensive one that handles most normal activity as well as the more expensive one, but suffers some performance degradation during peak load periods.

Note one difference from the traditional timesharing model. In a timesharing system, the most important shared resource is the processing power of a shared computer. But because each workstation user in a VAXcluster system has a dedicated computer, a user who runs large compute-bound jobs on that dedicated computer does not significantly affect users of other computers in the VAXcluster system.

For clustered workstations, the critical shared resource is a disk server. Thus, if a workstation user runs even a small I/O-intensive job, its effect on other workstations sharing the same disk server might be noticeable.

5.8.2.2 Concurrent Booting Activity

One of the rare times when all VAXcluster computers are simultaneously active is during a cluster reboot—for example, after a power failure. All satellites are waiting to reload the VMS operating system, and as soon as a boot server is available, they begin to boot in parallel. This booting activity places a significant I/O load on the system disk or disks.

For example, Table 5-5 shows system disk I/O activity and elapsed time until login for a single satellite with minimal startup procedures when the satellite is the only one booting. Table 5-6 shows system disk I/O activity and times elapsed between boot server response and login for various numbers of satellites booting from a single system disk. The disk in these examples has a capacity of 40 I/O operations per second.

Note that the numbers in the tables are fabricated and are meant to provide only a generalized picture of booting activity. Elapsed times until login on satellites in any particular cluster depend on the complexity of the site-specific system startup procedures. Computers in clusters with many layered products or site-specific applications require more system disk I/O operations to complete booting operations.

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

Table 5-5 System Disk I/O Activity and Boot Time for Single Satellite

Total I/O Requests to System Disk	Average System Disk I/O Operations per Second	Elapsed Time Until Login (Minutes)
4200	6	12

Table 5-6 System Disk I/O Activity and Boot Times for Multiple Satellites

Number of Satellites	I/Os per Second Requested	I/Os per Second Serviced	Elapsed Time Until Login (Minutes)
1	6	6	12
2	12	12	12
4	24	24	12
6	36	36	12
8	48	40	14
12	72	40	21
16	96	40	28
24	144	40	42
32	192	40	56
48	288	40	84
64	384	40	112
96	576	40	168

While the elapsed times shown in Table 5-6 do not include the time required for the boot server itself to reload, they illustrate that the I/O capacity of a single system disk can be the limiting factor for cluster reboot time.

Note that you can reduce overall cluster boot time by configuring multiple system disks and distributing system roots for computers evenly across those disks. This technique has the advantage of increasing overall system disk I/O capacity but the disadvantage of requiring additional system management effort. For example, layered product installation or VMS operating system upgrades must be repeated once for each system disk.

In a mixed-interconnect VAXcluster system with HSC-connected disks, VMS Volume Shadowing Software can be used to increase the I/O capacity of a single system disk. Installations or updates need only be applied once to a volume-shadowed system disk. For clusters with substantial system disk I/O requirements, you can use multiple system disks, each configured as a shadow set.

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

5.8.2.3 Boot Time Costs

When configuring a VAXcluster system for minimum boot times, consider the following:

- Costs of workstations being unavailable during a cluster reboot
- Hardware costs of additional disk drives
- Cost of VMS Volume Shadowing software, if needed
- System management effort required to maintain multiple system disks
- Probability of power interruptions

Note: Sites with stringent demands for uptime should investigate power conditioning options to minimize power interruption problems.

5.8.2.4 Moving High-Activity Files off System Disks

To reduce I/O activity on system disks, you can move page and swap files for computers off system disks, and you can set up page and swap files for satellites on the satellites' local disks, if such disks are available. You specify the sizes and locations of page and swap files when you run CLUSTER_CONFIG.COM to add computers.

You should also move off the system disk such high-activity files as the following:

```
SYSUAF.DAT      NETPROXY.DAT   RIGHTSLLIST.DAT
JBCSYSQUE.DAT  ACCOUNTNG.DAT  VMSMAIL_PROFILE.DATA
```

To specify the location of the files, follow instructions in Chapter 2.

5.8.2.5 Controlling Dump File Size and Creation

Whether your VAXcluster system uses a single common system disk or multiple system disks, you should plan a strategy to manage dump files. Dump file management is especially important for large clusters with a single system disk.

In the event of a software-detected system failure, each computer normally writes the contents of memory to a full dump file on its system disk for analysis. By default, this full dump file is the size of physical memory plus a small number of pages. If system disk space is limited (as is probably the case if a single system disk is used for a large cluster), you may want to specify that no dump file be created for satellites, or that AUTOGEN create a selective dump file. The selective dump file is typically 30% to 60% the size of a full dump file.

You can control dump file size and creation for each computer by specifying appropriate values for the AUTOGEN symbols DUMPSTYLE and DUMPFIL in the computer's MODPARAMS.DAT file. Dump files are specified as shown in Table 5-7.

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

Table 5-7 AUTOGEN Dump File Symbols

Value Specified	Effect
DUMPSTYLE = 0	Full dump file created (default)
DUMPSTYLE = 1	Selective dump file created
DUMPFIL = 0	No dump file created

Caution: While it is possible to configure computers without dump files, the lack of a dump file can make it difficult or impossible to determine the cause of a system failure.

5.8.2.6 Sharing Dump Files

Another option for saving dump file space is to share a single dump file among multiple computers. This technique makes it possible to analyze isolated computer failures. But dumps are lost if multiple computers fail at the same time or if a second computer fails before you can analyze the first failure. Because boot server failures have a greater impact on cluster operation than failures of other computers, you should configure full dump files on boot servers to help ensure speedy problem analysis.

The VMS operating system attempts to ensure that dump files are not unintentionally shared. However, if you want to share dump files, you can follow these steps:

- 1 Decide whether to use full or selective dump files.
- 2 Determine the size of the largest dump file needed by any satellite.
- 3 Select a satellite whose memory configuration is the largest of any in the cluster and do the following:
 - a. Set DUMPSTYLE = 0 (or DUMPSTYLE = 1) in that satellite's MODPARAMS.DAT file.
 - b. Remove any DUMPFIL symbol from the satellite's MODPARAMS.DAT file.
 - c. Run AUTOGEN on that satellite to create a dump file.
- 4 Rename the dump file to SYS\$COMMON:[SYSEXE]SYSDUMP-COMMON.DMP, or create a new dump file named SYSDUMP-COMMON.DMP in SYS\$COMMON:[SYSEXE].
- 5 For each satellite that is to share the dump file, do the following:
 - a. Create a file synonym entry for the dump file in the system-specific root. For example, to create a synonym for the satellite using root SYS1E, you could include a command like the following in the appropriate startup procedure:

```
$ SET FILE SYS$COMMON:[SYSEXE]SYSDUMP-COMMON.DMP -  
/ENTER=SYS$SYSDEVICE:[SYS1E.SYSEXE]SYSDUMP.DMP
```

- b. Add the following lines to the satellite's MODPARAMS.DAT file:

```
DUMPFIL = 0  
DUMPSTYLE = 0 (or DUMPSTYLE = 1)
```

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

- 6 After a satellite has rebooted, you can delete any SYSDUMP.DMP file in its SYS\$SPECIFIC directory. Note that if the old dump file is deleted before the satellite reboots, the disk space is lost. You can recover the space by entering the DCL command ANALYZE/DISK/REPAIR.

5.8.3 Adding Computers to an Existing Cluster

When a computer is first added to a cluster, SYSGEN parameters that control the computer's system resources are normally adjusted in several steps, as follows:

- 1 CLUSTER_CONFIG.COM sets initial parameters that are adequate to boot the computer in a minimum environment.
- 2 When the computer boots, AUTOGEN runs automatically to size the static operating system (without using any dynamic FEEDBACK data), and the computer reboots into the production environment.
- 3 After the newly added computer has been subjected to typical use for a day or more, you should manually run AUTOGEN with FEEDBACK to adjust parameters for the production environment.
- 4 At regular intervals, and whenever a major change occurs in the cluster configuration or production environment, you should manually run AUTOGEN with FEEDBACK to readjust parameters for the changes.

Because, however, the first AUTOGEN run (initiated by CLUSTER_CONFIG.COM) is performed both in the minimum environment and without FEEDBACK, a newly added computer may be inadequately configured to run in the production environment of some large clusters. For this reason, you might want to implement additional configuration measures like those described in Section 5.8.3.1 and Section 5.8.3.2.

5.8.3.1 Running AUTOGEN with FEEDBACK for Initial Configuration

To ensure that computers are adequately configured for production use when they first join the cluster, you can run AUTOGEN with FEEDBACK automatically as part of the initial boot sequence. While this step adds an additional reboot before the computer can be used for production work, the computer's performance can be substantially improved for the first few days of use.

When a computer first boots into a large cluster, much of the computer's resource utilization is determined by the current cluster configuration. Factors such as the number of computers, the number of disk servers, and the number of disks available or mounted contribute to a fixed minimum resource requirement. Because this minimum does not change with continued use of the computer, FEEDBACK information on the required resources is immediately valid.

Other FEEDBACK information, however, such as that influenced by normal user activity, is not immediately available, because the only "user" has been the system startup process. If AUTOGEN were run with FEEDBACK at this point, some system values might be set too low.

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

By running a simulated user load at the end of the first production boot, you can ensure that AUTOGEN has reasonable FEEDBACK information. The User Environment Test Package (UETP) supplied with the VMS operating system contains a test that simulates such a load. You can run this test (the UETP LOAD phase) as part of the initial production boot and then run AUTOGEN with FEEDBACK before a user is allowed to log in.

To implement this technique, you can create a command file like that in step 1 of the following procedure and submit the file to the computer's local batch queue from the cluster common SYSTARTUP procedure. Your command file conditionally runs the UETP LOAD phase and then reboots the computer with AUTOGEN FEEDBACK.

5.8.3.2 Creating a Command File to Run AUTOGEN with FEEDBACK

As shown in the sample file, UETP lets you specify a typical user load to be run on the computer when it first joins the cluster. The UETP run generates data that AUTOGEN uses to set appropriate SYSGEN values for the computer when rebooting it with FEEDBACK. Note, however, that the default setting for the UETP user load assumes that the computer is used as a timesharing system. This calculation can produce SYSGEN values that might be excessive for a single-user workstation, especially if the workstation has large memory resources. Therefore, you might want to modify the default user load setting, as shown in the sample file.

Follow these steps:

- 1 Create a command file like the following:

```
$!  
$! ***** SYS$COMMON:[SYSMGR]UETP_AUTOGEN.COM *****  
$!  
$! For initial boot only, run UETP LOAD phase and  
$! reboot with AUTOGEN FEEDBACK.  
$!  
$ SET NOON  
$ SET PROCESS/PRIVILEGES=ALL  
$!  
$! Run UETP to simulate a user load for a satellite  
$! with 8 simultaneously active user processes. For a  
$! CI-connected computer, allow UETP to calculate the load.  
$!  
$ LOADS = "8"  
$ IF F$GETDVI("PAA0:", "EXISTS") THEN LOADS = ""  
$ @UETP LOAD 1 'loads'  
$!  
$! Create amarker file to prevent resubmission of  
$! UETP_AUTOGEN.COM at subsequent reboots.  
$!  
$ CREATE SYS$SPECIFIC:[SYSMGR]UETP_AUTOGEN.DONE  
$!  
$! Reboot with AUTOGEN to set SYSGEN values.  
$!  
$ @SYS$UPDATE:AUTOGEN SAVPARAMS REBOOT FEEDBACK  
$!  
$ EXIT
```

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

- 2 Edit the cluster common SYSTARTUP file and add commands like the following at the end of the file. Assume that queues have been started and that a batch queue is running on the newly added computer. Submit UETP_AUTOGEN.COM to the computer's local batch queue:

```
$!  
$ NODE = F$GETSYI("NODE")  
$ IF F$SEARCH ("SYS$SPECIFIC:[SYSMGR]UETP_AUTOGEN.DONE") .EQS. ""  
$ THEN  
$   SUBMIT /NOPRINT /NOTIFY /USERNAME=SYSTEST -  
$     /QUEUE='NODE'_BATCH SYS$MANAGER:UETP_AUTOGEN  
  
$ WAIT_FOR_UETP:  
$ WRITE SYS$OUTPUT "Waiting for UETP and AUTOGEN... ''F$TIME()''"  
$   WAIT 00:05:00.00           ! Wait 5 minutes  
$   GOTO WAIT_FOR_UETP  
$ ENDIF  
$!
```

Note that UETP must be run under the username SYSTEST.

- 3 Execute CLUSTER_CONFIG.COM to add the computer.

When you boot the computer, it runs UETP_AUTOGEN.COM to simulate the user load you have specified and then reboots with AUTOGEN FEEDBACK to set appropriate SYSGEN values.

5.8.4 Setting Up a New Large VAXcluster System

When building a new large cluster, you must be prepared to run AUTOGEN and reboot the cluster several times during the installation. The parameters that AUTOGEN sets for the first computers added to the cluster will probably be inadequate when additional computers are added. Readjustment of parameters is especially critical for boot and disk servers.

One solution to this potential problem is to run UETP_AUTOGEN.COM to reboot computers at regular intervals as new computers are added. You should run the procedure according to the percentage of growth. For example, each time there is a significant percentage increase in the number of computers (from 5 to 10, from 10 to 20, and so forth), you should run UETP_AUTOGEN.COM. For best results, the cluster environment should be as close as possible to the final production environment when you run the procedure.

To set up the cluster, you can follow these steps:

- 1 Configure boot and disk servers, using the CLUSTER_CONFIG.COM command procedure.
- 2 Install all layered products and site-specific applications required for the cluster production environment, or as many as possible.
- 3 Prepare the cluster startup procedures so that they are as close as possible to those to be used in the final production environment.
- 4 Add a small number of satellites (perhaps 2 or 3), using CLUSTER_CONFIG.COM.
- 5 Reboot the cluster to verify that the startup procedures work as expected.

Building and Maintaining the Cluster

5.8 Guidelines for Configuring Large Clusters

- 6 After you have verified that startup procedures work, run UETP_AUTOGEN.COM on every computer's local batch queue to reboot the cluster again and to set initial production environment values. When the cluster has rebooted, all computers should have reasonable parameter settings. However, check the settings to be sure.
- 7 Add additional satellites to double their number, and then rerun UETP_AUTOGEN on each computer's local batch queue to reboot the cluster and set values appropriately to accommodate the newly added satellites.
- 8 Repeat the previous step until all satellites have been added.
- 9 When all satellites have been added, run UETP_AUTOGEN a final time on each computer's local batch queue to reboot the cluster and to set new values for the production environment.

Note that for best performance, you might not want to run UETP_AUTOGEN on every computer simultaneously, because the procedure simulates a user load that is probably more demanding than that for the final production environment. A better method is to run UETP_AUTOGEN on several satellites (those with the least recently adjusted parameters) while adding new computers. This technique increases efficiency, because little is gained when a satellite reruns AUTOGEN shortly after joining the cluster. For example, if the entire cluster is rebooted after 30 satellites have been added, few adjustments are made to system parameter values for the 28th satellite added—only two satellites have joined the cluster since that satellite ran UETP_AUTOGEN as part of its initial configuration.

5.8.5 Defining the VAXcluster Alias

The VAXcluster alias acts as a single network node identifier for a VAXcluster system. Computers in the cluster can use the alias for communications with other computers in a DECnet-VAX network. A maximum of 64 VAXcluster computers can participate in a VAXcluster alias. If your cluster includes more than 64 computers, you must determine which 64 should participate in the alias and then define it on those computers. For detailed information on the VAXcluster alias, refer to the *VMS Networking Manual*.

A

Cluster SYSGEN Parameters

For systems to boot properly into a cluster, certain system parameters must be set on each cluster computer. Table A-1 lists SYSGEN parameters used in cluster configurations.

Table A-1 Cluster SYSGEN Parameters

Parameter	Description
ALLOCLASS	Specifies a numeric value from 0 to 255 to be assigned as the allocation class for the computer. The default value is 0.
DISK_QUORUM	The physical device name, in ASCII, of an optional quorum disk. ASCII spaces indicate that no quorum disk is being used. DISK_QUORUM must be defined on one or more cluster computers capable of having a direct (non-MSCP served) connection to the disk. These computers are called quorum disk watchers . The remaining computers (computers with a blank value for DISK_QUORUM) recognize the name defined by the first watcher computer which which they communicate.
EXPECTED_VOTES	Specifies a setting that is used to derive the initial quorum value. This setting is the sum of all VOTES held by potential cluster members. By default, the value is 1. The connection manager sets a quorum value to a number that will prevent cluster partitioning (see Section 1.5). To calculate quorum, the system uses the following formula: $\text{estimated quorum} = (\text{EXPECTED_VOTES} + 2) / 2$
MSCP_LOAD	Controls whether the MSCP server is loaded. Specify 1 to load the server. By default, the value is set to zero, and the server is not loaded.
MSCP_SERVE_ALL	Specifies MSCP disk-serving functions when the MSCP server is loaded. The default value of zero specifies that no disks are served. A value of 1 specifies that all available disks are served. A value of 2 specifies that only locally connected (non-HSC) disks are served.
NISCS_CONV_BOOT	Specifies whether conversational bootstraps are enabled on the computer. The default value of zero specifies that conversational bootstraps are disabled. A value of 1 enables conversational bootstraps.
NISCS_LOAD_PEA0	Specifies whether the VAXport driver PEDRIVER is to be loaded to enable cluster communications over the Ethernet. The default value of zero specifies that the driver is not loaded. A value of 1 specifies that that driver is loaded.
NISCS_PORT_SERV	Specifies whether data checking is enabled for the computer. The default value of zero specifies that data checking is disabled.
QDSKVOTES	Specifies the number of votes contributed to the cluster votes total by a quorum disk. The maximum is 127, the minimum is 0, and the default is 1. This parameter is used only when DISK_QUORUM is defined.

(continued on next page)

Cluster SYSGEN Parameters

Table A-1 (Cont.) Cluster SYSGEN Parameters

Parameter	Description
QDSKINTERVAL	<p>Specifies the disk quorum polling interval, in seconds. The maximum value is 32767, the minimum value is 1, and the default is 10. Lower values trade increased overhead cost for greater responsiveness.</p> <p>Digital recommends that this parameter be set to the same value on each cluster computer.</p>
RECNXINTERVAL	<p>Specifies, in seconds, the interval during which the connection manager attempts to reconnect a broken connection to another computer. If a new connection cannot be established during this period, the connection is declared irrevocably broken, and either this computer or the other must leave the cluster. This parameter trades faster response to certain types of system failures against the ability to survive transient faults of increasing duration.</p> <p>Digital recommends that this parameter be set to the same value on each cluster computer.</p>
VAXCLUSTER	<p>Controls whether the computer should join or form a cluster. This parameter accepts the following three values:</p> <ul style="list-style-type: none"> • 0—Specifies that the computer will not participate in a cluster. • 1—Specifies that the computer should participate in a cluster if hardware supporting SCS is present (CI, UDA, HSC50). • 2—Specifies that the computer should participate in a cluster <p>You should always set this parameter to 2 on computers intended to run in a cluster, 0 on computers that boot from a UDA and are not intended to be part of a cluster, and 1 (the default) otherwise.</p>
VOTES	<p>Specifies the number of votes towards a quorum to be contributed by the computer. By default, the value is 1.</p>
SCS Parameters	
PANUMPOLL	<p>Specifies the number of ports to poll at each interval. Digital recommends that this parameter be set to the same value on each cluster computer.</p>
PASTIMOUT	<p>Specifies the interval at which the CI port driver performs time-based bookkeeping operations. This interval is also the period after which a start handshake datagram is assumed to have timed out.</p> <p>Normally the default value is adequate. Digital recommends that this parameter be set to the same value on each cluster computer.</p>
PASTDGBUF	<p>Specifies the number of datagram receive buffers to queue for the CI port driver's configuration poller, that is, the maximum number of start handshakes that can be in progress simultaneously.</p> <p>Normally the default value is adequate. Digital recommends that this parameter be set to the same value on each cluster computer.</p>

(continued on next page)

Cluster SYSGEN Parameters

Table A-1 (Cont.) Cluster SYSGEN Parameters

Parameter	Description
PAMAXPORT	<p>Specifies the maximum number of CI ports the CI port driver polls for a broken port-to-port virtual circuit, or a failed remote computer.</p> <p>You can decrease this parameter in order to reduce polling activity if the hardware configuration has fewer than 16 ports. For example, if the configuration has a total of five ports assigned port numbers 0-4, then you should set PAMAXPORT to 4.</p> <p>The default for this parameter is 15 (poll for all possible ports 0 through 15). Digital recommends that this parameter be set to the same value on each cluster computer.</p>
PANOPOLL	<p>Disables CI polling for ports if set to 1. (The default is 0.) When PANOPOLL is set, a computer will not discover that another computer has shut down or powered down promptly and will not discover a new computer that has booted. This parameter is useful when you want to bring up a computer detached from the rest of the cluster for checkout purposes. It is roughly equivalent to uncabing the computer from the star coupler.</p> <p>PANOPOLL = 0 is the normal setting and is required if you are booting from an HSC.</p>
PAPOLLINTERVAL	<p>Specifies in seconds, the polling interval the computer interconnect (CI) port driver uses to poll for a newly booted computer, a broken port-to-port virtual circuit, or a failed remote computer.</p> <p>This parameter trades polling overhead against quick response to virtual circuit failures. Digital recommends that you use default value for this parameter.</p> <p>Digital recommends that this parameter be set to the same value on each cluster computer.</p>
PAPOOLINTERVAL	<p>Specifies in seconds, the interval at which the PA port driver checks for available nonpaged pool after a failure to allocate.</p> <p>Normally the default value is adequate.</p>
PASANITY	<p>Controls whether the port sanity timer is enabled to permit remote computers to detect a computer that has been halted or retained at IPL 7 for a prolonged period. This parameter is normally set to 1 and should only be set to 0 when debugging with XDELTA.</p> <p>PASANITY is a dynamic parameter (altered the next time the port is initialized) and has a default value of 1.</p>
PRCPOLINTERVAL	<p>Specifies, in seconds, the polling interval used to look for SCS applications, such as the connection manager and MSCP disks, on other computers. Each computer is polled, at most, once each interval.</p> <p>This parameter trades polling overhead against quick recognition of new computers or servers as they appear. Digital recommends that you set this parameter to 15, which is the default.</p>
SCSBUFFCNT	<p>Specifies the number of computer interconnect (CI) buffer descriptors configured for all CI ports on the computer.</p>
SCSCONNCNT	<p>Specifies the total number of SCS connections that are configured for use by all system applications.</p> <p>Normally, the default value is adequate.</p>

(continued on next page)

Cluster SYSGEN Parameters

Table A-1 (Cont.) Cluster SYSGEN Parameters

Parameter	Description
SCSMAXMSG	Specifies the SCS maximum sequenced message size. Normally, the default value is adequate.
SCSMAXDG	Specifies the maximum number of bytes of application data in one datagram. Normally the default value is adequate.
SCSFLOWCUSH	Specifies the lower limit for receive buffers at which point SCS starts to notify the remote SCS of new receive buffers. For each connection, SCS tracks the number of receive buffers available. SCS communicates this number to the SCS at the remote end of the connection. However, SCS does not need to do this for each new receive buffer added. Instead, SCS notifies the remote SCS of new receive buffers if the number of receive buffers falls as low as the SCSFLOWCUSH value. Normally the default value is adequate.
SCSSYSTEMID	Specifies the lower-order 32 bits of the 48-bit system identification number. This parameter is not dynamic and must be the same as the DECnet computer number ($1024 * \text{<DECnet area>} + \text{DECnet computer number}$).
SCSSYSTEMIDH	Specifies the high-order 16 bits of the 48 bit system identification number. This parameter must be set to 0. It is reserved by Digital for future use.
SCSNODE	Specifies the SCS system name. This parameter is not dynamic. You should use a name that is the same as the DECnet computer name (limited to six characters) since the name must be unique among all computers in the cluster. Note that once a computer has been recognized by another computer in the cluster, you cannot change the SCSSYSTEMID or SCSNODE parameter without changing both.
SCSRESPCNT	Specifies the total number of response descriptor table entries configured for use by all system applications.

B

Building a Common SYSUAF.DAT File

This appendix provides guidelines for building a common user authorization file from computer-specific files. For more detailed information on how to set up a computer-specific authorization file, see the descriptions in the *VMS Authorize Utility Manual* and in the *Guide to Setting Up a VMS System*.

To build a common SYSUAF.DAT file, follows these steps:

- 1 Print a listing of SYSUAF.DAT on each computer. To print this listing, invoke AUTHORIZE and specify the AUTHORIZE command LIST as follows:

```
$ SET DEF SYSS$SYSTEM
$ RUN AUTHORIZE
UAF> LIST/FULL [*,*]
```

- 2 Use the listings to compare the accounts from each computer. On the listings, mark down any necessary changes.

One such change is to delete any accounts that you no longer need. You should also make sure that each user account in the cluster has a unique UIC.

For example, VAXcluster member VENUS may have a user account JONES that has the same UIC as user account SMITH on computer MARS. When computers VENUS and MARS are joined to form a cluster, accounts JONES and SMITH will exist in the cluster environment with the same UIC. If the UICs of these accounts are not differentiated, each user will have the same access rights to various objects in the cluster. In this case you should assign each account a unique UIC.

Make sure that accounts that perform the same type of work have the same group UIC. Accounts in a single-computer environment probably follow this convention. However, there may be groups of users on each computer that will perform the same work in the cluster but have group UICs unique to their local computer. As a rule, the group UIC for any given work category should be the same on each computer in the cluster. For example, data entry accounts on VENUS should have the same group UIC as data entry accounts on MARS.

Note that if you change the UIC for a particular user, you should also change the owner UICs for that user's existing files and directories. You can use the DCL commands SET FILE and SET DIRECTORY to make these changes. These commands are described in detail in the *VMS DCL Dictionary*.

- 3 Choose the SYSUAF.DAT file from one of the computers to be a master SYSUAF.DAT.

Building a Common SYSUAF.DAT File

- 4 Merge the SYSUAF.DAT files from the other computers to the master SYSUAF.DAT by running the Convert Utility (CONVERT) on the computer that owns the master SYSUAF.DAT. (See the *VMS Convert and Convert/Reclaim Utility Manual* for a description of CONVERT.) To use CONVERT to merge the files, each SYSUAF.DAT file must be accessible to the computer that is running CONVERT.

To merge the UAFs into the master SYSUAF.DAT file, specify the CONVERT command in the following format:

```
CONVERT SYSUAF1,SYSUAF2,...SYSUAFn MASTER_SYSUAF
```

Note that if a given user name appears in more than one source file, only the first occurrence of that name will appear in the merged file.

The command sequence in the following example adds the SYSUAF.DAT file from two VAXcluster computers to the master SYSUAF.DAT in the current default directory:

```
$ SET DEFAULT SYS$SYSTEM
$ CONVERT [SYS1.SYSEXE]SYSUAF.DAT, -
_ $ [SYS2.SYSEXE]SYSUAF.DAT SYSUAF.DAT
```

The CONVERT command in this example adds the records from the files [SYS1.SYSEXE]SYSUAF.DAT and [SYS2.SYSEXE]SYSUAF.DAT to the file SYSUAF.DAT on the local computer.

After you run CONVERT, you have a master SYSUAF.DAT that contains records from the other SYSUAF.DAT files.

- 5 Use AUTHORIZE to modify the accounts in the master SYSUAF.DAT according to the changes you marked on the initial listings of the SYSUAF.DAT files from each computer.

Merging RIGHTSLIST.DAT Files

If you need to merge RIGHTSLIST.DAT files, you can use a command sequence like the following:

```
$ SET DEFAULT SYS$SYSTEM
$ ANALYZE/RMS/FDL RIGHTSLIST.DAT
$ CONVERT/STATISTICS/FDL=RIGHTSLIST -
_ $ [SYS1.SYSEXE]RIGHTSLIST.DAT, [SYS2.SYSEXE]RIGHTSLIST.DAT -
_ $ RIGHTSLIST.DAT
```

The commands in this example add the RIGHTSLIST.DAT files from two VAXcluster computers to the master RIGHTSLIST.DAT file in the current default directory. For detailed information of creating and maintaining RIGHTSLIST.DAT files, refer to the *Guide to VMS System Security*.

C

Cluster Troubleshooting Information

This appendix contains information to help you perform troubleshooting operations for the following:

- Failures of computers to boot or to join the cluster
- Cluster hangs
- CLUEXIT bugchecks
- VAXport device problems

C.1 Diagnosing Failures of Computers to Boot or to Join the Cluster

Before you initiate diagnostic procedures, be sure to verify that these conditions are met:

- All cluster hardware components are correctly connected and checked for proper operation.
- VAXcluster computers and mass storage devices are configured according to requirements specified in the *VAXcluster Software Product Description* (SPD).

When attempting to add a new or recently repaired CI-connected computer to the cluster, you must verify that the CI cables are correctly connected, as described in Section C.4.2.2.

When attempting to add a satellite to a local area or mixed-interconnect cluster, you must verify that the Ethernet is configured according to requirements specified in the VAXcluster SPD, and that the machine's memory resources and Ethernet adapter device meet the requirements specified in that document. You must also verify that you have correctly configured and started the DECnet-VAX network, following the procedures described in Section 2.3.

If after performing preliminary checks and taking appropriate corrective action, you find that a computer still fails to boot or to join the cluster, you can follow the procedures in Sections C.1.2 through C.1.4 to attempt recovery.

C.1.1 Summary of Events for Computers Booting and Joining the Cluster

To perform diagnostic and recovery procedures effectively, you must understand the events that occur when a computer boots and attempts to join the cluster. This section outlines those events and shows typical messages displayed at the console.

Cluster Troubleshooting Information

C.1 Diagnosing Failures of Computers to Boot or to Join the Cluster

Note that events vary, depending on whether a computer is the first to boot in a new cluster or whether it is booting in an active cluster. Note further that some events (such as loading the cluster security database) occur only in local area and mixed-interconnect clusters.

The normal sequence of events is as follows:

- 1 The computer boots. If the computer is a satellite, a message like the following shows the name and Ethernet address of the boot server that has downline loaded the satellite:

```
%VAXcluster-I-SYSLOAD, system loaded from Node X... (XX-XX-XX-XX-XX-XX)
```

For any booting computer, the VMS "banner message" is displayed in the following format:

```
VAX/VMS Version n.n DD-MMM-YYYY hh:mm:ss
```

- 2 The computer attempts to form or join the cluster, and the following message appears:

```
waiting to form or join a VAXcluster system
```

If the computer is a member of a local area or mixed-interconnect cluster, the cluster security database is loaded. Optionally, the MSCP server may be loaded:

```
%VAXcluster-I-LOADSECDB, loading the cluster security database  
%MSCPLOAD-I-LOADMSCP, loading the MSCP disk server
```

- 3 If the computer discovers a cluster, the computer attempts to join. If a cluster is found, the connection manager displays one or more messages in the following format:

```
%CNXMAN, Sending VAXcluster membership request to system X...
```

Otherwise, the connection manager forms the cluster when it has enough votes to establish quorum (that is, when enough voting computers have booted).

- 4 As the booting computer joins the cluster, the connection manager displays a message in the following format:

```
%CNXMAN, now a VAXcluster member -- system X...
```

Note that if quorum is lost while the computer is booting, or if a computer is unable to join the cluster within 2 minutes of booting, the connection manager displays messages like the following:

```
%CNXMAN, Discovered system X...  
%CNXMAN, Deleting CSB for system X...  
%CNXMAN, Established "connection" to quorum disk  
%CNXMAN, Have connection to system X...  
%CNXMAN, Have "connection" to quorum disk
```

The last two messages show any connections that have already been formed.

If the cluster includes a quorum disk, you may also see messages like the following:

```
%CNXMAN, Using remote access method for quorum disk  
%CNXMAN, Using local access method for quorum disk
```

Cluster Troubleshooting Information

C.1 Diagnosing Failures of Computers to Boot or to Join the Cluster

The first message indicates that the connection manager is unable to access the quorum disk directly, either because the disk is unavailable or because it is accessed through the MSCP server. Another computer in the cluster that can access the disk directly must verify that a reliable connection to the disk exists.

The second message indicates that the connection manager can access the quorum disk directly and can supply information about the status of the disk to computers that cannot access the disk directly.

Note that the connection manager may not see the quorum disk initially, because the disk may not yet be configured. In that case, the connection manager first uses remote access, then switches to local access.

- 5 Once the computer has joined the cluster, normal startup procedures execute. One of the first functions is to start the OPCOM process:

```
%%%%%%%%%% OPCOM 15-APR-1990 16:33:55.33 %%%%%%%%%%%
Logfile has been initialized by operator _X...$OPA0:
Logfile is SYS$SYSROOT:[SYSMGR]OPERATOR.LOG;17

%%%%%%%%%% OPCOM 15-APR-1990 16:33:56.43 %%%%%%%%%%%
16:32:32.93 Node X... (csid 0002000E) is now a VAXcluster member
```

When other computers join the cluster, OPCOM displays messages like the following:

```
%%%%%%%%%% OPCOM 15-APR-1990 16:34:25.23 %%%%%%%%%%% (from node X... at 16:34:25.23)
16:34:24.42 Node X... (csid 000100F3) received VAXcluster membership request from node X...
```

As startup procedures continue, various messages report startup events.

Note: For troubleshooting purposes, you may want to include in your site-specific startup procedures messages announcing each phase of the startup process—for example, mounting disks or starting queues.

C.1.2 CI-Connected Computer Fails to Boot

If a CI-connected computer fails to boot, perform the following checks:

- Verify that the computer's SCSNODE and SYSSYSTEMID parameters are unique in the cluster. If they are not, you must either alter *both* values or reboot all other computers.
- Verify that you are using the correct bootstrap command file. This file must specify the internal bus computer number (if applicable), the HSC node number, and the HSC disk from which the computer is to boot. Refer to your processor-specific installation and operations guide for information on setting values in default bootstrap command procedures.
- Verify that the SYSGEN parameter PAMAXPORT is set to a value greater than or equal to the largest CI port number.
- Verify that the HSC subsystem is ONLINE. The ONLINE switch on the HSC operator control panel should be pressed in.

Cluster Troubleshooting Information

C.1 Diagnosing Failures of Computers to Boot or to Join the Cluster

- Verify that the disk is available. The correct port switches on the disk's operator control panel should be pressed in.
- Verify that the computer has access to the HSC subsystem. The SHOW HOSTS command of the HSC SETSHO Utility displays status for all VAX computers (hosts) in the cluster. (For complete information on the SETSHO Utility, consult the HSC hardware documentation.) If the computer in question appears in the display as DISABLED, use the SETSHO Utility to set the computer to the ENABLED state.
- Verify that the HSC subsystem allows access to the boot disk. Invoke the SETSHO Utility to ensure that the boot disk is available to the HSC subsystem. The utility's SHOW DISKS command displays the current state of all disks visible to the HSC subsystem and displays all disks in the no-host-access table. If the boot disk appears in the no-host-access table, use the SETSHO Utility to set the boot disk to host-access. If the boot disk is AVAILABLE or MOUNTED and host-access ENABLED, but does not appear in the no-host-access table, contact your Customer Services representative and explain both the problem and the steps you have taken.

C.1.3 Satellite Fails to Boot

To boot successfully, a satellite must communicate with a boot server over the Ethernet. You can use DECnet event logging to verify this communication. Proceed as follows:

- 1 Log in as system manager on the boot server.
- 2 If event logging for management layer events is not already enabled, enter the following NCP commands to enable it:

```
NCP> SET LOGGING MONITOR EVENT 0.*  
NCP> SET LOGGING MONITOR STATE ON
```

- 3 Enter the following DCL command:

```
$ REPLY/ENABLE=NETWORK
```

This command enables the terminal to receive DECnet messages reporting downline load events.

- 4 Boot the satellite. If the satellite and the boot server can communicate, and if all boot parameters are correctly set, messages like the following are displayed at the boot server's terminal:

```
DECnet event 0.3, automatic line service  
From node 2.4 (URANUS), 15-APR-1990 09:42:15.12  
Circuit QNA-0, Load, Requested, Node = 2.42 (OBERON)  
File = SYS$SYSDEVICE:<SYS10.>, Operating system  
Ethernet address = 08-00-2B-07-AC-03
```

```
DECnet event 0.3, automatic line service  
From node 2.4 (URANUS), 15-APR-1990 09:42:16.76  
Circuit QNA-0, Load, Successful, Node = 2.42 (ARIEL)  
File = SYS$SYSDEVICE:<SYS11.>, Operating system  
Ethernet address = 08-00-2B-07-AC-13
```

Cluster Troubleshooting Information

C.1 Diagnosing Failures of Computers to Boot or to Join the Cluster

If the satellite cannot communicate with the boot server, no message for that satellite appears. There may be a problem with an Ethernet cable connection or adapter service.

If the satellite's data in the DECnet database is incorrectly specified (for example, if the hardware address is incorrect), a message like the following displays the correct address and indicates that a load was requested:

```
DECnet event 0.7, aborted service request
From node 2.4 (URANUS), 15-APR-1990 09:42:09.67
Circuit QNA-0, Line open error, Ethernet address = 08-00-2B-03-29-99
```

Note the absence of the node name, node address, and system root.

If a satellite fails to boot, perform the following checks:

- Verify that the boot device is available. This check is particularly important for local area and mixed-interconnect clusters in which satellites boot from multiple system disks.
- Verify that the satellite's SCSNODE and SCSSYSTEMID values and its DECnet node name and address are unique in the cluster.
- Verify that the DECnet-VAX network is up and running.
- Verify that circuit service is enabled for the boot server's Ethernet adapter device. Invoke the NCP Utility and enter an NCP command in the following format, where *circuit-id* is the name of the Ethernet adapter circuit that the boot server uses to service downline load requests from satellites:

```
NCP> SHOW CIRCUIT circuit-id
```

If service is not enabled, you can enter NCP commands like the following to enable it:

```
NCP> SET CIRCUIT circuit-id STATE OFF
NCP> DEFINE CIRCUIT circuit-id SERVICE ENABLED
NCP> SET CIRCUIT circuit-id SERVICE ENABLED STATE ON
```

The DEFINE command updates the permanent database and ensures that service is enabled the next time you start the network. Note that DECnet traffic will be interrupted while the circuit is off.

- Verify that you have specified the correct Ethernet hardware address for the satellite. Proceed as follows:
 - 1 Enter an NCP command in the following format on the boot server, specifying the satellite's node name:

```
NCP> SHOW NODE X... CHARACTERISTICS
```

The system displays data like the following:

```
Node Volatile Characteristics as of 15-APR-1990 13:15:28
Remote node =      2.41 (ARIEL)

Hardware address           = 08-00-2B-03-27-95
Tertiary loader            = SYS$SYSTEM:TERTIARY_VMB.EXE
Load Assist Agent         = SYS$SHARE:NISCS_LAA.EXE
Load Assist Parameter     = DISK$VAXVMSRL5:<SYS12.>
```

Cluster Troubleshooting Information

C.1 Diagnosing Failures of Computers to Boot or to Join the Cluster

- 2 At the satellite's console prompt (>>>), enter the commands shown in Table 5-2 to display the satellite's current Ethernet hardware address.
 - 3 Compare the hardware address values displayed by NCP and at the satellite's console. The values should be identical and should also match the value shown in the file `SYS$MANAGER:NETNODE_UPDATE.COM`. If the values do not match, you must make appropriate adjustments. For example, if you have recently replaced the satellite's Ethernet adapter device, you must execute `CLUSTER_CONFIG`'s `CHANGE` function to update the network database and `NETNODE_UPDATE.COM` on the appropriate boot server.
- Verify that the satellite's load assist parameter specifies the correct device and root directory name and that the satellite's root is unique in the cluster. If changes are needed, you can use `CLUSTER_CONFIG.COM` to remove the satellite and then add it again with correct values.

C.1.4 Computer Fails to Join the Cluster

If a computer boots but fails to join the cluster, proceed as follows:

- Verify that VAXcluster software has been loaded. Look for connection manager (`%CNXMAN`) messages like those shown in Section C.1.1. If no such messages are displayed, it is likely that VAXcluster software was not loaded at boot time. Reboot the computer in conversational mode. At the `SYSBOOT>` prompt, set the `VAXCLUSTER` parameter to 2. (In local area or mixed-interconnect clusters, you must also set `NISCS_LOAD_PEA0` to 1.) Note that these parameters should also be set in the computer's `MODPARAMS.DAT` file. For more information on booting a computer in conversational mode, consult your processor-specific installation and operations guide.

In local area and mixed-interconnect clusters, verify that the cluster security database file (`SYS$COMMON:CLUSTER_AUTHORIZE.DAT`) exists and that you have specified the correct group number for this cluster.

- Verify that the computer has booted from the correct disk and system root. If `%CNXMAN` messages are displayed, and if after the conversational reboot the computer still does not join the cluster, check the console output on all active computers and look for messages indicating that one or more computers found a remote computer that conflicted with a known or local computer. Such messages suggest that two computers have booted from the same system root.

Review the boot command files for all CI-connected computers and ensure that all are booting from the correct disks and from unique system roots. If you find it necessary to modify the computer's bootstrap command procedure (console media), you may be able to do so on another processor that is already running in the cluster. Replace the running processor's console media with the media to be modified, and use the Exchange Utility and a text editor to make

Cluster Troubleshooting Information

C.1 Diagnosing Failures of Computers to Boot or to Join the Cluster

the required changes. Consult the appropriate processor-specific installation and operations guide for information on examining and editing boot command files.

- Verify that the computer's SCSNODE and SCSSYSTEMID parameters are unique in the cluster. To be eligible to join a cluster, a computer must have unique SCSNODE and SYSSYSTEMID parameter values. Check that the current values do not duplicate any values set for existing VAXcluster computers. Note that if you discover that one or the other value is not unique, you must alter *both* values or reboot all other computers. To check or modify values, you can perform a conversational bootstrap operation. However, for reliable future bootstrap operations, you must specify appropriate values for these parameters in the computer's MODPARAMS.DAT file.

C.1.5 --- Startup Procedures Fail to Complete

If a computer boots and joins the cluster but appears to hang before startup procedures complete—that is, before you are able to log in to the system, be sure that you have allowed sufficient time for the startup procedures to execute.

If the startup procedures fail to complete after a period that is normal for your site, try to access the procedures from another VAXcluster computer and make appropriate adjustments. For example, verify that all required devices are configured and available.

One potential cause of such a failure is the lack of some system resource such as NPAGEDYN or page file space. If you suspect that the value for the NPAGEDYN parameter is set too low, you can perform a conversational bootstrap operation to increase it. Use SYSBOOT to check the current value, and then double the value. If this procedure is unsuccessful, double the value once more.

If you suspect a shortage of page file space, and if another VAXcluster computer is available, you can log in on that computer and use the System Generation Utility (SYSGEN) to provide adequate page file space for the problem computer. (Note that insufficient page file space on the booting computer may cause other computers to hang.) If the computer still cannot complete the startup procedures, contact your Customer Services representative.

C.2 --- Diagnosing Cluster Hangs

Conditions like the following can cause a VAXcluster computer to suspend process or system activity (that is, to hang):

- Cluster quorum is lost.
- A shared cluster resource is inaccessible.

Sections C.2.1 and C.2.2 discuss these conditions.

Cluster Troubleshooting Information

C.2 Diagnosing Cluster Hangs

C.2.1 Cluster Quorum Is Lost

The VAXcluster quorum scheme coordinates activity among VAXcluster computers and ensures the integrity of shared cluster resources. (The quorum scheme is described fully in Section 1.5.1.) Quorum is checked after any change to the cluster configuration—for example, when a voting computer leaves or joins the cluster. If quorum is lost, process creation and I/O activity on all computers in the cluster are blocked.

Information about the loss of quorum and clusterwide events that cause loss of quorum are sent to the OPCOM process, which broadcasts messages to designated operator terminals. The information is also broadcast to each computer's operator console (OPA0), unless broadcast activity is explicitly disabled on that terminal. Because, however, quorum may be lost before OPCOM has been able to inform the operator terminals, the messages sent to OPA0 are the most reliable source of information about events that may cause loss of quorum.

If quorum is lost, you can follow instructions in Section 5.7.4 to recover.

C.2.2 A Shared Cluster Resource Is Inaccessible

Access to shared cluster resources is coordinated by the distributed lock manager. If a particular process is granted a lock on a resource (for example, a shared data file), other processes in the cluster that request incompatible locks on that resource must wait until the original lock is released. If the original process retains its lock for an extended period, other processes waiting for the lock to be released may appear to hang.

Occasionally a system activity must acquire a restrictive lock on a resource for an extended period. For example, to perform a volume rebuild, system software takes out an exclusive lock on the volume being rebuilt. While this lock is held, no processes can allocate space on the disk volume. If they attempt to do so, they may appear to hang.

Access to files that contain data necessary for the operation of the system itself is coordinated by the distributed lock manager. For this reason, a process that acquires a lock on one of these resources and is then unable to proceed may cause the cluster to appear to hang.

For example, this condition may occur if a process locks a portion of the system authorization file (SYS\$SYSTEM:SYSUAF.DAT) for write access. Any activity that requires access to that portion of the file, such as logging into an account with the same or similar user name or sending mail to that username, will be blocked until the original lock is released. Normally this lock would be released quickly, and users would not notice the locking operation.

However, if the process holding the lock is unable to proceed, other processes could enter a wait state. Because the authorization file is used during login and for most process creation operations (for example, batch and network jobs) blocked processes could rapidly accumulate in the cluster. Because the distributed lock manager is functioning normally under these conditions, users are not notified by broadcast messages or other means that a problem has occurred.

C.3 Diagnosing CLUEXIT Bugchecks

The VMS operating system performs **bugcheck** operations only when it detects conditions that could compromise normal system activity or endanger data integrity. A **CLUEXIT bugcheck** is a type of bugcheck initiated by the connection manager, the VAXcluster software component that manages the interaction of cooperating VAXcluster computers. Most such bugchecks are triggered by conditions resulting from hardware failures (particularly failures in communications paths), configuration errors, or system management errors.

The conditions that most commonly result in CLUEXIT bugchecks are as follows:

- The cluster connection between two computers is broken for longer than `RECNXINTERVAL` seconds. Thereafter, the connection is declared irrevocably broken. If the connection is later reestablished, either or both of the computers shut down with a CLUEXIT bugcheck.

This condition can occur upon power failure recovery with battery backup, after the repair of an SCS communication link, or after the computer was halted for a period longer than `RECNXINTERVAL` seconds, and was restarted with a `CONTINUE` command entered at the operator console. You must determine the cause of the interrupted connection and correct the problem. For example, if recovery from a power failure is longer than `RECNXINTERVAL` seconds, you may want to increase the value of the `RECNXINTERVAL` parameter on all computers.

- Cluster partitioning occurs. A member of a cluster discovers or establishes connection to a member of another cluster, or a foreign cluster is detected in the quorum file. In this case, you must review the setting of `EXPECTED_VOTES` on all computers.
- The value specified for the `SYSGEN` parameter `SCSMAXMSG` on a computer is too small. Verify that the value of `SCSMAXMSG` on all VAXcluster computers is set to a value that is at the least the default value.

C.4 Diagnosing VAXport Device Problems

The following sections present information on the CI and Ethernet VAXport devices. Information is also provided on entries in the system error log and on corrective actions to take when errors occur. Topics include the following:

- VAXport communication mechanisms
- Port failures
- VAXcluster error log entries
- OPA0 error messages

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

C.4.1 **VAXport Communication Mechanisms**

This section describes CI and Ethernet port communication mechanisms and System Communications Services (SCS) connections.

Port Polling

Shortly after a CI-connected computer boots, the CI port driver (PADRIVER) begins configuration polling to discover other active ports on the CI. Normally the poller runs every 5 seconds (the default value of the SYSGEN parameter PAPOLLINTERVAL). In the first polling pass, all addresses are probed over cable path A; on the second pass all addresses are probed over path B; on the third pass path A is probed again, and so on.

The poller probes by sending request ID (REQID) packets to all possible port numbers, including itself. Active ports receiving the REQIDs return ID packets (IDREC) to the port issuing the REQID. A port may respond to a REQID even if the computer attached to the port is not running.

In any CI-only, local area, or mixed-interconnect cluster, the port drivers perform a start handshake when a pair of ports and port drivers has successfully exchanged ID packets. The port drivers exchange datagrams containing information about the computers, such as the type of computer and the operating system version. If this exchange is successful, each computer declares a virtual circuit open. An open virtual circuit is prerequisite to all other activity.

Ethernet Communications

In local area and mixed-interconnect clusters, a multicast scheme is used to locate computers on the Ethernet. Every 3 seconds the port emulator driver (PEDRIVER) sends HELLO messages to a cluster-specific multicast address that is derived from the cluster group number. The driver also enables the reception of these messages from other computers. When the driver receives a HELLO message from a computer with which it does not currently share an open virtual circuit, it attempts to create a circuit. HELLO messages received from a computer with a currently open virtual circuit indicate that the remote computer is operational.

A standard three-message exchange handshake is used to create a virtual circuit. The handshake messages contain information about the transmitting computer and its record of the cluster password. These parameters are verified at the receiving computer, which continues the handshake only if its verification is successful. Thus, each computer authenticates the other. After the final message, the virtual circuit is opened for use by both computers.

System Communications Services (SCS) Connections

System services such as the disk class driver, connection manager, and the MSCP server communicate between computers with a protocol called System Communications Services (SCS). Primarily, SCS is responsible for the formation and breaking of intersystem process connections and for flow control of message traffic over those connections. SCS is implemented in the VAXport driver (for example, PADRIVER, PBDRIVER, PEDRIVER),

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

and in a loadable piece of the VMS operating system called SCSLOA.EXE (loaded automatically during system initialization).

When a virtual circuit has been opened, a computer periodically probes a remote computer for system services that the remote computer may be offering. The SCS directory service, which makes known services that a computer is offering, is always present both on computers and HSC subsystems. As system services discover their counterparts on other computers and HSC subsystems, they establish SCS connections to each other. These connections are full duplex and are associated with a particular virtual circuit. Multiple connections are typically associated with a virtual circuit.

C.4.2 Port Failures

Taken together, SCS, the VAXport drivers, and the port itself support a hierarchy of communications paths. Working up from the most fundamental level, these are as follows:

- The physical wires. The Ethernet is a single coaxial cable. The CI has two pairs of transmit and receive cables (path A transmit and receive and path B transmit and receive). For the CI, VMS software normally sends traffic in automatic path select mode. The port chooses the free path or, if both are free, an arbitrary path (implemented in the cables and star coupler, and managed by the port).
- The virtual circuit (implemented partly in the CI port or Ethernet port emulator driver (PEDRIVER) and partly in SCS software).
- The SCS connections (implemented in system software).

Failures can occur at each communications level and in each component. Failures at one level translate into failures at other levels as follows:

- **Wires.** If the Ethernet fails or is disconnected, Ethernet traffic stops or is interrupted, depending on the nature of the failure. For the CI, either path A or B can fail while the virtual circuit remains intact. All traffic is directed over the remaining good path. When the wire is repaired, the repair is detected automatically by port polling, and normal operations resume on all ports.
- **Virtual circuit.** If no path works between a pair of ports, the virtual circuit fails and is closed. A path failure is discovered as follows:
 - For the CI, when polling fails, or when attempts are made to send normal traffic, and the port reports that neither path yielded transmit success.
 - For the Ethernet, when no multicast HELLO message or incoming traffic is received from another computer.

When a virtual circuit fails, every SCS connection on it fails. The software automatically reestablishes connections when the virtual circuit is reestablished. Normally, reestablishing a virtual circuit takes several seconds after the problem is corrected.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

- **CI port.** If a port fails, all virtual circuits to that port fail, and all SCS connections on those virtual circuits fail. If the port is successfully reinitialized, virtual circuits and connections are reestablished automatically. Normally, port reinitialization and reestablishment of connections take several seconds.
- **Ethernet adapter.** If an Ethernet adapter device fails, attempts are made to restart it. If repeated attempts fail, all virtual circuits time out, and their connections are broken.
- **SCS connection.** When the software protocols fail or, in some instances, when the software detects a hardware malfunction, a connection is terminated. Other connections are normally unaffected, as is the virtual circuit. Breaking of connections is also used under certain conditions as an error recovery mechanism—most commonly when there is insufficient nonpaged pool available on the computer.
- **Computer.** If a computer fails because of operator shutdown, bugcheck, or halt and reboot, all other computers in the cluster record the failure as failures of their virtual circuits to the port on the failed computer.

C.4.2.1 Verifying CI Port Functions

Before you boot in a cluster a CI-connected computer that is new, just repaired, or suspected of having a problem, you should have Customer Services verify that the computer runs correctly on its own.

To diagnose communication problems, you can invoke the Show Cluster Utility and tailor the SHOW CLUSTER report by entering the SHOW CLUSTER command ADD CIRCUIT CABLE_ST. This command adds a class of information about all the virtual circuits as seen from the computer on which you are running SHOW CLUSTER. Primarily, you are checking whether there is a virtual circuit in the OPEN state to the failing computer. Common causes of failure to open a virtual circuit and keep it open are the following:

- Port errors on one side or the other
- Cabling errors
- A port set off line because of software problems
- Insufficient nonpaged pool available on both sides
- Failure to set correct values for the SYSGEN parameters SCSNODE, SCSSYSTEMID, PAMAXPORT, PANOPOLL, PASTIMOUT, and PAPOLLINTERVAL.

Run SHOW CLUSTER from each active computer in the cluster to verify whether each computer's view of the failing computer is consistent with every other computer's view. If all the active computers have a consistent view of the failing computer, the problem may be in the failing computer. If, on the other hand, only one of several active computers detects that the newcomer is failing, that particular computer may be experiencing a problem.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

If no virtual circuit is open to the failing computer, check the bottom of the SHOW CLUSTER display for information on circuits to the port of the failing computer. Virtual circuits in partially open states are shown at the bottom of the display. If the circuit is shown in a state other than OPEN, communications between the local and remote ports are taking place, and the failure is probably at a higher level than in port or cable hardware. Next, check that both paths A and B are good to the failing port. The loss of one path should not prevent a computer from participating in a cluster.

C.4.2.2 Verifying CI Cable Connections

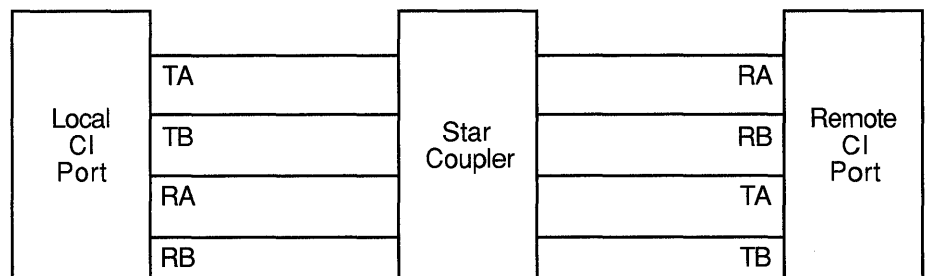
Whenever the configuration poller finds that no virtual circuits are open and that no handshake procedures are currently opening virtual circuits, the poller analyzes its environment. It does so by using the send-loopback-datagram facility of the CI port.

The send-loopback-datagram facility tests the connections between the CI port and the star coupler by routing messages across them. The messages are called loopback datagrams. (The port processes other self-directed messages without using the star coupler or external cables.)

The configuration poller makes entries in the error log whenever it detects a change in the state of a circuit. Note, however, that it is possible for two changed-to-failed-state messages to be entered in the log without an intervening changed-to-succeeded-state message. Such a series of entries means that the circuit state continues to be faulty.

The following paragraphs discuss various incorrect CI cabling configurations and the entries made in the error log when these configurations exist. Figure C-1 shows a two-computer configuration with all cables correctly connected. Figure C-2 shows a CI cluster with a pair of crossed cables.

Figure C-1 A Correctly Connected Two-Computer CI Cluster

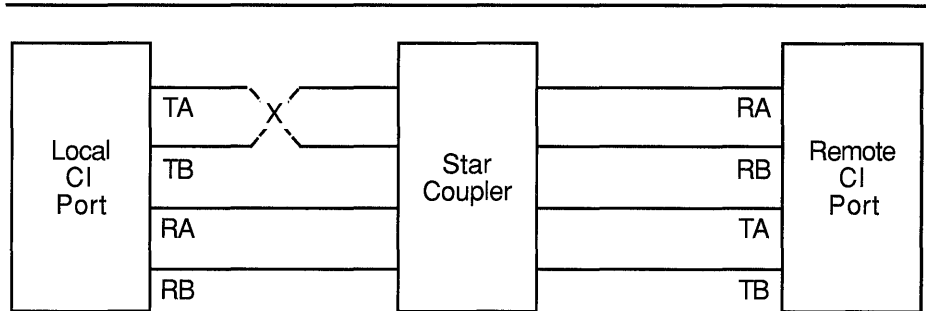


ZK-1924-GE

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

Figure C-2 Crossed CI Cable Pair



ZK-1925-GE

If a pair of transmitting cables or a pair of receiving cables is crossed, a message sent on TA is received on RB, and a message sent on TB is received on RA. This is a hardware error condition from which the port cannot recover. An entry is made in the error log to say that a single pair of crossed cables exists. The entry contains the following lines:

```
DATA CABLE(S) CHANGE OF STATE
PATH 1. LOOPBACK HAS GONE FROM GOOD TO BAD
```

If this situation exists, you can correct it by reconnecting the cables properly. The cables could be misconnected in several places. The coaxial cables that connect the port boards to the bulkhead cable connectors can be crossed, or the cables can be misconnected to the bulkhead or the star coupler.

The information in Figure C-2 can be represented more simply. Configuration 1 shows the cables positioned as in Figure C-2, but it does not show the star coupler or the computers. The letters LOC and REM indicate the pairs of transmitting (T) and receiving (R) cables on the local and remote computers, respectively.

Configuration 1

```
T x   = R
R =   = T
LOC   REM
```

The pair of crossed cables causes loopback datagrams to fail on the local computer, but succeed on the remote computer. Crossed pairs of transmitting cables and crossed pairs of receiving cables cause the same behavior.

Note that only an odd number of crossed-cable pairs causes these problems. If an even number of cable pairs is crossed, communications succeed. An error log entry is made in some cases, however, and the contents of the entry depends on which pairs of cables are crossed.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

Configuration 2 shows two-computer clusters with the combinations of two pairs of crossed-cable pairs. These crossed pairs cause the following entry to be made in the error log of the computer that has the cables crossed:

```
DATA CABLE(S) CHANGE OF STATE
CABLES HAVE GONE FROM UNCROSSED TO CROSSED
```

Loopback datagrams succeed on both computers, and communications are possible.

Configuration 2

T x	= R	T =	x R
R x	= T	R =	x T
LOC	REM	LOC	REM

Configuration 3 shows the possible combinations of two pairs of crossed cables that cause loopback datagrams to fail on both computers in the cluster. Communications can still take place between the computers. An entry stating that cables are crossed is made in the error log of each computer.

Configuration 3

T x	= R	T =	x R
R =	x T	R x	= T
LOC	REM	LOC	REM

Configuration 4 shows the possible combinations of two pairs of crossed cables that cause loopback datagrams to fail on both computers in the cluster, but allow communications. No entry stating that cables are crossed is made in the error log of either computer.

Configuration 4

T x	x R	T =	= R
R =	= T	R x	x T
LOC	REM	LOC	REM

Configuration 5 shows the possible combinations of three pairs of crossed cables. In each case, loopback datagrams fail on the computer that has only one crossed pair of cables. Loopback datagrams succeed on the computer with both pairs crossed. No communications are possible.

Configuration 5

T x	x R	T x	= R	T =	x R	T x	x R
R x	= T	R x	x T	R x	x T	R =	x T
LOC	REM	LOC	REM	LOC	REM	LOC	REM

If all four cable pairs between two computers are crossed, communications succeed, loopback datagrams succeed, and no crossed-cable message entries are made in the error log. Such a condition might be detected by noting error log entries made by a third computer in the cluster, but only if the third computer has one of the crossed-cable cases described.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

C.4.2.3 Repairing CI Cables

This section describes some ways in which Customer Services can make repairs on a running computer. This information is provided to aid system managers in scheduling repairs.

For cluster software to survive cable-checking activities or cable-replacement activities, you must be sure that either path A or path B is intact at all times between each port and between every other port in the cluster.

You can, for example, remove path A and path B in turn from a particular port to the star coupler. To make sure that the configuration poller finds a path that was previously faulty but is now operational, follow these steps:

- 1 Remove path B.
- 2 After the poller has discovered that path B is faulty, reconnect path B.
- 3 Wait two poller intervals, and then enter the DCL command `SHOW CLUSTER` to make sure that the poller has reestablished path B. Or, enter the DCL command `SHOW CLUSTER/CONTINUOUS` followed by the `SHOW CLUSTER` command `ADD CIRCUITS, CABLE_ST`. Wait until `SHOW CLUSTER` tells you that path B has been reestablished.
- 4 Remove path A.
- 5 After the poller has discovered that path A is faulty, reconnect path A.
- 6 Wait two poller intervals to make sure that the poller has reestablished path A.

If both paths are lost at the same time, the virtual circuits are lost between the port with the broken cables and all other ports in the cluster. This condition will in turn result in loss of SCS connections over the broken virtual circuits. However, recovery from this situation is automatic after an interruption in service on the affected computer. The length of the interruption varies, but it is usually approximately two poller intervals (or 10 seconds) at the default `SYSGEN` parameter settings.

C.4.3 Analyzing Error Log Entries for VAXport Devices

To anticipate and avoid potential problems, you must monitor events recorded in the error log. From the total error count, displayed by a DCL command in the format `SHOW DEVICE device-name`, you can determine whether errors are increasing. If so, you should examine the error log.

The DCL command `ANALYZE/ERROR_LOG` invokes the Error Log Utility to report the contents of an error log file. (For more information on the Error Log Utility, see the *VMS Error Log Utility Manual*.)

Note that some error log entries are informational only, and require no action. For example, If you shut down a computer in the cluster, all other active computers that have open virtual circuits between themselves and the computer that has been shut down make entries in their error logs.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

Such computers record up to three errors for the event:

- 1 Path A received no response.
- 2 Path B received no response.
- 3 The virtual circuit is being closed.

These messages are normal and reflect the change of state in the circuits to the computer that has been shut down.

On the other hand, some error log entries are made for problems that degrade operation, or for nonfatal hardware problems. The VMS operating system might continue to run satisfactorily under these conditions. The purpose of detecting these problems early is to prevent nonfatal problems (such as loss of a single CI path) from becoming serious problems (such as loss of both paths).

C.4.3.1 Error Log Entry Formats

Errors and other events on the CI or Ethernet cause VAXport drivers to enter information in the system error log. The two formats used for error log entries are the **device-attention** format and the **logged-message** format. Sections C.4.3.2 and C.4.3.3 describe those formats.

Device-attention entries for the CI record events that, in general, are indicated by the setting of a bit in a hardware register. For the Ethernet, device-attention entries typically record errors on an Ethernet adapter device. Logged-message entries record the receipt of a message packet that contains erroneous data or that signals an error condition.

C.4.3.2 Device-Attention Entries

Example C-1 shows device-attention entries for the CI. The left column gives the name of a device register or a memory location. The center column gives the value contained in that register or location, and the right column gives an interpretation of that value.

Example C-1 CI Device-Attention Entry

```
***** ENTRY      83. ***** ①
ERROR SEQUENCE 10.          LOGGED ON:      SID 0150400A
DATE/TIME 15-APR-1990 11:45:27.61          SYS_TYPE 01010000 ②
DEVICE ATTENTION   KA780                    ③
                   SCS NODE: MARS
CI SUB-SYSTEM, MARS$PAA0: - PORT POWER DOWN ④
```

(continued on next page)

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

Example C-1 (Cont.) CI Device-Attention Entry

CNFGR	00800038	ADAPTER IS CI ADAPTER POWER-DOWN	
PMCSR	000000CE	MAINTENANCE TIMER DISABLE MAINTENANCE INTERRUPT ENABLE MAINTENANCE INTERRUPT FLAG PROGRAMMABLE STARTING ADDRESS UNINITIALIZED STATE	
PSR	80000001	RESPONSE QUEUE AVAILABLE MAINTENANCE ERROR	
PFAR	00000000		
PESR	00000000		
PPR	03F80001		
UCB\$B_ERTCNT	32	50. RETRIES REMAINING	5
UCB\$B_ERTMAX	32	50. RETRIES ALLOWABLE	6
UCB\$L_CHAR	0C450000	SHAREABLE AVAILABLE ERROR LOGGING CAPABLE OF INPUT CAPABLE OF OUTPUT	
UCB\$W_STS	0010	ONLINE	
UCB\$W_ERRCNT	000B	11. ERRORS THIS UNIT	7

- 1 The first two lines are the entry heading. These lines contain the number of the entry in this error log file, the sequence number of this error, and the identification number (SID) of this computer. Each entry in the log file contains such a heading.
- 2 The next line contains the date and time, and the computer type.
- 3 The next two lines contain the entry type, the processor type (KA780), and the computer's SCS node name.
- 4 The line CI SUB-SYSTEM, MARS\$PAA0: - PORT POWER DOWN contains the name of the subsystem and the device that caused the entry, and the reason for the entry. The CI subsystem's device PAA0 on MARS was powered down.

The next 15 lines contain the names of hardware registers in the port, their contents, and interpretations of those contents. See the appropriate CI hardware manual for a description of all the CI port registers.

The CI port can recover from many errors, but not all. When an error occurs from which the CI cannot recover, the port notifies the port driver. The port driver logs the error and attempts to reinitialize the port. If the port fails after 50 such initialization attempts, the driver takes it off line, unless the system disk is connected to the failing port

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

or this computer is supposed to be a cluster member. If the CI port is required for system disk access or cluster participation and all 50 reinitialization attempts have been used, then the computer bugchecks with a CIPORT-type bugcheck. Once a CI port is off line, you can put the port back on line only by rebooting the computer.

- ⑤ The UCB\$B_ERTCNT field contains the number of reinitializations that the port driver can still attempt. The difference between this value and UCB\$B_ERTMAX is the number of reinitializations already attempted.
- ⑥ The UCB\$B_ERTMAX field contains the maximum number of times the port can be reinitialized by the port driver.
- ⑦ The UCB\$W_ERRCNT field contains the total number of errors that have occurred on this port since it was booted. This total includes both errors that caused reinitialization of the port and errors that did not.

Example C-2 shows device-attention entries for the Ethernet. The left column gives the name of a device register or a memory location. The center column gives the value contained in that register or location, and the right column gives an interpretation of that value.

Example C-2 Ethernet Device-Attention Entry

```

***** ENTRY      80. ***** ①
ERROR SEQUENCE 26.          LOGGED ON:      SID 08000000
DATE/TIME 15-APR-1990 11:30:53.07          SYS_TYPE 01010000 ②
DEVICE ATTENTION KA630 ③
                        SCS NODE: PHOBOS
NI-SCS SUB-SYSTEM, PHOBOS$PEAO: ④
  FATAL ERROR DETECTED BY DATALINK ⑤
  STATUS1          0000002C ⑥
  STATUS2          00000000
  DATALINK UNIT    0001 ⑦
  DATALINK NAME    41515803 ⑧
                        00000000
                        00000000
                        00000000
                        DATALINK NAME = XQA1:
  REMOTE NODE      00000000 ⑨
                        00000000
                        00000000
                        00000000
  REMOTE ADDR      00000000 ⑩
                        0000
  LOCAL ADDR       000400AA ⑪
                        4C07
                        ETHERNET ADDR = AA-00-04-00-07-4C ⑫
  ERROR CNT        0001
  UCB$W_ERRCNT     0007
                        1. ERROR OCCURRENCES THIS ENTRY
                        7. ERRORS THIS UNIT

```

- ① The first two lines are the entry heading. These lines contain the number of the entry in this error log file, the sequence number of this error, and the identification number (SID) of this computer. Each entry in the log file contains such a heading.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

- ② The next line contains the date and time, and the computer type.
- ③ The next two lines contain the entry type, the processor type (KA630), and the computer's SCS node name.
- ④ This line shows the name of the subsystem and component that caused the entry.
- ⑤ This line shows the reason for the entry. The Ethernet driver has shut down the data link because of a fatal error. The data link will be restarted automatically, if possible.
- ⑥ STATUS1 and STATUS 2 show the I/O completion status returned by the Ethernet driver. If a message transmit was involved, the status applies to that transmit.
- ⑦ DATALINK UNIT shows the unit number of the Ethernet device on which the error occurred.
- ⑧ DATALINK NAME is the name of the Ethernet device on which the error occurred.
- ⑨ REMOTE NODE is the name of the remote node to which the packet was being sent. If zero, no remote node was available or no packet was associated with the error.
- ⑩ REMOTE ADDR is the Ethernet address of the remote node to which the packet was being sent. If zero, no packet was associated with the error.
- ⑪ LOCAL ADDR is the Ethernet address of the local node.
- ⑫ ERROR CNT. Because some errors can occur at extremely high rates, some error log entries represent more than one occurrence of an error. This field indicates how many. The errors counted occurred in the 3 seconds preceding the time stamp on the entry.

C.4.3.3 Logged-Message Entries

Logged-message entries are made when the CI or Ethernet port receives a response that contains either data that the port driver cannot interpret or an error code in the status field of the response.

Example C-3 shows a CI logged-message entry with an error code in the status field PPD\$B_STATUS.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

Example C-3 CI Logged-Message Entry

```
***** ENTRY          3. ***** ①
ERROR SEQUENCE 3.          LOGGED ON SID 01188542
ERL$LOGMESSAGE, 15-APR-1990 13:40:25.13 ②
          KA780 REV #3. SERIAL #1346.    MFG PLANT 15. ③
CI SUB-SYSTEM, MARS$PAA0: ④
DATA CABLE(S) STATE CHANGE - PATH #0. WENT FROM GOOD TO BAD ⑤
    LOCAL STATION ADDRESS, 000000000002 (HEX) ⑥
    LOCAL SYSTEM ID, 000000000001 (HEX) ⑦
    REMOTE STATION ADDRESS, 000000000004 (HEX) ⑧
    REMOTE SYSTEM ID, 00000000000A9 (HEX) ⑨
UCB$B_ERTCNT          32 ⑩
                                50. RETRIES REMAINING
UCB$B_ERTMAX          32
                                50. RETRIES ALLOWABLE
UCB$W_ERRCNT          0001
PPD$B_PORT            04 ⑪
                                1. ERRORS THIS UNIT
PPD$B_STATUS          A5 ⑫
                                REMOTE NODE #4.
                                FAIL
                                PATH #0., NO RESPONSE
                                PATH #1., "ACK" OR NOT USED
                                NO PATH
PPD$B_OPC             05 ⑬
                                IDREQ
PPD$B_FLAGS           03 ⑭
                                RESPONSE QUEUE BIT
                                SELECT PATH #0.
"CI" MESSAGE ⑮
00000000
00000000
80000004
0000FE15
4F503000
00000507
00000000
00000000
00000000
00000000
00000000
00000000
00000000
00000000
00000000
00000000
00000000
00000000
00000000
00000000
```

- ① The first two lines are the entry heading. These lines contain the number of the entry in this error log file, the sequence number of the error, and the identification number (SID) of the computer. Each entry in the log file contains a heading.
- ② The next line contains the entry type, the date, and time.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

- ③ The next line contains the processor type (KA780), the hardware revision number of the computer (REV #3), the serial number of the computer (SERIAL #1346), and the plant number (15).
- ④ The line CI SUB-SYSTEM, MARS\$PAA0: contains the name of the subsystem and the device that caused the entry.
- ⑤ The next line gives the reason for the entry (one or more data cables have changed state), and a more detailed reason for the entry. Path 0, which the port used successfully before, cannot be used now.

Note: ANALYZE/ERROR_LOG uses the notation path 0 and path 1; cable labels use the notation path A (=0) and path B (=1).

- ⑥ The local (⑥) and remote (⑧) station addresses are the port numbers (range 0–15) of the local and remote ports. The port numbers are set in hardware switches by Customer Services. The local (⑦) and remote (⑨) system IDs are the SCS system IDs set by the SYSGEN parameter SCSSYSTEMID for the local and remote systems. For HSC subsystems, the system ID is set with the HSC console.
- ⑦ The rest of the entry, which consists of the entry fields that begin with UCB\$, gives information on the contents of the unit control block (UCB) for this CI device.

The following fields , which begin with PPD\$, are fields in the message packet that the local port has received.

- ⑧ PPD\$B_PORT contains the station address of the remote port. In a loopback datagram, however, this field contains the local station address.
- ⑨ The PPD\$B_STATUS field contains information on the nature of the failure that occurred during the current operation. When the operation completes without error, ERF prints the word NORMAL beside this field; otherwise, ERF decodes the error information contained in PPD\$B_STATUS. Here a NO PATH error occurred because of a lack of response on path 0, the selected path.
- ⑩ The PPD\$B_OPC field contains the code for the operation that the port was attempting when the error occurred. The port was trying to send a request-for-ID message.
- ⑪ The PPD\$B_FLAGS field contains bits that indicate, among other things, the path that was selected for the operation.
- ⑫ The “CI” MESSAGE is a hexadecimal listing of bytes 16 through 83 (decimal) of the response (message or datagram). Since responses are of variable length depending upon the port opcode, bytes 16 through 83 may contain either more or fewer bytes than actually belong to the message. Here the request-for-id contains no information in bytes 16 through 83.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

C.4.3.4 Error Log Entry Descriptions

This section describes error log entries for the CI and Ethernet ports. Each entry shown is followed by a brief description of what the associated VAXport driver (PADRIVER, PBDRIVER, PEDRIVER) does, and the suggested action a system manager should take. In cases where Software Performance Reports (SPRs) with crash dumps are requested, it is important to capture the crash dumps as soon as possible after the error. For CI entries, note that path A and path 0 are the same path, and that path B and path 1 are the same path.

BIIC FAILURE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services.

CI PORT TIMEOUT

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device offline.

User Action: First, increase the SYSGEN parameter PAPOLLINTERVAL. If the problem disappears and you are not running privileged user-written software, submit an SPR. Otherwise, call Customer Services.

11/750 CPU MICROCODE NOT ADEQUATE FOR PORT

Explanation: The VAXport driver sets the port off line with no retries attempted. In addition, if this port is needed because the computer is booted from an HSC subsystem or is participating in a cluster, the computer bugchecks with a UCODEREV code bugcheck.

User Action: Read the appropriate section in the current VAXcluster SPD for information on required computer microcode revisions. Call Customer Services if necessary.

PORT MICROCODE REV NOT CURRENT, BUT SUPPORTED

Explanation: The VAXport driver detected that the microcode is not at the current level, but will continue normally. This error is logged as a warning only.

User Action: Contact Customer Services when it is convenient to have the microcode updated.

PORT MICROCODE REV NOT SUPPORTED

Explanation: The VAXport driver sets the port off line without attempting any retries.

User Action: Read the VAXcluster SPD for information on the required CI port microcode revisions. Contact Customer Services if necessary.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

DATA CABLE(S) STATE CHANGE
CABLES HAVE GONE FROM CROSSED TO UNCROSSED

Explanation: The VAXport driver logs this event.

User Action: No action needed.

DATA CABLE(S) STATE CHANGE
CABLES HAVE GONE FROM UNCROSSED TO CROSSED

Explanation: The VAXport driver logs this event.

User Action: Check for crossed-cable pairs. See Section C.4.2.2.

DATA CABLE(S) STATE CHANGE
PATH 0. WENT FROM BAD TO GOOD

Explanation: The VAXport driver logs this event.

User Action: No action needed.

DATA CABLE(S) STATE CHANGE
PATH 0. WENT FROM GOOD TO BAD

Explanation: The VAXport driver logs this event.

User Action: Check path A cables to see that they are not broken or improperly connected.

DATA CABLE(S) STATE CHANGE
PATH 0. LOOPBACK IS NOW GOOD, UNCROSSED

Explanation: The VAXport driver logs this event.

User Action: No action needed.

DATA CABLE(S) STATE CHANGE
PATH 0. LOOPBACK WENT FROM GOOD TO BAD

Explanation: The VAXport driver logs this event.

User Action: Check for crossed-cable pairs or faulty CI hardware. See Sections C.4.2.1 and C.4.2.2.

DATA CABLE(S) STATE CHANGE
PATH 1. WENT FROM BAD TO GOOD

Explanation: The VAXport driver logs this event.

User Action: No action needed.

DATA CABLE(S) STATE CHANGE
PATH 1. WENT FROM GOOD TO BAD

Explanation: The VAXport driver logs this event.

User Action: Check path B cables to see that they are not broken or improperly connected.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

DATA CABLE(S) STATE CHANGE
PATH 1. LOOPBACK IS NOW GOOD, UNCROSSED

Explanation: The VAXport driver logs this event.

User Action: No action needed.

DATA CABLE(S) STATE CHANGE
PATH 1. LOOPBACK WENT FROM GOOD TO BAD

Explanation: The VAXport driver logs this event.

User Action: Check for crossed-cable pairs or faulty CI hardware. See Sections C.4.2.1 and C.4.2.2.

DATAGRAM FREE QUEUE INSERT FAILURE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.

DATAGRAM FREE QUEUE REMOVE FAILURE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures, or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.

FAILED TO LOCATE PORT MICRO-CODE IMAGE

Explanation: The VAXport driver marks device off line and makes no retries.

User Action: Make sure console volume contains the microcode file CI780.BIN (for the CI780, CI750, or CIBCI) or the microcode file CIBCA.BIN for the CIBCA-AA, and then reboot the computer.

HIGH PRIORITY COMMAND QUEUE INSERT FAILURE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, 8800) contention.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

MSCP ERROR LOGGING DATAGRAM RECEIVED

Explanation: On receipt of an error message from the HSC subsystem, the VAXport driver logs the error and takes no other action. It is recommended that you disable the sending of HSC informational error log datagrams with the appropriate HSC console command. Informational error log datagrams take considerable space in the error log data file.

User Action: They are useful to read only if they are not captured on the HSC console for some reason (for example, the HSC console ran out of paper.) This logged information is a duplicate of the messages logged on the HSC console.

INAPPROPRIATE SCA CONTROL MESSAGE

Explanation: The VAXport driver closes the port-to-port virtual circuit to the remote port.

User Action: Submit a Software Performance Report to Digital including the error logs and the crash dumps from the local and remote computers.

INSUFFICIENT NON-PAGED POOL FOR INITIALIZATION

Explanation: The VAXport driver marks device off line and makes no retries.

User Action: Reboot the computer with a larger value for NPAGEDYN or NPAGEVIR.

LOW PRIORITY CMD QUEUE INSERT FAILURE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.

MESSAGE FREE QUEUE INSERT FAILURE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.

MESSAGE FREE QUEUE REMOVE FAILURE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

MICRO-CODE VERIFICATION ERROR

Explanation: The VAXport driver detected an error while reading the microcode that it just loaded into the port. The driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services.

NO PATH-BLOCK DURING VIRTUAL CIRCUIT CLOSE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Submit a Software Performance Report to Digital including the error log and a crash dump from the local computer.

NO TRANSITION FROM UNINITIALIZED TO DISABLED

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services.

PORT ERROR BIT(S) SET

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: For CI microcode version 7 or later, a **maintenance timer expiration** bit may mean that the PASTIMOUT SYSGEN parameter is set too low, especially if the local computer is running privileged user-written software. For all other bits, call Customer Services.

PORT HAS CLOSED VIRTUAL CIRCUIT

Explanation: The VAXport driver closes the virtual circuit that the local CI port opened to the remote port.

User Action: Check the PPD\$B_STATUS field of the error log entry for the reason the virtual circuit was closed. This error is normal if the remote computer crashed or was shut down.

PORT POWER DOWN

Explanation: The VAXport driver halts port operations, and then waits for power to return to the port hardware.

User Action: Restore power to the port hardware.

PORT POWER UP

Explanation: The VAXport driver reinitializes the port and restarts port operations.

User Action: No action needed.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

RECEIVED CONNECT WITHOUT PATH-BLOCK

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Submit a Software Performance Report to Digital including the error log and a crash dump from the local computer.

REMOTE SYSTEM CONFLICTS WITH KNOWN SYSTEM

Explanation: The configuration poller discovered a remote computer with SCSSYSTEMID and/or SCSNODE equal to that of another computer to which a virtual circuit is already open.

User Action: Shut the new computer down as soon as possible. Reboot it with a unique SCSSYSTEMID and SCSNODE. Do not leave the new computer up any longer than necessary. If you are running a cluster and two computers with conflicting identity are polling when any other virtual circuit failure takes place in the cluster, then computers in the cluster may crash with a CLUEXIT bugcheck.

RESPONSE QUEUE REMOVE FAILURE

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.

SCSSYSTEMID MUST BE SET TO NON-ZERO VALUE

Explanation: The VAXport driver sets the port off line without attempting any retries.

User Action: Reboot the computer with a conversational boot and set the SCSSYSTEMID to the correct value. At the same time, check that SCSNODE has been set to the correct nonblank value.

SOFTWARE IS CLOSING VIRTUAL CIRCUIT

Explanation: The VAXport driver closes the virtual circuit to the remote port.

User Action: Check error log entries for the cause of the virtual circuit closure. Faulty transmission or reception on both paths, for example, causes this error and may be detected from the one or two previous error log entries noting bad paths to this remote computer.

SOFTWARE SHUTTING DOWN PORT

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Check other error log entries for the possible cause of the port reinitialization failure.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

UNEXPECTED INTERRUPT

Explanation: The VAXport driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.

User Action: Call Customer Services.

UNRECOGNIZED SCA PACKET

Explanation: The VAXport driver closes the virtual circuit to the remote port. If the virtual circuit is already closed, the port driver inhibits datagram reception from the remote port.

User Action: Submit a Software Performance Report to Digital, including the error log file that contains this entry and the crash dumps from both the local and remote computers.

VIRTUAL CIRCUIT TIMEOUT

Explanation: The VAXport driver closes the virtual circuit that the local CI port opened to the remote port. This closure occurs if the remote computer is running CI microcode version 7 or later, and if the remote computer has failed to respond to any messages sent by the local computer.

User Action: This error is normal if the remote computer has halted, crashed, or was shut down. This error may mean that the local computer's PASTIMOUT SYSGEN parameter is set too low, especially if the remote computer is running privileged user-written software.

INSUFFICIENT NON-PAGED POOL FOR VIRTUAL CIRCUITS

Explanation: The VAXport driver closes virtual circuits because of insufficient pool.

User Action: Enter the DCL command SHOW MEMORY to determine pool requirements, and then adjust the appropriate SYSGEN requirements.

Note: The following descriptions apply only for Ethernet devices.

FATAL ERROR DETECTED BY DATALINK

Completion status: SS\$_ABORT (0000002C)

Explanation: The Ethernet driver has shut down the device because of a fatal error and is returning all outstanding transmits with SS\$_OPINCOMPL. The Ethernet device is automatically restarted, and all the aborted transmits are logged in the error log.

User Action: Infrequent occurrences of this error are probably not a problem. If the error occurs frequently or is accompanied by loss or reestablishment of connections to remote computers, there is probably a hardware problem. Check for the proper Ethernet adapter revision level or call Customer Services.

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

TRANSMIT ERROR FROM DATALINK

Completion status: SS\$_OPINCOMPL (000002D4)

Explanation: The Ethernet driver is in the process of restarting the data link because an error forced the driver to shut down the controller and all users (see FATAL ERROR DETECTED BY DATALINK).

Completion status: SS\$_DEVREQERR (00000334)

Explanation: The Ethernet controller tried to transmit the packet 16 times and failed because of defers and/or collisions. This condition indicates that Ethernet traffic is heavy.

Completion status: SS\$_DISCONNECT (0000204C)

Explanation: There was a loss of carrier during or after the transmit.

User Action: The port emulator automatically recovers from any of these errors, but many such errors indicate either that the Ethernet controller is faulty or that the Ethernet is overloaded. If you suspect either of these conditions, contact Customer Services.

INVALID CLUSTER PASSWORD RECEIVED

Explanation: A computer is trying to join the cluster using the correct cluster group number for this cluster but an invalid password. The Port Emulator discards the message. The probable cause is that another cluster on the Ethernet LAN is using the same cluster group number.

User Action: Provide all clusters on the same Ethernet LAN with unique cluster group numbers.

NISCS PROTOCOL VERSION MISMATCH RECEIVED

Explanation: A computer is trying to join the cluster using a version of the cluster Ethernet protocol that is incompatible with the one in use on this cluster.

User Action: Install a version of the VMS operating system that uses a compatible protocol, or change the cluster group number so that the computer joins a different cluster.

C.4.4 OPA0 Error Messages

VAXport drivers detect certain error conditions and attempt to log them. Under some circumstances, attempts to log errors to the error logging device can fail. Such failures can occur because the error logging device is not accessible when attempts are made to log the error condition. Because of the central role that the VAXport device plays in clusters, the loss of error-logged information in such cases makes it difficult to diagnose and fix problems.

A second, redundant method of error logging captures at least some of the information about VAXport device error conditions that would otherwise be lost. This second method consists of broadcasting selected information about the error condition to OPA0, in addition to the port driver's attempt to log the error condition to the error logging device. The VAXport driver

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

attempts both OPA0 error broadcasting and standard error logging under any of the following circumstances:

- The system disk has not yet been mounted.
- The system disk is undergoing mount verification.
- During mount verification, the system disk drive contains the wrong volume.
- Mount verification for the system disk has timed out.
- The local computer is participating in a cluster, and quorum has been lost.

Note the implicit assumption that the system and error logging devices are one and the same.

This second method of reporting errors is also not entirely reliable. Because of the way OPA0 error broadcasting is performed, some error conditions may not be reported. This situation occurs whenever a second error condition is detected before the VAXport driver has been able to broadcast the first error condition to OPA0. In such a case, only the first error condition is reported to OPA0, because that condition is deemed to be the more important one.

Certain error conditions are always broadcast to OPA0, regardless of whether the error logging device is accessible. In general, these are errors that cause the port to shut down either permanently or temporarily.

One OPA0 error message for each error condition is always logged. The text of each error message is similar to the text in the summary displayed by formatting the corresponding standard error log entry using the Error Log Utility. (See Section C.4.3.4 for a list of Error Log Utility summary messages and their explanations.)

Many of the OPA0 error messages contain some optional information such as the remote port number, CI packet information (flags, port operation code, response status, and port number fields), or specific CI port registers.

Following is a list of OPA0 error messages, subdivided by error type. See the CI hardware documentation for a detailed description of the CI port registers (CNF = Configuration Register; PMC = Port Maintenance and Control Register; PSR = Port Status Register), which are optionally displayed for certain of the error conditions. The codes, always file accessible, specify whether the message is *always* logged on OPA0 or is logged only when the system device is *inaccessible*.

Software Errors During Initialization (Always Logged on OPA0)

```
%Pxxn, Insufficient Non-Paged Pool for Initialization
%Pxxn, Failed to Locate Port Micro-code Image
%Pxxn, SCSSYSTEMID has NOT been set to a Non-Zero Value
```

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

Hardware Errors (Always Logged on OPA0)

%Pxxn, BIIC failure - BICSR/BER/CNF xxxxxx/xxxxxx/xxxxxx
%Pxxn, Micro-code Verification Error
%Pxxn, Port Transition Failure - CNF/PMC/PSR xxxxxx/xxxxxx/xxxxxx
%Pxxn, Port Error Bit(s) Set - CNF/PMC/PSR xxxxxx/xxxxxx/xxxxxx
%Pxxn, Port Power Down
%Pxxn, Port Power Up
%Pxxn, Unexpected Interrupt - CNF/PMC/PSR xxxxxx/xxxxxx/xxxxxx
%Pxxn, CI Port Timeout
%Pxxn, CI port ucode not at required rev level. RAM/PROM rev is xxxx/xxxx
%Pxxn, CI port ucode not at current rev level. RAM/PROM rev is xxxx/xxxx
%Pxxn, CPU ucode not at required rev level for CI activity

Queue Interlock Failures (Always Logged on OPA0)

%Pxxn, Message Free Queue Remove Failure
%Pxxn, Datagram Free Queue Remove Failure
%Pxxn, Response Queue Remove Failure
%Pxxn, High Priority Command Queue Insert Failure
%Pxxn, Low Priority Command Queue Insert Failure
%Pxxn, Message Free Queue Insert Failure
%Pxxn, Datagram Free Queue Insert Failure

Errors Signaled with a CI Packet

%Pxxn, Unrecognized SCA Packet - FLAGS/OPC/STATUS/PORT xx/xx/xx/xx
(ALWAYS)
%Pxxn, Port has Closed Virtual Circuit - REMOTE PORT xxx
(ALWAYS)
%Pxxn, Software Shutting Down Port
(ALWAYS)
%Pxxn, Software is Closing Virtual Circuit - REMOTE PORT xxx
(ALWAYS)
%Pxxn, Received Connect Without Path-Block - FLAGS/OPC/STATUS/PORT xx/xx/xx/xx
(ALWAYS)
%Pxxn, Inappropriate SCA Control Message - FLAGS/OPC/STATUS/PORT xx/xx/xx/xx
(ALWAYS)
%Pxxn, No Path-Block During Virtual Circuit Close - REMOTE PORT xxx
(ALWAYS)
%Pxxn, HSC Error Logging Datagram Received - REMOTE PORT xxx
(INACCESSIBLE)
%Pxxn, Remote System Conflicts with Known System - REMOTE PORT xxx
(ALWAYS)
%Pxxn, Virtual Circuit Timeout - REMOTE PORT xxx
(ALWAYS)

Cluster Troubleshooting Information

C.4 Diagnosing VAXport Device Problems

%Pxxn, Parallel Path is Closing Virtual Circuit - REMOTE PORT xxx
(ALWAYS)

%Pxxn, Insufficient Non-paged Pool for Virtual Circuits
(ALWAYS)

Cable Change-of-State Notification

%Pxxn, Path #0. Has gone from GOOD to BAD - REMOTE PORT xxx
(INACCESSIBLE)

%Pxxn, Path #1. Has gone from GOOD to BAD - REMOTE PORT xxx
(INACCESSIBLE)

%Pxxn, Path #0. Has gone from BAD to GOOD - REMOTE PORT xxx
(INACCESSIBLE)

%Pxxn, Path #1. Has gone from BAD to GOOD - REMOTE PORT xxx
(INACCESSIBLE)

%Pxxn, Cables have gone from UNCROSSED to CROSSED - REMOTE PORT xxx
(INACCESSIBLE)

%Pxxn, Cables have gone from CROSSED to UNCROSSED - REMOTE PORT xxx
(INACCESSIBLE)

%Pxxn, Path #0. Loopback has gone from GOOD to BAD - REMOTE PORT xxx
(ALWAYS)

%Pxxn, Path #1. Loopback has gone from GOOD to BAD - REMOTE PORT xxx
(ALWAYS)

%Pxxn, Path #0. Loopback has gone from BAD to GOOD - REMOTE PORT xxx
(ALWAYS)

%Pxxn, Path #1. Loopback has gone from BAD to GOOD - REMOTE PORT xxx
(ALWAYS)

%Pxxn, Path #0. Has become working but CROSSED to Path #1. - REMOTE PORT xxx
(INACCESSIBLE)

%Pxxn, Path #1. Has become working but CROSSED to Path #0. - REMOTE PORT xxx
(INACCESSIBLE)

Note that if the port driver can identify the remote SCS node name of the affected computer, the driver replaces the "REMOTE PORT xxx" text with "REMOTE SYSTEM X...", where X... is the value of the SYSGEN parameter SCSNODE on the remote computer. If the remote SCS node name is not available, the port driver uses the existing message format.

Two other messages concerning the CI port appear on OPA0. They are as follows:

%Pxxn, CI port is reinitializing (xxx retries left.)

%Pxxn, CI port is going off line.

The first message indicates that a previous error requiring the port to shut down is recoverable, and that the port will be reinitialized. The "xxx retries left information" specifies how many more reinitializations are allowed before the port must be left permanently off line. Each reinitialization of the port (for reasons other than power fail recovery) causes approximately 2Kb of nonpaged pool to be lost.

The second message indicates that a previous error is not recoverable, and that the port will be left off line. In this case, the only way to recover the port is to reboot the computer.

Index

A

- ACP_REBLDSYSD parameter • 3–13
- Adding a computer • 5–7, 5–23, 5–38
 - adjusting EXPECTED_VOTES • 5–23
- Allocation class • 3–7
 - assigning value to computers • 3–8
 - assigning value to HSC subsystems • 3–8
 - determining for mixed-interconnect cluster • 5–4
 - rules for specifying • 3–7
 - sample configurations • 3–8
- Authorize Utility (AUTHORIZE) • B–1
- AUTOGEN.COM command procedure
 - enabling or disabling disk server • 5–14
 - executed by CLUSTER_CONFIG.COM • 5–2
 - running with feedback option • 5–25, 5–38
 - specifying dump file • 5–36

B

- Batch queue • 4–6
 - assigning unique name to • 4–8
 - clusterwide generic • 4–8
 - initializing • 4–8
 - sample configuration • 4–6
 - setting up • 4–7
 - starting • 4–8
 - SYS\$BATCH • 4–8
- Boot events • C–1
- Boot node
 - See Boot server
- Boot server
 - functions • 1–6
 - selecting • 5–3
- Broadcast messages • 5–12

C

- CI (Computer Interconnect) • 1–3, C–1
 - analyzing error log entry • C–16
 - cable repair • C–16
 - communication path • C–11

- CI (Computer Interconnect) (Cont.)
 - connected computer
 - failure to boot • C–3
 - failure to join the cluster • C–6
 - device-attention entry • C–17
 - error log entry • C–23
 - error log entry formats • C–17
 - logged message entry • C–20
 - port
 - loopback datagram facility • C–13
 - polling • C–10
 - verifying function • C–12
- CI-connected computer
 - adding • 5–7
- CLUEXIT bugcheck
 - diagnosing • C–9
- Cluster
 - boot events • C–1
 - CLUEXIT bugcheck • C–9
 - common-environment • 2–1
 - communication mechanisms • 1–5
 - configuration planning • 1–17
 - configuration type • 1–5
 - changing to mixed-interconnect • 5–19
 - configuring • 5–5, 5–32
 - connection manager • 1–4
 - disk class driver • 1–5
 - disk devices • 3–1
 - distributed file system • 1–4
 - distributed job controller • 1–5
 - distributed lock manager • 1–4
 - distributed processing • 1–2
 - error log entries for VAXport device • C–16
 - failure of computer to boot • C–1
 - failure of computer to join the cluster • C–1, C–6
 - group number • 5–31
 - hang condition • C–7
 - hardware components • 1–3
 - installing license • 2–5
 - interconnect devices • 1–3
 - maintaining • 5–24
 - managing queues • 4–1
 - MSCP server • 1–5
 - multiple-environment • 2–1
 - password • 5–31
 - preparing operating environment • 2–1
 - queues • 1–2

Index

Cluster (Cont.)

- quorum disk • 1–14
- quorum disk watcher • 1–14
- quorum file (QUORUM.DAT) • 1–14
- rebooting after configuration change • 5–23
- reconfiguring • 5–23
- recording configuration data • 5–25
- recovering from startup procedure failure • C–7
- resource access • 1–4
- resource locking • 1–4
- restoring quorum • 5–26
- security management • 1–11, 5–30
- shared disk resources • 1–2
- shared processing and printer resources • 1–2
- summary of configuration procedures • 1–17
- System Communications Services (SCS) • 1–4, C–10
- tape devices • 3–1
- troubleshooting • C–1
- types of operating environments • 2–1
- updating MODPARAMS.DAT files • 5–23
- VAXport device error log entries • C–16
- VAXport driver • 1–4, C–10
- voting member • 1–13
 - adding • 5–3, 5–7, 5–23
 - removing • 5–3, 5–13, 5–23
- workload balancing • 1–2
- Cluster-accessible disks • 1–2, 3–1
 - and MSCP server • 3–2
- Cluster authorization file (CLUSTER_AUTHORIZE.DAT) • 1–12, 5–31
- Cluster SYSGEN parameters • A–1 to A–2
- CLUSTER_CONFIG.COM command procedure
 - adding computers • 5–6
 - change options • 5–14
 - converting standalone computer to cluster computer • 5–21
 - creating a duplicate system disk • 5–21
 - enabling disk server • 3–3, 5–16
 - functions • 5–1
 - modifying satellite Ethernet hardware address • 5–14
 - preparing to execute • 5–5
 - removing computers • 5–13
 - required information • 5–5
 - system files created for satellite • 5–2
- Common command procedure
 - coordinating • 2–9
 - creating • 2–10, 2–11
 - executing • 2–10
 - on cluster-accessible disks • 2–9
 - SYLOGIN.COM • 2–11

- Common-environment cluster • 2–1, 2–10
- Common file
 - coordinating for multiple boot servers • 2–15
 - coordinating for multiple system disks • 2–15
 - JBCSYSQUE.DAT • 4–2
 - mail database • 2–14
 - moving off system disk • 5–36
 - NETPROXY.DAT • 2–12
 - RIGHTSLIST.DAT • 2–13
 - system • 2–12
 - SYSUAF.DAT • 2–12
 - VMSMAIL_PROFILE.DATA • 2–14
- Common MAIL database • 2–14
- Common rights database • 2–13
- Common system disk
 - directory structure • 2–2
- Computer Interconnect (CI)
 - See CI
- Computer-specific startup functions • 2–11
- Configuration data
 - recording • 5–25
- Configuration type
 - changing from CI-only to mixed-interconnect • 5–19
 - changing from local area to mixed-interconnect • 5–20
- Connection manager • 1–4, 1–12
- Conversational bootstrap
 - controlling • 5–32
- Convert Utility (CONVERT) • B–2
 - using to merge SYSUAF.DAT files • B–1
- Crossed cable • C–13

D

- DECnet-VAX network
 - cluster functions • 1–5
 - configuring • 2–6
 - copying remote node databases in VAXcluster environments • 2–8
 - enabling circuit service for cluster boot server • 2–6
 - installing license • 2–5
 - making databases available clusterwide • 2–8
 - making remote node data available clusterwide • 2–6
 - maximum address value, defining for cluster boot server • 2–6
 - modifying satellite Ethernet hardware address • 5–14

DECnet-VAX network (Cont.)

- NETCONFIG.COM command procedure • 2-7
- NETNODE_REMOTE.DAT file, renaming to
SYS\$COMMON directory • 2-8
- Network Control Program (NCP) • 2-8
- restoring satellite configuration data • 5-12
- starting • 2-8
- tailoring • 2-6
- VAXcluster alias • 2-7, 2-9, 5-41

Device driver

- loading • 2-10

Directory structure

- on common system disk • 2-2

Disk

- See also Dual-pathed disk
- See also Dual-ported disk
- cluster-accessible • 3-1
 - storing common procedures on • 2-9
- command procedures for setting up • 2-11
- configuring • 3-12
- directory structure on common system disk • 2-2
- DSA • 3-5
- DSA controller • 3-2
- DSSI • 3-5
- dual-pathed • 3-1
- HSC • 3-1, 3-2, 3-8
- local • 3-2
- managing • 3-1
- MASSBUS • 3-6
- mounting • 3-12
- MOUNT/NOREBUILD • 3-12
- MSCP-served • 3-1
- naming conventions • 3-7
- paths • 3-7
- quorum • 1-14
- rebuilding • 3-12
- restricted access • 3-1
- setting up • 2-11

Disk class driver • 1-5

Disk controller • 1-2

Disk server

- configuring Ethernet adapter • 5-33
- configuring memory • 5-33
- functions • 1-6
- selecting • 5-3

DISK_QUORUM parameter • 1-14

Distributed file system • 1-4

Distributed job controller • 1-5

Distributed lock manager • 1-4

Distributed processing • 1-2, 4-1

DSA disk • 3-5

DSSI-based configuration

- See Dual-host VAXcluster configuration

DSSI disk • 3-5

Dual-host VAXcluster configuration • 1-9

- rules • 1-10

Dual-pathed disk • 3-4

- DSA • 3-5
- DSSI • 3-5
- HSC • 3-4, 3-8
- MASSBUS • 3-6

Dual-ported disk

- DSA • 3-5
- MASSBUS • 3-6
- setting up • 2-10

Dump file

- controlling size • 5-36
- managing • 5-36
- sharing • 5-37

DUMPFIL AUTOGEN symbol • 5-36

DUMPSYLE AUTOGEN symbol • 5-36

Duplicate system disk

- creating • 5-21

E

Ethernet

- configuring adapter • 5-33
- error log entry • C-23
- hardware address • 5-5
- monitoring activity • 5-26
- port • C-10

EXPECTED_VOTES parameter • 1-13, 5-23, 5-27

F

Failover

- dual-host VAXcluster configuration • 1-9
- dual-ported DSA disk • 3-5

File access

- controlling • 2-12

File system

- coordinating • 2-12

Index

G

Generic queue

- clusterwide batch • 4–8
- clusterwide printer • 4–4
- establishing • 4–4

Group number

See Security management

H

Hang condition

- diagnosing • C–7

Hierarchical storage controller (HSC) subsystem

See HSC subsystem

HSC disk • 1–2, 1–10, 3–2

- dual-pathed • 3–4, 3–8

HSC subsystem • 1–2

- changing allocation class values • 5–24

J

Job controller • 1–5

Job-controller queue file (JBCSYSQUE.DAT) • 1–2, 4–10

- sharing • 2–12
- specifying location of • 4–2

K

Known images

- installing • 2–11

L

Local area cluster • 1–6

- creating cluster security database • 1–11
- monitoring Ethernet activity • 5–26

Local disk

- setting up • 2–10

Logical name

- defining • 2–11
- defining for NETPROXY.DAT • 2–13
- defining for RIGHTSLIST.DAT • 2–13
- defining for SYLOGIN.COM • 2–10
- defining for SYSUAF.DAT • 2–13
- defining for VMSMAIL_PROFILE.DATA • 2–14
- system • 2–2

Login

- controlling • 2–12

M

MAIL database

- preparing common file • 2–14

Mail Utility (MAIL)

- controlling • 2–12
- preparing common database • 2–14

MASSBUS disk

- dual-ported • 3–6

Mixed-interconnect cluster

- changing allocation class values on HSC subsystems • 5–24
- creating cluster security database • 1–11
- determining allocation class values • 5–4
- monitoring Ethernet activity • 5–26
- MSCP-served HSC disk • 1–10

MODPARAMS.DAT file

- created by CLUSTER_CONFIG.COM • 5–2
- specifying dump file • 5–36
- specifying MSCP disk-serving parameters • 3–3
- updating • 5–23

Mounting disks • 3–12

MSCP server • 1–5

- and cluster-accessible disks • 3–2
- initializing • 3–3
- loading • 3–3
- load sharing • 3–3

MSCP_LOAD parameter • 3–3

MSCP_SERVE_ALL parameter • 3–3

Multiple-environment cluster • 2–1, 2–11

N

NETCONFIG.COM command procedure

See DECnet–VAX network

NETNODE_REMOTE.DAT
 renaming to SYS\$COMMON directory • 2–8

NETNODE_REMOTE.DAT file
 sharing • 2–12

NETNODE_UPDATE.COM command procedure • 5–12

NETPROXY.DAT file
 creating common version • 2–12
 defining logical name for • 2–13
 setting up • 2–12
 sharing • 2–12

Network
 See DECnet-VAX network

Network Control Program (NCP)
 See DECnet-VAX network

O

OPA0: workstation operator console terminal • 5–12

OPCOM messages • 5–12

Operating system
 coordinating files • 2–12
 installing • 2–4
 upgrading • 2–4

P

Page file (PAGEFILE.SYS)
 created by CLUSTER_CONFIG.COM • 5–2, 5–3

Partitioning of cluster • 1–12, C–9

Password
 See Security management

Port select button • 3–4

Printer queue • 4–2
 assigning unique name to • 4–3
 initializing • 4–4
 sample configuration • 4–2
 setting up • 4–2
 starting • 4–4

Proxy login
 controlling • 2–12
 records • 2–13

Q

QDSKVOTES parameter • 1–14

Queue
 batch
 See Batch queue

command procedures • 2–11

common command procedure • 4–10

controlling • 1–2, 4–1

printer
 See Printer queue

setting up • 2–11

single-computer and cluster • 4–1

Quorum
 adjusting EXPECTED_VOTES • 5–23

equation • 1–13

EXPECTED_VOTES parameter • 1–13, 5–23, 5–27

loss causes cluster hang condition • C–8

lowering value • 5–27

reasons for loss • C–8

restoring after unexpected computer failure • 5–26

VOTES parameter • 1–13

voting member • 1–13
 adding • 5–3, 5–7, 5–23
 removing • 5–3, 5–13, 5–23

QUORUM.DAT file • 1–14

Quorum disk • 1–14
 adding • 5–23
 adjusting EXPECTED_VOTES • 5–23
 disabling • 5–3
 enabling • 5–3
 mounting • 1–14
 removing • 5–23

Quorum disk watcher • 1–14

Quorum file (QUORUM.DAT) • 1–14

Quorum scheme • 1–12

R

Rebooting a satellite with operating system installed on local disk • 5–29

Reconfiguring the cluster • 5–23

Remote network node data
 controlling • 2–12

Remote node databases
 copying • 2–8

Removing a computer • 5–13, 5–23
 adjusting EXPECTED_VOTES • 5–23
 shutting down before removing from cluster • 5–13

Removing a satellite • 5–13

Resource sharing • 1–12

Index

Restoring quorum • 5–26
Restoring satellite configuration data • 5–12
Restricted access disk • 3–1
RIGHTSLIST.DAT file
 defining logical name for • 2–13
 merging • B–2
 preparing common version of • 2–13
 sharing • 2–12

S

Satellite
 adding • 5–9
 disabling conversational bootstrap operations •
 5–32
 failure to boot • C–4
 failure to join the cluster • C–6
 functions • 1–7
 local disk used for paging and swapping • 1–7
 maintaining network configuration data • 5–12
 modifying Ethernet hardware address • 5–14
 obtaining Ethernet hardware address • 5–5
 rebooting if operating system installed on local
 disk • 5–29
 removing • 5–13
 restoring network configuration data • 5–12
 system files created by CLUSTER_CONFIG.COM
 • 5–2
SCS SYSGEN parameters • A–2 to A–4
Search list • 2–2
Security management
 controlling conversational bootstrap operations •
 5–32
 modifying cluster group number • 5–31
 modifying cluster password • 5–31
 overview • 5–30
SET CLUSTER/EXPECTED_VOTES command •
 5–27
Shared command procedure files • 2–9
Shared disk volume • 3–11
 for job-controller queue file • 4–10
 mounting • 3–11
Shared file
 NETPROXY.DAT • 2–12
 SYSUAF.DAT • 2–12
Shared files • 2–12
Shared queues • 4–1
Show Cluster Utility (SHOW CLUSTER) • 5–27
 CL_QUORUM • 5–27
 CL_VOTES • 5–27
Shutting down the cluster • 5–27
Site-specific startup command file
 elements • 2–11
Standalone computer
 converting to cluster computer • 5–21
Star coupler • 1–3
Star coupler expander (CISCE) • 1–3
START/QUEUE/MANAGER command • 4–2
Startup
 computer-specific function • 2–11
Startup command file
 coordinating • 2–9
 creating common version • 2–10, 2–11
 site-specific elements • 2–11
Startup procedure
 failure to complete • C–7
Swap file (SWAPFILE.SYS)
 created by CLUSTER_CONFIG.COM • 5–2, 5–3
SYLOGIN.COM file
 coordinating • 2–10
 creating common version • 2–10, 2–11
 defining logical name for • 2–10
SYSBOOT.EXE image
 renaming before rebooting satellite • 5–30
SYSGEN parameters
 cluster parameters • A–1 to A–2
 SCS parameters • A–2 to A–4
SYSMAN Utility
 enabling VAXcluster alias operations • 2–9
 modifying cluster security data • 5–31
System command procedures
 coordinating • 2–9
System Communications Services (SCS) • 1–4, C–10
System directory • 2–2
System disk
 configuring in large cluster • 5–33, 5–36
 creating duplicate • 5–21
 directory structure • 2–2
 moving high-activity files • 5–36
 rebuilding • 3–13
System files • 2–12
SYSUAF.DAT file
 creating common version • 2–12
 defining logical name for • 2–13
 merging • B–1
 printing listing of • B–1
 setting up • 2–12
 sharing • 2–12

T

Troubleshooting

- analyzing VAXport error log entries • C-16
- CLUEXIT bugcheck • C-9
- error log entries for CI and Ethernet ports • C-23
- failure of computer to boot • C-1
- failure of computer to join the cluster • C-1, C-6
- failure of startup procedure to complete • C-7
- hang condition • C-7
- loss of quorum • C-8
- OPA0 error messages • C-30
- repairing CI cables • C-16
- shared resource is inaccessible • C-8
- VAXport device problem • C-9
- verifying CI cable connections • C-13
- verifying CI port • C-12

U

- Upgraded systems • 2-4
- User accounts
 - comparing • B-1
 - coordinating • 2-12, B-1
 - group UIC • B-1
- User environment
 - common-environment cluster • 2-1
 - creating common-environment cluster • 2-10
 - defining • 2-12
 - multiple-environment cluster • 2-1
- User Environment Test Package (UETP)
 - creating command procedure to run • 5-39
 - running in large cluster • 5-39
 - specifying values for LOAD phase • 5-39
- User identification code (UIC) • B-1

V

- VAXcluster alias
 - defining • 2-7, 5-41
 - enabling operations • 2-9
- VAXport communication • C-10
- VAXport driver • 1-4, C-10
- VAXVMSSYS.PAR file
 - created by CLUSTER_CONFIG.COM • 5-2
- Virtual circuit • C-10

- VMSMAIL_PROFILE.DATA
 - defining logical name for • 2-14
- VMSMAIL_PROFILE.DATA file
 - preparing common version of • 2-14
 - sharing • 2-12
- VMS operating system
 - installing license • 2-5
- VMS RMS distributed file system • 1-4
- Volume label
 - modifying for satellite's local disk • 5-3
- Volume shadowing
 - in mixed-interconnect cluster • 5-35
- VOTES parameter • 1-13
- Voting member • 1-13
 - adding • 5-3, 5-7, 5-23
 - removing • 5-3, 5-13, 5-23

W

- Workload balancing • 1-2, 4-1

How to Order Additional Documentation

Technical Support

If you need help deciding which documentation best meets your needs, call 800-343-4040 before placing your electronic, telephone, or direct mail order.

Electronic Orders

To place an order at the Electronic Store, dial 800-DEC-DEMO (800-332-3366) using a 1200- or 2400-baud modem. If you need assistance using the Electronic Store, call 800-DIGITAL (800-344-4825).

Telephone and Direct Mail Orders

Your Location	Call	Contact
Continental USA, Alaska, or Hawaii	800-DIGITAL	Digital Equipment Corporation P.O. Box CS2008 Nashua, New Hampshire 03061
Puerto Rico	809-754-7575	Local Digital sub- sidiary
Canada	800-267-6215	Digital Equipment of Canada Attn: DECdirect Operations KAO2/2 P.O. Box 13000 100 Herzberg Road Kanata, Ontario, Canada K2K 2A6
International	_____	Local Digital sub- sidiary or approved distributor
Internal ¹	_____	USASSB Order Processing - WMO /E15 <i>or</i> U.S. Area Software Supply Business Digital Equipment Corporation Westminster, Massachusetts 01473

¹For internal orders, you must submit an Internal Software Order Form (EN-01740-07).

Reader's Comments

VMS VAXcluster Manual
AA-LA27B-TE

Please use this postage-paid form to comment on this manual. If you require a written reply to a software problem and are eligible to receive one under Software Performance Report (SPR) service, submit your comments on an SPR form.

Thank you for your assistance.

I rate this manual's:	Excellent	Good	Fair	Poor
Accuracy (software works as manual says)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completeness (enough information)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clarity (easy to understand)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Organization (structure of subject matter)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Figures (useful)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Examples (useful)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Index (ability to find topic)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Page layout (easy to find information)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I would like to see more/less _____

What I like best about this manual is _____

What I like least about this manual is _____

I found the following errors in this manual:

Page	Description
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

Additional comments or suggestions to improve this manual:

I am using **Version** _____ of the software this manual describes.

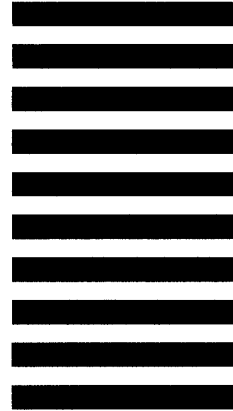
Name/Title _____ Dept. _____
Company _____ Date _____
Mailing Address _____
_____ Phone _____

--- Do Not Tear - Fold Here and Tape ---

digitalTM



No Postage
Necessary
if Mailed
in the
United States



BUSINESS REPLY MAIL
FIRST CLASS PERMIT NO. 33 MAYNARD MASS.

POSTAGE WILL BE PAID BY ADRESSEE

DIGITAL EQUIPMENT CORPORATION
Corporate User Publications — Spit Brook
ZKO1-3/J35 110 SPIT BROOK ROAD
NASHUA, NH 03062-9987



--- Do Not Tear - Fold Here ---

Cut Along Dotted Line

Reader's Comments

VMS VAXcluster Manual
AA-LA27B-TE

Please use this postage-paid form to comment on this manual. If you require a written reply to a software problem and are eligible to receive one under Software Performance Report (SPR) service, submit your comments on an SPR form.

Thank you for your assistance.

I rate this manual's:	Excellent	Good	Fair	Poor
Accuracy (software works as manual says)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completeness (enough information)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clarity (easy to understand)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Organization (structure of subject matter)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Figures (useful)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Examples (useful)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Index (ability to find topic)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Page layout (easy to find information)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I would like to see more/less _____

What I like best about this manual is _____

What I like least about this manual is _____

I found the following errors in this manual:

Page	Description
_____	_____
_____	_____
_____	_____
_____	_____

Additional comments or suggestions to improve this manual:

I am using **Version** _____ of the software this manual describes.

Name/Title _____ Dept. _____
Company _____ Date _____
Mailing Address _____
Phone _____

----- Do Not Tear - Fold Here and Tape -----

digitalTM



No Postage
Necessary
if Mailed
in the
United States



BUSINESS REPLY MAIL
FIRST CLASS PERMIT NO. 33 MAYNARD MASS.

POSTAGE WILL BE PAID BY ADDRESSEE

DIGITAL EQUIPMENT CORPORATION
Corporate User Publications — Spit Brook
ZKO1-3/J35 110 SPIT BROOK ROAD
NASHUA, NH 03062-9987



----- Do Not Tear - Fold Here -----