

VAXcluster Systems

Guidelines for VAXcluster System Configurations

Part Number: EK-VAXCT-CG-006

September 1992

NOTE

Because there are few absolute rules for VAXcluster system configurations, readers should use the material in this document to supplement their basic understanding of the VAXcluster environment.

Revision/Update Information: This revised document supersedes *Guidelines for VAXcluster System Configurations*, Part Number EK-VAXCS-CG-005.

Operating System and Version: VMS Version 5.5-2

Digital Equipment Corporation
Maynard, Massachusetts

September 1992

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.

Copyright © Digital Equipment Corporation 1992. All rights reserved.

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: ACMS, ALL-IN-1, BI, CI, CMI, DEC, DECalert, DECcams, DECbridge, DECcp, DECdirect, DECdtm, DECintact, DECmate, DECmcc, DECnet, DECpage, DECperformance, DECrouter, DECscheduler, DECserver, DECsystem, DECtalk, DEC VTX, DECwindows, DELUA, DEQNA, DEUNA, Digital, Electronic Store, HSC, KDA, KDM, LA, LAT, LN03, LP27, MASSBUS, MicroPower/Pascal MicroVAX, MSCP, MUXserver, NMI, PrintServer, Q-bus, RA, RK, RL, RM, RP, RQC25, RQDX3, RRD50, RV20, SA, SBI, SDI, SPM, STI, TA, TK, TMSCP, TU, UDA, UNIBUS, VAX, VAX Ada, VAX APL, VAX BASIC, VAX C, VAX COBOL, VAX COBOL Generator, VAX DATATRIEVE, VAX DBMS, VAX DOCUMENT, VAX DSM, VAX FMS, VAX FORTRAN, VAX LIMS/SM, VAX LISP, VAX MAILGATE, VAX NOTES, VAX OPS5, VAX PASCAL, VAX Performance Advisor, VAX RALLY, VAX Rdb/ELN, VAX RMS, VAX SCAN, VAX ScriptPrinter, VAX SQL, VAX TEAMNDATA, VAX VALU, VAX Xway, VAXBI, VAXcluster, VAXELN, VAXft, VAXinfo I, VAXinfo II, VAXinfo III, VAXmail, VAXserver, VAXset, VAXsimPLUS, VAXstation, VIDA, VMS, VNXset, VT, WPS-PLUS, and the DIGITAL logo.

The following are third-party trademarks: IBM is a registered trademark of International Business Machines Corporation; PostScript is a registered trademark of Adobe Systems, Inc. StorageTek is a registered trademark of Storage Technology Corporation.

This document is available on CD-ROM.

This document was prepared using VAX DOCUMENT, Version 2.1.

Contents

Preface	ix
1 Digital Computing Environment	
1.1 VAXcluster System Environment	1-1
1.1.1 VAXcluster System Components	1-2
1.1.2 VAXcluster System Configuration Types	1-3
2 Establishing Your VAXcluster Requirements	
2.1 Defining Your Computing Environment	2-1
2.2 Considering Your Application Requirements	2-1
2.3 Determining Your Overall System Requirements	2-2
3 Choosing Your VAX CPUs	
3.1 CPU Selection Guidelines	3-1
3.1.1 Determining Application Requirements for CPUs	3-1
3.1.1.1 Computing Style	3-2
3.1.1.2 Availability	3-2
3.1.1.3 Growth	3-2
3.1.1.4 I/O Requirements	3-2
3.1.2 Selecting Your VAX CPU Configuration	3-3
3.1.2.1 VAXcluster System and Fault-Tolerant CPUs	3-3
3.1.2.2 VAXcluster Multi-Datacenter Facility Systems	3-3
3.1.2.3 Determining the Number of CPUs	3-3
3.1.2.4 Determining Memory Requirements	3-4
3.2 VAX CPU Characteristics and Positioning	3-4
4 Choosing Your VAXcluster Interconnect	
4.1 Interconnect Characteristics	4-1
4.1.1 Throughput	4-1
4.1.2 CPU Overhead	4-1
4.2 Interconnect Types	4-2
4.2.1 Ethernet Interconnect	4-2
4.2.2 CI Interconnect	4-3
4.2.3 Digital Storage Systems Interconnect	4-5
4.2.4 FDDI Interconnect	4-6
4.3 Mixed-Interconnect Configurations	4-6
4.4 Multiple Interconnects of the Same Type	4-7
4.4.1 Multiple CI Interconnects	4-7
4.4.2 Multiple DSSIs	4-7
4.4.3 Multiple Ethernets	4-7

4.4.4	Multiple FDDIs	4-8
4.5	SCSI Bus	4-8

5 Designing Your Storage Subsystem

5.1	Storage Subsystem Design Preliminaries	5-1
5.2	Description of a Storage Hierarchy	5-2
5.3	How to Design Your Storage Hierarchy	5-4
5.3.1	Gather Application I/O Requirements	5-5
5.3.2	Estimate Storage Capacity Requirements	5-5
5.3.3	Consider Additional Storage Subsystem Attributes	5-8
5.3.3.1	How CPU Selections Affect Storage Subsystem Design	5-8
5.3.3.2	Select Storage Devices to Meet Your Requirements	5-8
5.3.4	Gather Additional Data	5-8
5.3.4.1	Online Storage Performance and Availability Work Sheet	5-9
5.3.4.2	Modify the Storage Hierarchy According to Performance Requirements	5-10
5.3.4.3	Performance Considerations for HSC Controller Subsystems	5-10
5.3.4.4	Backup Storage Performance	5-11
5.3.4.5	Performance Considerations When Including Newer Technology Storage	5-11
5.3.5	Meet Availability Requirements	5-12
5.3.5.1	Complete Your Online Storage Performance and Availability Work Sheet	5-12
5.3.5.2	Device Availability	5-13
5.3.5.3	Data Availability	5-13
5.3.5.4	System Disk Redundancy	5-15
5.3.5.5	Site Redundancy	5-15
5.3.5.6	VMS Volume Shadowing Software	5-16
5.3.5.7	Redundancy Through Backup Strategy	5-18
5.3.6	Storage Management Considerations	5-18
5.3.6.1	Disk Utilization and Fragmentation	5-19
5.3.6.2	Backup	5-19
5.3.6.3	VAX SLS	5-20
5.3.6.4	Floor Space for your Storage Devices	5-20
5.4	Storage Device Characteristics	5-21

6 VAXcluster Configuration Rules and Guidelines

6.1	General VAXcluster Configuration Rules	6-1
6.2	Configuration Rules for CI VAXcluster Systems	6-1
6.2.1	Configuration Rules for CPUs with Multiple CI Connections	6-2
6.2.2	Additional Guidelines for CI VAXcluster Systems	6-2
6.3	Configuration Rules for DSSI VAXcluster Systems	6-4
6.3.1	Configuration Rules for CPUs with Multiple DSSI Connections	6-5
6.3.2	Additional Guidelines for DSSI VAXcluster Systems	6-5
6.4	Configuration Rules for Ethernet VAXcluster Systems	6-7
6.5	Configuration Rules for FDDI VAXcluster Systems	6-7
6.6	Configuration Rules for CPUs with Multiple LAN (Ethernet or FDDI) Adapters	6-7

7 Optimizing VAXcluster System Design

7.1	Increasing Availability	7-1
7.1.1	Hardware Redundancy Methods	7-2
7.1.2	Failover Mechanisms	7-3
7.1.3	Environmental Protection	7-4
7.1.4	Quorum scheme	7-4
7.1.5	VAXcluster State Transitions	7-5
7.2	Guidelines for Selecting Disk Servers and Satellites	7-5
7.2.1	Disk Server I/O Capacity	7-5
7.2.2	CPU I/O Capacity	7-6
7.2.3	Ethernet I/O Capacity	7-7
7.2.4	Disk Drive I/O Capacity	7-7
7.2.5	Summary of Disk Server I/O Capacity	7-7
7.3	Lock Manager	7-8
7.4	Backup Strategy	7-8
7.4.1	Using HSC BACKUP or VMS BACKUP	7-9
7.5	Configuring System Disks	7-9
7.5.1	Booting Activity	7-10
7.6	Print Services in Your VAXcluster System	7-10
7.7	Tools for Managing your VAXcluster System	7-11
7.7.1	CLUSTER_CONFIG.COM Command Procedure	7-11
7.7.2	Show Cluster Utility	7-12
7.7.3	System Management Utility	7-12
7.7.4	VMS Local Area VAXcluster Failure Analysis	7-12
7.7.5	Optional Tools and Products	7-13

A SPD Disk Storage Requirements

B VAXcluster Software Product Description

B.1	Description	B-1
B.2	Hardware Support	B-6
B.3	Software Requirements	B-8
B.4	Optional Software	B-8
B.5	Growth Considerations	B-8
B.6	Ordering Information	B-9
B.7	Software Licensing	B-9
B.8	Software Product Services	B-9
B.9	Software Warranty	B-9

C Specifications for Mature Products

C.1	CPU Information	C-1
C.2	Adapter Information	C-3
C.3	Storage Information	C-4
C.4	Printer Information	C-7

Glossary

Index

Figures

1–1	Ethernet VAXcluster Configuration	1–5
1–2	Typical VAXcluster Configuration with CI	1–6
1–3	DSSI VAXcluster Configuration	1–6
1–4	Multiple DSSI VAXcluster Configuration	1–7
1–5	FDDI Multi-Datacenter LAN-based VAXcluster	1–8
1–6	Large VAXcluster Configuration with Multiple Interconnects	1–9
5–1	Storage Hierarchy	5–2
5–2	Storage Hierarchy in a Simple Configuration	5–3
5–3	Complex Storage Subsystem	5–4
5–4	DSSI Shadow Sets	5–15
5–5	Dual-Hosted Mixed-interconnect Disks	5–16
5–6	Dual-Ported Mixed-interconnect Disks	5–17
6–1	Multi-CI VAXcluster System	6–3
6–2	Dual-Host MicroVAX 3400 Configuration	6–4
6–3	Multiple DSSI Segments in a VAXcluster System	6–6
6–4	Multiple LAN VAXcluster System	6–8

Tables

1–1	Attributes of VAXcluster Configuration Types	1–4
3–1	VAX System CPU Performance Characteristics	3–5
3–2	CPU Packaging Information	3–7
3–3	VAX System I/O Performance Characteristics	3–7
3–4	CPU Interconnect Options	3–8
3–5	CPU Storage Bus Options	3–8
4–1	Interconnect Connections	4–2
4–2	Ethernet Adapters	4–3
4–3	CI Adapters	4–4
4–4	DSSI Adapters for VAXcluster Systems	4–5
4–5	FDDI Adapter	4–6
5–1	Online Storage Capacity Work Sheet	5–7
5–2	Online Storage Performance and Availability Work Sheet	5–10
5–3	HSC Controller Channel/Drive Support	5–14
5–4	Storage Devices Listed by Interconnect	5–21
5–5	Disk Attributes	5–22
5–6	Storage Array Attributes	5–23
5–7	Library Attributes	5–23
5–8	Tape Attributes	5–24
5–9	Bus Type and I/O Rates for SDI Controllers	5–24
6–1	Maximum CI Adapters Per VAXcluster CPU	6–2
6–2	DSSI Adapters Per CPU	6–5

6-3	DSSI Bus ID Assignments	6-6
7-1	Disk Server I/O Capacity Based on 80% CPU Utilization	7-6
7-2	Ethernet Adapter I/O Capacity	7-7
7-3	Disk Server Capacity — Average (4-Block) I/O Operations Per Second	7-8
7-4	HSC BACKUP Versus VMS BACKUP	7-9
7-5	Selected Digital Printers	7-11
A-1	Space Required on VMS System Disk	A-1
B-1	Number and Type of Adapters Supported	B-6
B-2	LAN Adapters Supported	B-7
B-3	Trademark Information	B-9
C-1	VAX System CPU Performance Characteristics	C-1
C-2	VAX System I/O Performance Characteristics	C-2
C-3	Disk Server Capacity — Average (4-Block) I/O Operations Per Second	C-3
C-4	Ethernet Adapters	C-3
C-5	CI Adapters	C-3
C-6	Maximum CI Adapters Per VAXcluster CPU	C-4
C-7	DSSI Adapters Per CPU	C-4
C-8	Disk Attributes for Mature Drives	C-4
C-9	Library Attributes	C-5
C-10	Tape Attributes	C-5
C-11	Bus Type and I/O Rates for SDI Controllers	C-5
C-12	UNIBUS Storage Devices	C-5
C-13	MASSBUS Storage Devices	C-6
C-14	Disk Server I/O Capacity Based on 80 Percent CPU Utilization	C-6
C-15	Selected Digital Printers	C-7

Preface

In the Digital computing environment, there are several ways to increase the capabilities of a VAX computer system beyond those of a single processor. You can use the following options in many combinations:

- A multiprocessor central processing unit (CPU)
- A multi-CPU VAXcluster system
- A network of CPUs or VAXcluster systems

This document briefly compares these options and provides information on sizing and configuring VAXcluster systems.

If you determine that the VAXcluster solution is the one that best satisfies your computing needs, this document will help you develop a configuration for your purposes. Because of the flexibility of VAXcluster systems and the variety of components that you can combine in a VAXcluster system, there may be several configurations that will serve your purposes equally well.

Use of This Document

Use this document by going through it systematically from beginning to end. Each chapter addresses a specific area of software or hardware that you must consider when you configure a system. As you read each chapter, make a note of information that is pertinent to your system and salient product features that are vital to your configuration.

Once you accumulate that information, you can design a working model of the configuration you desire. You can then work with application and environmental information to refine the model and establish your VAXcluster configuration. VAXcluster design experts can provide insight into how your configuration can best meet your needs, and they can inform you of any new developments in the areas of product or configuration. Your sales representative can put you in touch with design and configuration experts to assist in your configuration design.

Intended Audience

This document is intended for those who would like assistance in designing VAXcluster configurations to meet their needs. It assumes an understanding of the concepts described in the *Guide to VAXcluster Systems*, Part Number EC-H0929-57.

Related Documents

The following publications will help you design your configuration:

- *VAX Systems/DECsystems Systems and Options Catalog* and its periodic supplements¹
- *Telecommunications and Networks Buyer's Guide*
- Software Product Descriptions (SPDs)
- *Digital Self Maintenance Services Catalog* catalog (for those maintaining VAXcluster systems themselves)

Most of these documents are available to you through the Digital Reference Service. Contact your Digital sales representative for assistance in obtaining a subscription to this service.

Other documents that may be useful as you develop your configuration include the following:

- *DECdirect Plus Catalog*
- *VAX Software Buyer's Guide*
- *Software Documentation Catalog*
- *VAXcluster Systems Quorum*, a technical journal for VAXcluster system management
- *VAXcluster Supplementary Documentation Kit* including:
 - *Introduction to VAXcluster Application Design*
 - *VAXcluster Operator's Guide*
 - *VAXcluster Documents*
- *DSSI VAXcluster Installation and Troubleshooting Manual*
- *VAXcluster Multi-Datacenter Facility: Configuration Guidelines*
- The most current SPD for VMS, VAXcluster, and DECnet software. The current VAXcluster SPD appears in Appendix B
- *VMS VAXcluster Manual*²
- *Guide to Setting Up a VMS System*²

For information on Digital support services (hardware, software, self maintenance), see the *VAX Systems/DECsystems Systems and Options Catalog*. To obtain a copy of the *DECdirect Plus Catalog*, which contains accessories, supplies, add-on products, and upgrade products, call toll-free 1-800-258-1710 (in the U.S.). To obtain access to the Digital Electronic Store, a free online computer service to help evaluate, select, and purchase certain Digital products, register for an account (in the U.S.) by calling 1-800-332-3366 at 1200/2400 baud³ from 8:00 a.m. to midnight eastern standard time. The Electronic Store provides a menu-driven system from which you can select and order many Digital products.

¹ Some components included in this manual are components that Digital no longer ships. Many such components can be used alongside newer equipment in VAXcluster systems.

² Part of the VMS documentation set.

³ Use any VT100, VT200, VT300, Rainbow, DECmate, PRO, VAXstation, or personal computer that emulates a VT100.

Digital Computing Environment

The Digital computing environment offers a broad range of compatible options. The diversity of products and capabilities of these options can permit flexibility in VAXcluster system design, but proper configuration is vital for success in meeting application needs.

VAXcluster systems offer a high degree of resource integration and system management control. VAXcluster systems also provide availability features that support critical applications. In addition, VAXcluster systems support the growth requirements, from a small, incremental increase in processing power to a quantum leap that can support a totally new application. Because these options are compatible, you can implement them at any time, in any combination, to keep pace with your computing needs.

This document provides information to help you design your VAXcluster system configuration.

1.1 VAXcluster System Environment

A VAXcluster system is a highly integrated configuration of VAX and MicroVAX computer systems and storage subsystems. All VAX and MicroVAX central processing units (CPUs) run the VMS operating system. As members of a VAXcluster system, the CPUs share processing resources, queues, and data storage under a single VMS security and management domain, and can boot or fail independently. VAXcluster systems provide you with the following features and benefits:

- You have a single security and management domain.
- You can use mass storage efficiently, because multiple CPUs can access the same storage device. You can use VMS Volume Shadowing for enhanced disk redundancy and access.
- Your users can share files clusterwide from any CPU at the record level with full read/write access. In addition, you can distribute applications across multiple CPUs.
- You can add processing and storage resources, without disturbing the rest of the system, to keep pace with user demand as the computing needs of your enterprise grow.
- You can add redundant devices and interconnects to increase availability or increase disaster tolerance.
- You can distribute batch and print job processing across the VAXcluster system. Jobs that access shared resources can execute on any CPU.
- You can optimize availability and redundancy of shared resources.

1.1.1 VAXcluster System Components

VAXcluster systems include the following software and hardware components:

- VMS operating system — This is the component of the VAXcluster system that provides support for the orderly sharing of VAXcluster resources, such as disks, files and queues. The VMS software components that coordinate sharing of VAXcluster resources are as follows:
 - System communication services (SCS) implement internode VAXcluster communications using Digital's System Communications Architecture (SCA).
 - The Connection Manager controls membership of the VAXcluster system.
 - The Distributed Lock Manager synchronizes access by many users to shared resources.
 - The Distributed Job Controller enables clusterwide sharing of print and batch queues to optimize usage of these resources.
 - The VMS file system and Record Management Services (RMS) provide transparent, shared read/write access to all files on all disks in a distributed VAXcluster environment.
 - The MSCP server makes locally connected disks available clusterwide. The TMSCP server makes locally connected tapes available to the VAXcluster system.
 - Clusterwide Process Services let VMS system management commands, such as SHOW SYSTEM or SHOW USERS, operate clusterwide.
 - DECnet-VAX network software — This software ensures that system managers can access each CPU in the VAXcluster system from a single terminal, without terminal-switching facilities. DECnet-VAX software also lets the VAXcluster system communicate with other resources as network nodes. In some configurations, DECnet-VAX software is also used to downline load the VMS operating system. In these configurations, DECnet-VAX software and SCS coexist on the same Ethernet.
 - DECnet System Services (DSS) products — DSS products include VAX Distributed File Service (DFS), VAX Distributed Name Service (DNS), VAX Distributed Queuing Service (DQS), and Remote System Manager (RSM). These products are appropriate when you need to communicate among systems over extended distances, perhaps worldwide.
 - CPUs — VAXcluster systems can include any currently supported VAX, MicroVAX, or workstation, subject to conditions specified in the current VAXcluster Software Product Description (SPD).
- Symmetrical multiprocessing (SMP) systems — These systems are effective for multistreamed and compute-intensive work loads, with enough processes ready to be scheduled to keep all processors busy. An application that requires a lot of processing time to manipulate pieces of data may benefit from a VMS multiprocessing system, rather than running in the traditional uniprocessing VAX CPU system.
- Storage subsystems — All Digital disk storage subsystems can be configured for local or clusterwide access.

- Interconnects — Depending on your VAXcluster configuration, internode communications can occur over the CI bus, Ethernet, Digital Storage Systems Interconnect (DSSI), Fiber Distributed Data Interface (FDDI), or any combination of these. In configurations where multiple interconnects are available, VAXcluster software first selects the CI or DSSI bus, then the FDDI or Ethernet.
- Terminal servers — VAXcluster systems can include DECserver terminal servers — terminal switches that use the Ethernet to connect terminals, printers, and modems to one or more CPUs.

1.1.2 VAXcluster System Configuration Types

You can create your VAXcluster system all at once or in stages, by adding CPUs, storage subsystems, and other components as needed. VAXcluster systems can be configured using the following interconnects:

- Ethernet
- CI
- DSSI
- FDDI
- Mixed-interconnect combination

NOTE

Configuration characteristics described in the next subsections conform to the VMS VAXcluster Software Version 5.5 SPD. Note that these characteristics can change with future releases of the VMS operating system. For detailed information on supported configurations, see the current VMS VAXcluster Software SPD.

Of the maximum possible 96 VMS nodes in a VAXcluster system, some are configured as *server nodes* and others are configured as *satellite nodes*.

There are two types of server nodes: disk servers and Maintenance Operation Protocol (MOP) servers. A disk server grants other nodes access to disk storage that the other nodes cannot directly access. Using the MOP protocol, a MOP server serves information from disks that enables satellites to access a system disk and load the VMS operating system from its own system disk to join the configuration. For best performance, MOP servers are usually the most powerful CPUs in an Ethernet VAXcluster system and should use the highest bandwidth Ethernet adapters you can employ.

Satellite nodes have no direct connection to a system disk. Instead, they are often booted remotely from a system disk that is directly accessible to the MOP server. Generally, these satellite nodes are consumers of VAXcluster resources, though they may also provide disk serving and batch processing resources.

Table 1–1 summarizes features of VAXcluster configurations that use the CI, Ethernet, FDDI, or DSSI for VAXcluster communications. Note that VAXcluster configurations with multiple interconnects often provide maximum work group integration, because such configurations can support the widest range of CPUs, from VAXstations to VAX 4000 and VAX 6000 series CPUs. You can use the information in the table as an overview as you determine which

configuration types best suit your needs. For more detailed information on supported configurations, see the current VAXcluster Software SPD.

Table 1–1 Attributes of VAXcluster Configuration Types

Configuration Attributes	Interconnect			
	CI	DSSI	Ethernet	FDDI
Availability	Highest	Medium	Medium	High
Power range of CPUs ¹	Medium-high	Low-high	Low-high	High
Include workstations	No	No	Yes	No
Interconnect speed (peak megabytes per second)	17.5 ²	4	1.25	12.5
Storage capacity ³	Large	Medium-high	Served storage only	Served storage only
Maximum number of CPU nodes per bus	16	4 ⁴	96	16
Maximum distance between nodes (in meters)	90 ⁵	27	2,800 ⁶	40,000 ⁷

¹Low = < 3 VUPs, medium = 4 to 12 VUPs, high = > 12 VUPs.

²Per Star Coupler.

³For simultaneously shared data.

⁴8 nodes at most on a DSSI bus, including up to 4 VMS nodes. The rest can be disk or tape devices.

⁵45 meters is the maximum length of CI cables, so the distance between nodes includes two lengths of CI cable and the Star Coupler.

⁶Can be greater depending on configuration.

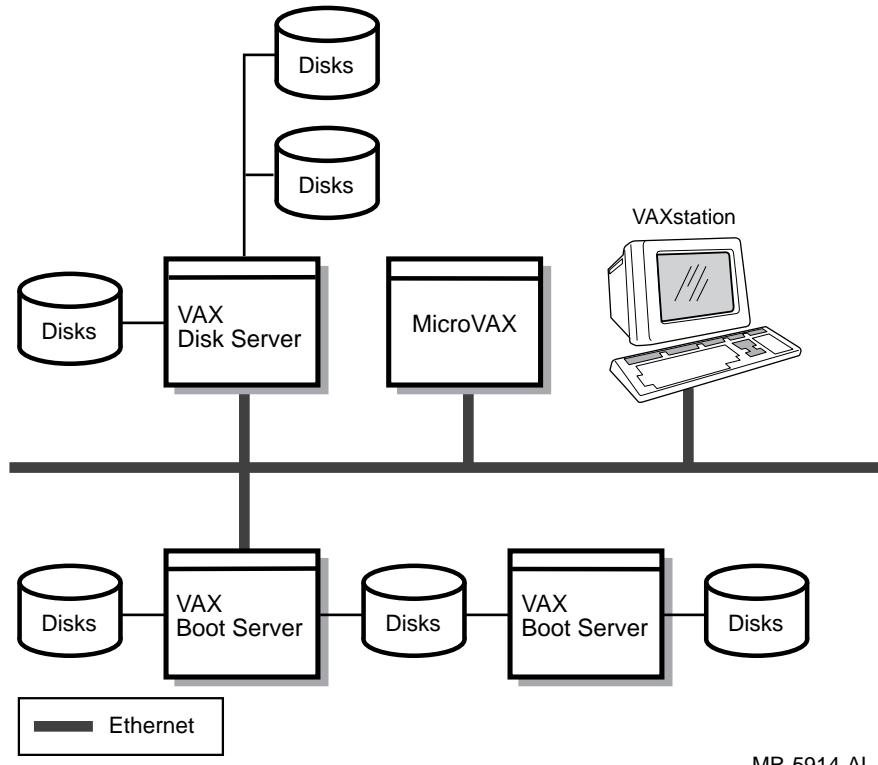
⁷This figure is a limit associated with VAXcluster MDF configurations. The maximum limit is 100,000 meters.

Ethernet VAXcluster Configurations

A VAXcluster configuration can also use Ethernet for VAXcluster communications, as shown in Figure 1–1. A single Ethernet can support many VAXcluster systems based on utilization and bridging. Each Ethernet VAXcluster system can include up to 96 nodes. VMS SMP systems count as one node.

Ethernet is often used as an interconnect medium for several VAXcluster types, forming a mixed-interconnect VAXcluster system. For instance, if you have 50 nodes at your site, you may have several DSSI segments, several CI segments, and some FDDI links, as components of the larger Ethernet-based VAXcluster system. Mixed-interconnect VAXcluster systems are described in more detail in the section titled Mixed-Interconnect VAXcluster Configurations and Section 4.3.

Figure 1–1 Ethernet VAXcluster Configuration



MR-5914-AL

CI VAXcluster Configurations

A CI VAXcluster configuration uses the CI to connect VAX CPUs to other VAX systems, to disks, and to tapes. Up to 16 VMS systems can be attached to a CI¹. Figure 1–2 shows a VAXcluster system configured with the CI. Some storage devices are dual-ported to two controllers (HSC systems). This provides failover capability and enhances data availability. When nodes are added to a CI VAXcluster, all systems remain on line and functional as the new member joins the VAXcluster.

If you want to add MicroVAX CPUs or VAXstations to a VAXcluster system that uses the CI bus, you can add other types of interconnects, such as Ethernet or DSSI, for VAXcluster communication.

DSSI VAXcluster Configurations

A DSSI VAXcluster configuration uses DSSI to connect up to four VAX CPUs to disks and tapes. It also requires connections for DECnet communications between VAX CPUs. Typically, this communication occurs over an Ethernet connection. Figure 1–3 shows a typical DSSI VAXcluster configuration. Note that all storage components connected to the DSSI are Integrated Storage Elements (ISEs) that include an internal controller for each disk or tape drive.

¹ These figures do not include MicroVAX or VAXstation systems, which do not participate as nodes in CI VAXcluster systems.

Figure 1–2 Typical VAXcluster Configuration with CI

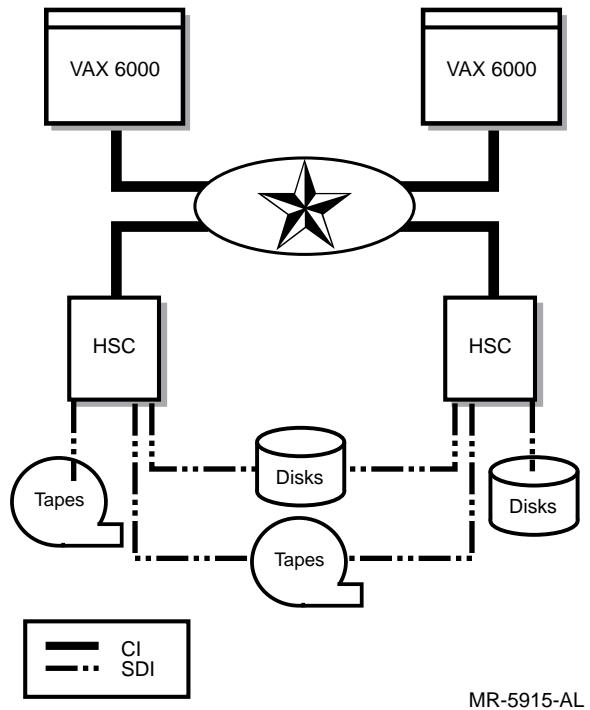


Figure 1–3 DSSI VAXcluster Configuration

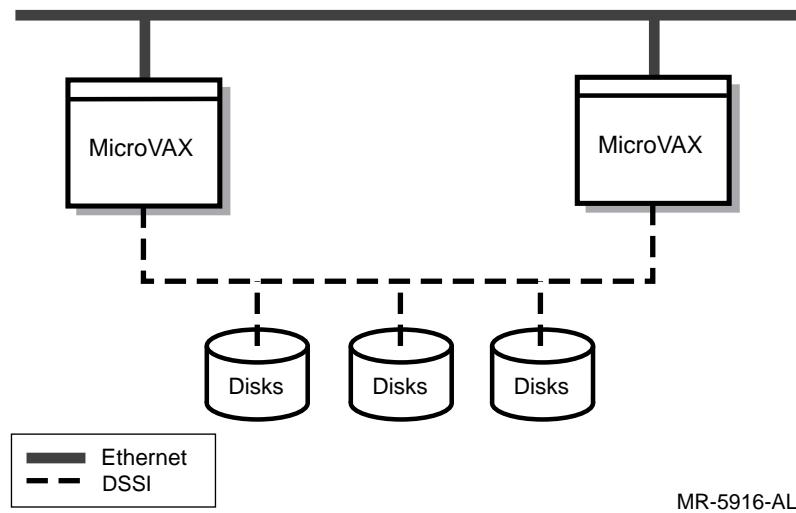
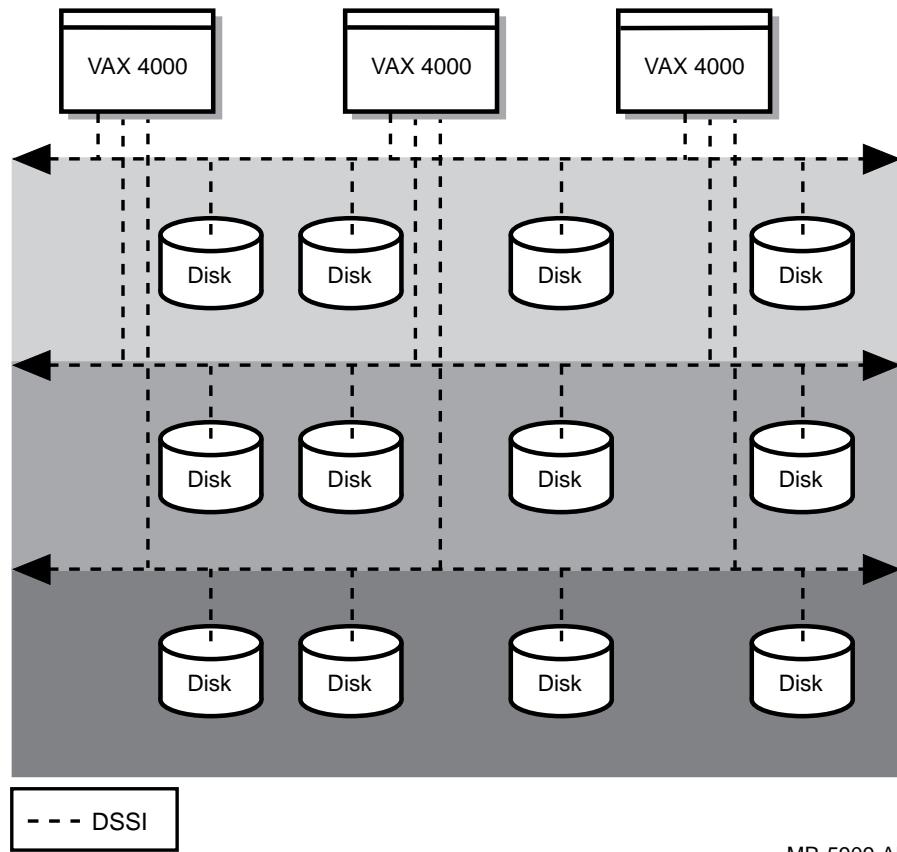


Figure 1–4 shows a DSSI configuration with three host CPUs.

Figure 1–4 Multiple DSSI VAXcluster Configuration



MR-5909-AL

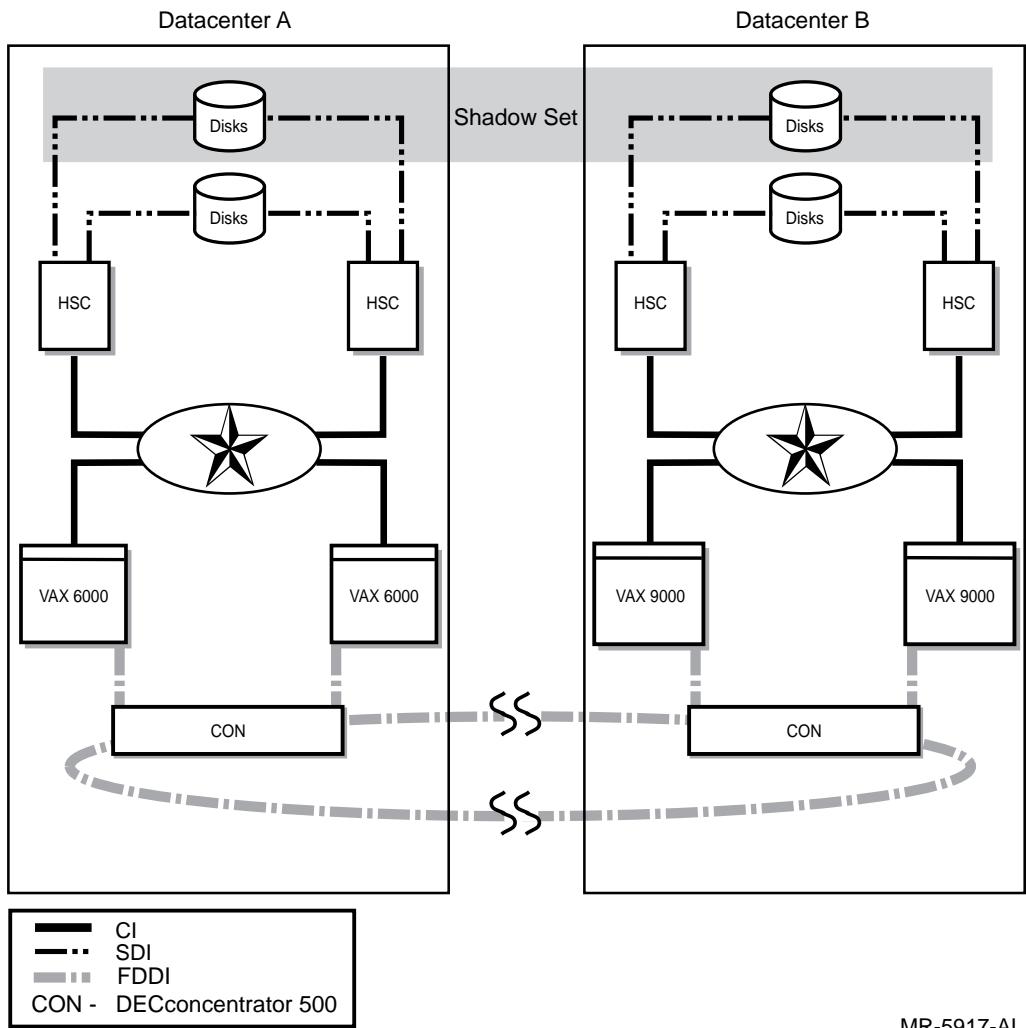
FDDI VAXcluster Configurations

Using FDDI as an interconnect extends the capabilities of VAXcluster systems. FDDI can extend further than CI and DSSI and provide higher bandwidth than Ethernet or DSSI. With the distances supported by FDDI, high-end processors (for example, the VAX 6000 and VAX 9000) that do not reside in the same computer room can be connected in a VAXcluster system. Data is accessed and MSCP served by CPUs across the FDDI. Figure 1–5 is an example of an FDDI VAXcluster configuration.

FDDI is used as a VAXcluster interconnect for the following reasons:

- To effect disaster tolerance. This can be accomplished by the purchase of the VAXcluster Multi-Datacenter Facility (MDF) package.
- To consolidate datacenter management. This entails combining autonomous, physically separated groups of DSSI-connected, CI connected, or standalone CPUs into a single VAXcluster system.
- To create a local area network (LAN)-based VAXcluster or an extended local area VAXcluster system.

Figure 1–5 FDDI Multi-Datacenter LAN-based VAXcluster



MR-5917-AL

Mixed-Interconnect VAXcluster Configurations

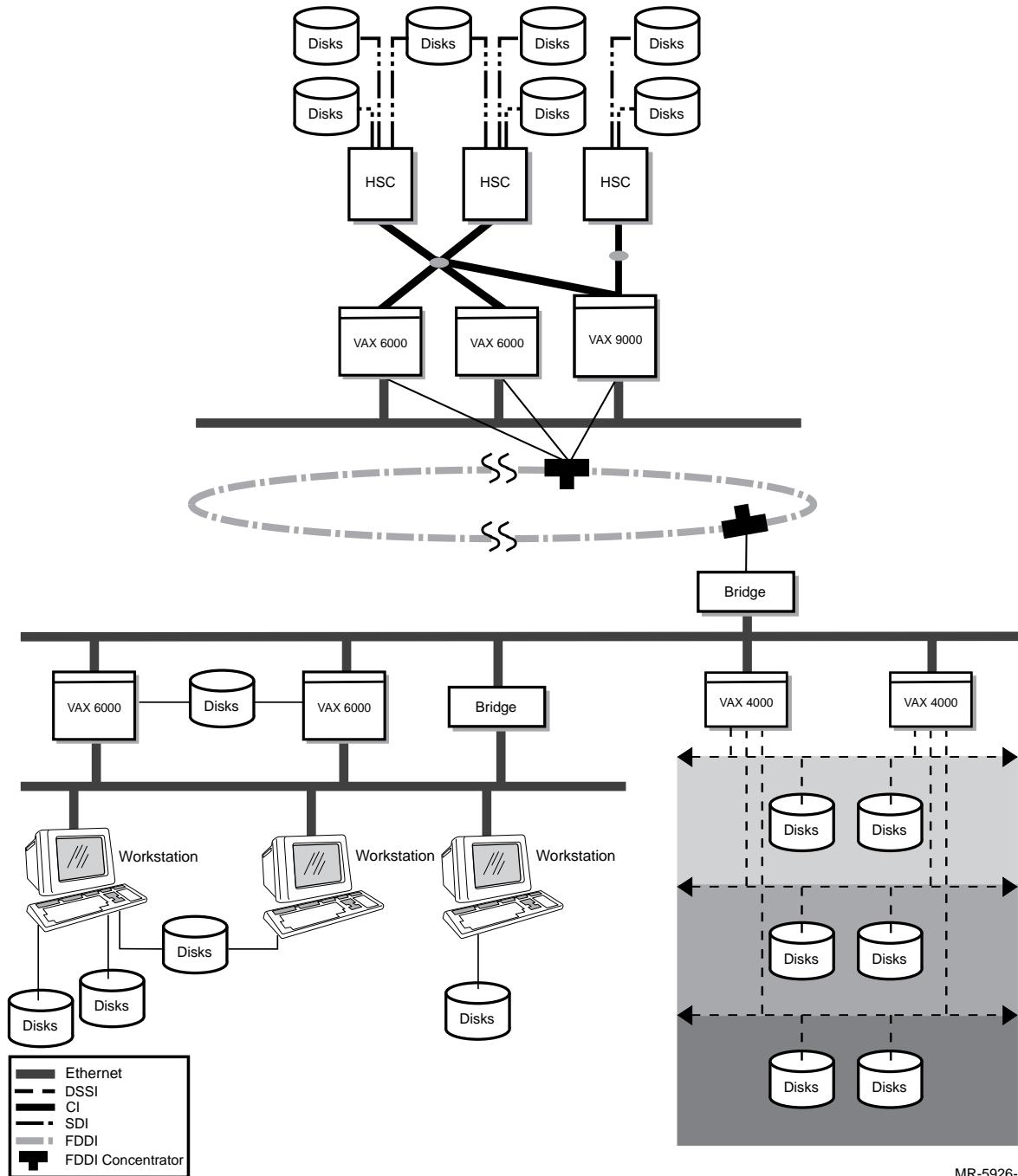
A VAXcluster configuration with mixed interconnects can use the CI, Ethernet, FDDI, and DSSI for VAXcluster communications. Configurations with mixed interconnects can contain up to 96 VMS nodes. A mixed-interconnect VAXcluster system can contain multiple CI segments (each with up to 16 nodes) and DSSI segments (each with up to three nodes). In this way, you can combine the following advantages of VAXcluster configurations that use the CI, FDDI, Ethernet, and DSSI as interconnects:

- Use of HSC subsystems for mass storage, so that satellites can access the large amounts of storage available through HSC subsystems
- Support for the full range of VAX systems, from workstations to mainframes
- High availability of system resources in appropriate configurations
- Use of graphics workstations in conjunction with large databases

Typically, one or more of the powerful CPUs are configured as MOP and disk servers to enhance resource availability and to balance the work load.

Figure 1–6 shows a VAXcluster configuration with multiple interconnects. Note that it combines configurations similar to those previously represented as separate interconnect configurations into one larger VAXcluster system.

Figure 1–6 Large VAXcluster Configuration with Multiple Interconnects



MR-5926-AL

Establishing Your VAXcluster Requirements

The process of configuring a VAXcluster system entails careful planning. You must define the characteristics of your computing environment and determine your application and system requirements. This chapter provides guidelines to help you determine those requirements. Topics include the following:

- Defining your computing environment
- Determining your application requirements
- Determining your overall system requirements

2.1 Defining Your Computing Environment

The first step in planning your VAXcluster system is to define your computing environment. You can have an *interactive* (timesharing) environment, where a number of active users are simultaneously accessing a system, a *batch* environment, where users expect a system to perform a job with no human intervention other than to start the job and use the final results, or a *client/server* environment, where distributed applications are split between a shared server and many clients.

In a general *timesharing environment*, such as program development, document preparation, or office automation, workload types tend to be an even balance of central processing unit (CPU) and input/output (I/O) use.

Applications can be *compute-intensive* (for example, when using the computer for simulation, modeling, or calculation). Some batch environments are compute-intensive. Client-server environments can be *I/O-intensive* (for example, when using the VAXcluster system for transferring and tracking data). Transaction processing (TP) systems and applications that manipulate large databases tend to be I/O intensive.

If you are configuring a VAXcluster system for scientific calculations, computer aided design/manufacturing (CAD/CAM) applications, or any environment likely to include VAXstation CPUs, your environment is probably compute-intensive. For this type of environment, you may determine that CPUs in your VAXcluster system need the maximum amount of configurable memory or that multiple high-powered CPUs are appropriate.

2.2 Considering Your Application Requirements

I/O-intensive environments typically run TP applications, such as financial or reservation systems. Such applications extract information from a shared database, display reports, and update the database. Different applications generate different I/O loads.

Once you define the general characteristics of your computing environment, determine your application requirements in the following key areas:

- Memory
- Compute power
- I/O throughput

The amount of memory is vital, since it must permit storage of all the information your applications require, plus additional storage for normal CPU functions.

Computer power must be proportional to how much calculation your applications perform, with enough additional CPU power to oversee data transfer between CPUs, and between CPUs and storage. You will find more information on establishing your memory and compute power requirements in Chapter 3.

I/O throughput is governed by adapters that feed information to interconnect data buses and by the inherent speed of the interconnect medium. Chapter 4 will help you establish your I/O requirements.

These three requirements are vital to the design, and their interdependencies require careful design consideration. A VAXcluster design expert can help you comply with budgetary and space restrictions while meeting your immediate needs and enabling the VAXcluster system to grow with any application expansion in the future.

For information on storage requirements for Digital system products, see Appendix A or the most current Software Product Descriptions (SPDs). See the *Guide to Software Handbook* for non-Digital software. This information can help you establish storage requirements for specific applications. To determine overall storage requirements, follow the guidelines in Chapter 5, which contains detailed information on storage devices and controllers.

2.3 Determining Your Overall System Requirements

After determining your application requirements, establish your overall system requirements for:

- System availability
- Disk availability and data redundancy
- Printer availability and redundancy
- System security
- Cost and projected growth
- Available physical space (computer room or office)

These requirements are discussed briefly in the following sections.

System Availability

VAXcluster configurations provide a range of enhanced availability. For example, a configuration that includes three or more CPUs (or two CPUs and a quorum disk) can continue to function after shutdown of one or more CPUs, depending on the configuration.

Properly configured VAXcluster systems can withstand the failure of various components.

Disk Availability and Data Redundancy

The VMS operating system supports multiple access paths to disks and the failover of disks between pairs of HSC subsystems, between local controllers, and between disk servers. Failover can occur when an access path fails and an alternative path is available. When the path breaks, the device using that path automatically fails over to another path. Failover of disk drives between HSC subsystems, local controllers (UDA50, KDA50, KDB50, KDM70), or disk servers helps to provide high data availability. When a disk server fails in configurations that include multiple servers, satellite CPU access to disks resumes through another server.

VMS Volume Shadowing provides redundancy for critical data and for system and application software. VMS Volume Shadowing provides transparent data access in the presence of disk, media, controller and communication failures. It can also decrease overall read access time because it maintains the same data on more than one disk drive. For more information on Volume Shadowing, see Section 5.3.5.6 and the *VMS Volume Shadowing Manual*.

Printer Availability and Redundancy

Many VAXcluster systems use multiple printers for redundancy. To optimize printer resources, connect printers to the network and set up clusterwide generic print queues, as described in the *VMS VAXcluster Manual*.

System Security

Multiple CPUs and multiple symmetrical processors in a single CPU are always single security domains.

Ethernet VAXcluster configurations can include VAXstation CPUs that are located outside secure areas. For truly stringent security precautions, consider using Digital Ethernet Secure Network Controller (DESNC) devices to connect these VAXstation CPUs to the Ethernet. These devices encrypt packets traveling across the Ethernet. You may want to locate all CPUs in secure areas, which is usually the case for CI or DSSI VAXcluster systems.

Cost and Projected Growth

In preparing your VAXcluster system budget, evaluate both current and future computing needs. Choose CPUs and configurations that can grow in tandem with your needs. VAXcluster systems offer excellent scalability — the equipment you purchase now can continue to be used as more resources, CPU power, and storage are added. Remember to consider the cost of ownership and management over time.

Capacity planning and management can begin before you have a VAXcluster system. Capacity planning products are available to help predict system performance and size configurations. For more information on these tools, see Section 7.7.5.

Available Physical Space

If physical space is limited by either availability or high cost, select more powerful CPUs or CPUs with a smaller footprint (CPUs that take up fewer square feet of floor space).

Choosing Your VAX CPUs

This chapter provides information to help you select CPUs for your VAXcluster system to satisfy your application requirements. This chapter covers the following topics:

- CPU selection criteria and guidelines
- Characteristics of VAX CPUs

To establish a common terminology, this chapter, as well as the rest of this book, uses the term *processor* to refer to an electronic unit that acts upon instructions to compute a result. The term *CPU* refers to the whole cabinet, which includes one or more processors, memory, and input/output (I/O) adapters and acts as a central controlling body. For example, a VAX 6000-530 CPU includes three processors.

3.1 CPU Selection Guidelines

VAX systems span a range of computing environments from desktop workstations (the VAXstation family) to entry-level departmental systems (the VAX 4000 family), to general-purpose datacenter systems (the VAX 6000 family), and mainframes (the VAX 9000 family). Furthermore, the VMS environment offers a wide range of alternative ways to grow and expand the processing capabilities of the datacenter. Many VAX CPUs can be expanded to include additional memory, processors, vector processors, or I/O subsystems.

Moreover, VMS VAXcluster Software offers many ways to interconnect VAX CPUs into integrated VAXcluster configurations. Major benefits of VAXcluster configurations include resource sharing, increased overall system availability, and the use of the combined compute power of multiple CPUs.

The following portions of this section discuss application requirements and the different types of VAX CPUs you can select.

3.1.1 Determining Application Requirements for CPUs

To establish your application requirements, you must evaluate the work your systems will perform. This includes understanding the following aspects of your application needs:

- Computing style
- Availability
- Growth
- I/O requirements

3.1.1.1 Computing Style

Determine what type of computing style your environment requires. Users share one or more primary systems in a general timesharing environment. In a client-server environment, users each have smaller systems that are networked together.

Users who want their jobs to run non-interactively submit their jobs to a batch queue. In contrast, users who directly and interactively use CPU power are said to work in a timesharing environment.

3.1.1.2 Availability

Your application will have a requirement for system availability. An application that is critical to business may benefit by being distributed over several VAX CPUs in a VAXcluster system. Then, if a CPU fails, the remaining members of the VAXcluster system are still available, and continue to access the disks, tapes, printers, and other peripheral devices that they need. Fiber Distributed Data Interface (FDDI) can be used for improving disaster tolerance and increasing availability with the VAXcluster Multi-Datacenter Facility (MDF). See *Building Dependable Systems: The VMS Approach* for complete guidelines for application availability. See the *VAXcluster Multi-Datacenter Facility: Configuration Guidelines* manual for more information on MDF.

3.1.1.3 Growth

When assessing the size of a new system, consider the applications you will be running, the users' needs, and how large a storage subsystem you require. Design with leeway for growth. CPU power, memory, I/O, and interconnect bandwidth can limit or reduce your VAXcluster system's performance if you design too close to your needs. Consider the cost of upgrading the system if you do not accurately gauge how your needs may expand. To ensure that unforeseen requirements are met, design with some slack in mind to accommodate these requirements.

Expanding CPU Power

For applications that experience rapid growth in their computing requirements, a number of different growth paths are available. Many existing systems can grow by adding another processor to an existing CPU.

However, when you approach the limits of a single computer system, even after upgrading to the most powerful CPU, VAXcluster systems offer the ideal next step. VAXcluster systems are effective when applications are *distributed* across multiple CPUs, so that the pieces run somewhat independently, but are still able to share access to stored data. Alternately, VAXcluster systems can be useful with multiple applications that share the same data.

Scaling I/O Performance

Another consideration in system expansion is the performance of the I/O subsystem. Spreading I/Os over disks and using disk striping software can maintain or improve performance levels as I/O rates increase.

3.1.1.4 I/O Requirements

An application places demands on the throughput (I/O request rate per second) and bandwidth (bytes of data transferred per second) of the I/O subsystem. A closer look at the I/O requirements reveals that you should consider two factors:

- How fast data is read or written
- How fast data is processed

Limits on read/write activity depend on the throughput capacity of the storage subsystem. Latency, the measurement of how long an application must wait for the data, must be kept low. Latency can be reduced by ensuring the I/O subsystem (interconnect and adapters) meets the needs of the application.

3.1.2 Selecting Your VAX CPU Configuration

Having established your applications requirements, you are now ready to select your CPUs. These next sections help you consider:

- VAXcluster Systems and Fault-Tolerant CPUs
- VAXcluster Multi-Datacenter Facility Systems

3.1.2.1 VAXcluster System and Fault-Tolerant CPUs

For applications that require high availability of services, the choice is between a VAXcluster system, a fault-tolerant VAX CPU, or a combination of both.

The major distinction, in terms of availability of services and data access, between VAXcluster systems and fault-tolerant processors is that failure of a CPU is not fully transparent to an application running in a VAXcluster environment. When a failure occurs, an application must reconnect to the VAXcluster server or establish a new session on another member of the VAXcluster system. In either case, there is a brief pause in service, in the range of seconds in duration.

With a fault-tolerant CPU, a failure of any system component, including a processor, memory, or peripheral is absolutely transparent to the application. Because fault tolerance is implemented in VAXft hardware, an application continues to run with no interruption of services.

3.1.2.2 VAXcluster Multi-Datacenter Facility Systems

The VAXcluster MDF provides the software, services, training, and license to support automatic failover and predictable recovery in the event of a disaster. VAXcluster MDF supports the capability of developing disaster-tolerant platforms and datacenter consolidation through managing multiple sites from multiple locations. Using specially designed Operations Management Station (OMS) software, two separate datacenters or VAXcluster systems connected with FDDI can be consolidated into one single VAXcluster that can be managed from either site.

FDDI can be used in conjunction with Ethernet to tie together CPUs at different geographic locations that cannot connect directly to FDDI. These groups of CPUs are connected to FDDI via an FDDI-Ethernet bridge. This way, CPUs that cannot connect directly to FDDI can participate in an MDF VAXcluster system.

3.1.3 Determining the Number of CPUs

One of the main advantages of a VAXcluster system is improved availability through redundancy of key components. If one VAX CPU in a VAXcluster system fails, the VAXcluster system continues to operate with the remaining CPUs if the VAXcluster system retains quorum. In a VAXcluster system with two identical VAX CPUs, loss of either VAX CPU leaves 50 percent of the original CPU power available. In a VAXcluster system with three identical VAX CPUs, loss of one VAX CPU leaves 66 percent of the original CPU power, and so on. More information is available in *Building Dependable Systems: The VMS Approach*.

You can also scale your VAXcluster system to protect your investment. Starting with the configuration your work requires today, your equipment can be reconfigured into a larger VAXcluster system tomorrow, or into multiple VAXcluster systems. The vast range of CPUs, from high-end symmetrical multiprocessing (SMP) systems to smaller workstations, can interconnect and be reconfigured easily to meet growing needs.

3.1.4 Determining Memory Requirements

CPUs in VAXcluster systems require 0.5 to 1.5 megabytes more memory in each CPU than in standalone systems. This additional memory is used to support the resource base in the VAXcluster system, which is larger than that of a standalone system. With added memory, a CPU in a VAXcluster system can generally support the same number of users or applications it supported as a standalone system. Some of this memory is devoted to VAXcluster functions and coordination of VAXcluster resources, but this memory may also be used when data being served by a CPU must be buffered enroute to the requesting CPU.

As a VAXcluster system configuration grows, there may also be modest increases in the amount of memory used for system work by each CPU. Because the per-CPU increase depends on the level of data sharing in the VAXcluster system and the distribution of resource management, that increase is not subject to fixed rules. If the CPU is a resource manager for a heavily utilized resource, additional memory may provide increased performance for VAXcluster users of that resource.

For more information on using additional memory to improve performance, refer to the *Guide to VMS Performance Management* in the VMS documentation set.

3.2 VAX CPU Characteristics and Positioning

Table 3-1 lists VAX CPUs that can be included in a VAXcluster configuration. It also indicates which of these CPUs can be attached to CI, DSSI, and LAN interconnects and what adapters can be used in each case. The families of VAX CPUs, as well as the individual VAX systems, are compared by their performance and memory capacity. For more details on these products, refer to the *VAX Systems/DECsystems Systems and Options Catalog*. Table 3-2 lists packaging information for all VAXcluster CPUs.

Note

Performance is highly dependent on configuration, application, and operating environment. You must carefully evaluate individual work loads to make reasonable performance estimates for specific applications. In the tables in this document, no guarantee of system performance is expressed or implied.

Table 3-3 lists the I/O characteristics of VAX CPUs and template systems, with typical disk capacities and raw I/O bandwidth. Actual I/O performance can never exceed the raw bandwidth of a particular I/O channel and is generally less. Table 3-4 lists the interconnect options for VAXcluster CPUs. Table 3-5 lists storage bus options for VAXcluster CPUs.

For information about mature CPUs no longer shipped by Digital, see Tables C-1 and C-2.

The processing power of a single VAX 11/780 CPU is called a VAX Unit of Processing (VUP). Where available, CPU processing power is also listed in SPECmarks. SPEC is the System Performance Evaluation Cooperative, a multivendor committee that defined a rigid set of benchmarks and optimization environments.

Table 3–1 VAX System CPU Performance Characteristics

CPU Name	VUPs	SPECmarks	Maximum Memory ¹
VAXstation CPUs			
3100–30/38	5/3.8	6.1/–	32
3100–40/48	5/3.8	6.1/–	32
3100–76	7.6	6.8	32
3100–80	10	10.5	64
3100–90	24	–	128
3200/3500	2.7	–	16/32
Entry-Level VAX CPUs			
3300/3400	2.4	–	52
3500/3600	2.7	–	64
3800/3900	3.8	–	64
4000–100	24	–	128
4000–200	5.0	5.6	64
4000–300	8.0	9.1	256
4000–400	16	–	512
4000–500	24	30.5	512
4000–600	32	40.9	512
Fault-Tolerant VAX CPUs			
VAXft Model 110	2.4	–	96
VAXft Model 310	3.8	–	128
VAXft Model 410	6.0	–	256
VAXft Model 610	6.0	–	256
VAXft Model 612	12	–	256
VAXft Model 810	35	–	256
Mid-Range and High-End VAX CPUs			
6000–210/220 ²	2.8/5.5	–/5.6	512
6000–230/240 ²	5.5/11.0	–	512
6000–310/320 ²	3.8/7.5	–/9.2	512

¹In megabytes.

²Some VAX 6000 and VAX 9000 family CPUs can host vector coprocessors.

(continued on next page)

Table 3–1 (Cont.) VAX System CPU Performance Characteristics

CPU Name	VUPs	SPECmarks	Maximum Memory ¹
Mid-Range and High-End VAX CPUs			
6000–330/340 ²	11.3/15.0	—	512
6000–350/360 ²	18.6/22.0	—	512
6000–410/420 ²	7/13	8.2/—	512
6000–430/440 ²	19/25	—	512
6000–450/460 ²	31/36	—	512
6000–510/520 ²	13/25	15.6/—	512
6000–530/540 ²	37/49	—	512
6000–550/560 ²	61/72	—	512
6000–610/620 ²	32/58	42.1/—	512
6000–630/640 ²	84/106	—	512
6000–650/660 ²	128/150	—	512
7000–110/120	37/73	51/—	3500
7000–130/140	108/144	—	3500
7000–610/620	35/65	—	3500
7000–630/640	95/126	—	3500
9000–110 ²	40	35	512
9000–210 ²	40	—	512
9000–3xx ²	40	35	512
9000–410 ²	40	—	512
9000–420 ²	79	—	512
9000–430 ²	118	—	512
9000–440 ²	157	—	512
10000–110/120	37/73	51/—	3500
10000–130/140	108/144	—	3500

¹In megabytes.²Some VAX 6000 and VAX 9000 family CPUs can host vector coprocessors.

Table 3–2 CPU Packaging Information

Package Style	CPU Types
Desktop	3100 family, 4000–100
Pedestal compact cabinet	3200/3500, 3520, 3540
Pedestal cabinet	3300/3400, 3500/3600, 3800/3900, 4000–200 /300/400/500/600
Rack mount	4000–200/300/400/500/600, VAXft model 110 /310/410
Two pedestal cabinets	VAXft model 110/310/410
Compact cabinet	VAXft model 610/612, VAX 6000 family
Three cabinets plus optional UPC ¹	VAX 9000 family

¹Utility Port Conditioner. Used by customers whose power does not meet the required standards.

Note

Performance data is application dependent.

Table 3–3 VAX System I/O Performance Characteristics

CPU Name	Internal Bus	
	Type	Throughput ¹
VAXstations		
3100	Integral	4.0
3200/3500	Integral	3.3
3520/3540	Integral	3.3
Entry-Level VAX CPUs		
3300/3400	Q-bus	3.3
3500/3600	Q-bus	3.3
3800/3900	Q-bus	3.3
VAX 4000 family	Integral Q-bus	3.3
Fault-Tolerant VAX CPUs		
VAXft 110	Integral	3.3
VAXft 310	Integral	3.3
VAXft 410	Integral	3.3
VAXft 610	Integral	80
VAXft 612	Integral	80

¹Maximum bus throughput in megabytes per second.

(continued on next page)

Table 3–3 (Cont.) VAX System I/O Performance Characteristics

CPU Name	Internal Bus	
	Type	Throughput ¹
Mid-Range and High-End VAX CPUs		
6000 family	XMI	80
9000–110	XMI	80
9000–210	XMI	80
9000–310/320	XMI	80/160
9000–330/340	2 XMI	160/320
9000–410/420	2 XMI	160
9000–430/440	4 XMI	320

¹Maximum bus throughput in megabytes per second.

Table 3–4 CPU Interconnect Options

Interconnect	CPUs
Ethernet	All
CI	VAX 6000 family, VAX 9000 family
DSSI	3300/3400, 3500/3600, 3800/3900, VAX 4000 family, VAXft 610, VAXft 612, VAX 6000 family, VAX 9000 family
FDDI	VAX 6000 family, VAX 9000 family

Table 3–5 CPU Storage Bus Options

Storage Bus	CPUs
SCSI ¹	3100, 3200/3500, 3520/3540
DSSI ²	VAXft 110/310/410/610/612, VAX 4000 family, VAX 6000 family, VAX 9000 family
SDI ³	3300/3400, 3500/3600, 3800/3900, 4000 series, VAX 6000 family, VAX 9000 family
STI ⁴	VAX 6000 family, VAX 9000 family

¹This bus can accommodate RRD-series, RZ-series, TK-series, TS-series, TU-series, and TZ-series storage devices.

²This bus can accommodate RF-series, SF-series, and TF-series storage devices.

³This bus can accommodate ESE20, RA-series, and SA-series storage devices.

⁴With a KDM70, this bus can connect to TA-series and TU-series storage devices.

Choosing Your VAXcluster Interconnect

An interconnect is a communication path used to exchange information among two or more nodes. These nodes can be VMS nodes or part of a storage subsystem. Communications within a VAXcluster system are handled by the system communication services (SCS) software. Messages between nodes travel over the interconnect(s) between those nodes.

In Chapter 3, you determined compute requirements for your VAXcluster system. This chapter gives you information about the different interconnects available to connect processors to each other and to storage subsystems.

4.1 Interconnect Characteristics

The key characteristics of an interconnect are its throughput (its capacity to carry data), the distance it can extend, and the number of nodes that can attach directly to it. Those features are summarized for VAXcluster interconnects in Table 1–1.

Use of interconnect is affected by:

- throughput
- CPU overhead

4.1.1 Throughput

The physical line rate is the maximum theoretical throughput of an interconnect. Every interconnect has to fulfill overhead requirements that make the maximum data carrying capacity (effective throughput) less than the physical line rate.

There are also factors that affect the effective throughput of an adapter. Larger data transfer sizes allow higher data rates, while smaller data transfer sizes produce a lower data rate. This is because overhead represents a small percentage of the total work when performing data transmission of larger messages.

4.1.2 CPU Overhead

The amount of central processing unit (CPU) overhead used by an interconnect depends on whether certain functions are handled internally by the adapter or by software. CI and Digital Storage Systems Interconnect (DSSI) adapters implement many functions in hardware. Therefore, less processing power is required to send messages over those interconnects. Ethernet and Fiber Distributed Data Interface (FDDI) adapters implement some of those same functions in software; so more processing power is required to send messages over these buses.

4.2 Interconnect Types

Interconnects are also distinguished by what types of processors and storage devices can connect to them. An adapter that connects the internal system bus to the interconnect is required for each processor or storage subsystem. Table 4–1 highlights the kind of connections each interconnect can make.

Table 4–1 Interconnect Connections

Type of Connections	Ethernet	CI	DSSI	FDDI
Connections for large CPUs ¹	Yes	Yes	Yes	Yes
Connections for medium CPUs ²	Yes	No	Yes	No
Connections for small CPUs ³	Yes	No	Yes	No
Connections for storage	No ⁴	Yes	Yes	No

¹Large CPUs include the VAX 6000 and VAX 9000 families.

²Medium CPUs include the VAX 4000 family.

³Small CPUs include VAXstation systems.

⁴InfoServer 100 can be attached to the Ethernet.

With these connections, each interconnect has certain attributes.

4.2.1 Ethernet Interconnect

The Ethernet bus is a 10-megabit per second interconnect that links VAX and MicroVAX CPUs in VAXcluster systems. The Ethernet interconnect:

- Supports all types of CPUs in a VAXcluster system.
- Allows some VMS systems to be downline loaded by other VMS systems.
- Supports multiple VAXcluster systems and up to 96 VMS nodes in a single VAXcluster system.
- Provides extended physical distribution of nodes. Nodes connected by an Ethernet can be anywhere on an extended local area network (LAN) and be kilometers apart.
- Costs less than any other interconnect.

The Ethernet carries LAT (terminal) traffic and DECnet traffic between VAX CPUs. This traffic can reduce the throughput available for VAXcluster communication. As a rough estimate, each active terminal on the LAT can use about 1 kilobyte per second of Ethernet throughput. PrintServer and serial printers connected to the Ethernet can also use a portion of its throughput. For all these reasons, the Ethernet can become an input/output (I/O) bottleneck in a VAXcluster system.

VMS supports multiple Ethernet adapters on some VMS systems. Using multiple Ethernet adapters and multiple Ethernet segments can help alleviate congestion and increase the throughput of VAXcluster traffic. Therefore, consider the capacity of the total network design when configuring a VAXcluster system with many Ethernet-connected nodes or when the Ethernet also supports a large number of terminals or printers.

Table 4–2 shows Ethernet adapters and the VAX CPUs to which they connect. Table C–4 lists information for adapters which Digital no longer ships.

Table 4–2 Ethernet Adapters

Device	Internal Bus	VAX CPU
DEBNA	VAXBI	88xx, 8700, 85xx, 8350, 8250, 6000 family, VAXstation 8000
DEBNI	VAXBI	VAX 8xxx, VAX 6000 family
DELQA	Q–bus	VAXstation 3xx0, VAXstation II, VAXstation II/GPX, MicroVAX II, MicroVAX 3500/3600, MicroVAX 3800 /3900
DEMNA	XMI	VAX 9000 family, VAX 6000 family
DESQA	Q–bus	MicroVAX 3500/3600, MicroVAX 3800/3900, VAX 4000 family
DESVA	CPU-module	MicroVAX 2000/3100, VAXft systems
SGEC	CPU-module	VAX 4000 family
DEQTA	Q–bus	VAXstation 3xx0, VAXstation II, VAXstation II/GPX, MicroVAX II, MicroVAX 3500/3600, MicroVAX 3800 /3900

4.2.2 CI Interconnect

The CI consists of several components:

- Star Coupler
- Optional Star Coupler Expander (CISCE)
- CI adapters
- CI cables

Star Coupler and Star Coupler Expander

The Star Coupler and CISCE provide a common connection point for signals from all VMS nodes and HSC subsystems to which they are connected by CI cables. The Star Coupler is a passive device; the CISCE consists of two amplifiers, one for each of its two paths. The Star Coupler lets you have up to 16 CPUs or HSC subsystems connected in a single CI based VAXcluster. The CISCE lets you add up to 16 more nodes, for a total of 32 nodes (16 of which can be VMS nodes).

CI Interconnect

The CI bus provides high-speed, highly available, multiple access path connections between all CI based nodes in a VAXcluster system, usually within a single computer room. The CI has two independent data paths for redundancy.

Each path uses a transmit cable and a receive cable. Both paths are used to handle traffic. If one path to a node is unavailable, traffic uses the remaining path. The failed path is periodically tested and automatically used for traffic as soon as the path becomes available again.

The primary advantages of the CI are as follows:

- Allows high-speed communication paths for larger processors and I/O-intensive applications.

- Permits efficient access to large amounts of storage. HSC subsystems can connect large numbers of disk and tape drives to the VAXcluster system, with direct access from all VAX nodes on a CI.
- Has minimal CPU overhead for communication. CI adapters are intelligent interfaces that perform much of the work required for communication among VMS nodes and storage. The VAXcluster topology allows all VMS nodes attached to a CI bus to communicate directly with the HSC subsystems on the same CI bus.
- Provides high availability through redundant data paths. Each CI adapter is connected with two pairs of CI cables. The loss of a single cable connection does not break the link to other VAXcluster nodes.

The effective throughput of the CI bus is high, and a single CI bus is not likely to be a bottleneck in a large VAXcluster configuration. If a single CI is not sufficient, multiple CI buses can be used. Section 4.4.1 provides more information on using multiple CI buses.

CI Adapters

Table 4–3 shows the CI adapters and the maximum data rate measured under various test conditions for each one. The higher values were obtained in tests that performed large I/O transfers (for example, 20 to 40 blocks per I/O request). Total data rate is lower with smaller transfers (for example, lock management traffic). Table C–5 lists adapters that Digital no longer ships.

The speeds of the CPUs supported by these adapters can range from roughly 1 VAX Unit of Processing (VUP) (the equivalent processing power of a VAX-11/780) to over 100 VUPs (VAX 9000-4x0). The ratio of adapter speed to CPU speed is lowest for fast processors.

Table 4–3 CI Adapters

Device	Internal Bus	Data Rate ¹	VAX CPU
CIBCA-B	VAXBI	0.54 to 2.6	88x0, 8700, 85x0, 8350, 8250, 6000-xxx, VAXstation 8000
CIXCD	XMI	2.1 to 8.0+	6000-xxx, 9000-xxx

¹Megabytes per second. Data rate capacity varies with speed of CPU and message size.

DECnet cannot run over the CIXCD. Therefore, nodes communicating over CIXCDs must have another connection, for example, Ethernet, to allow DECnet communications.

CI Connected Storage

CI connected configurations have storage devices accessed by VMS nodes through the HSC controller, Star Coupler, and CI bus. These configurations provide direct multihost disk and tape access for all CPUs attached to the CI. They can also allow use of a wide range of storage devices, online maintenance, and backup capability. Configurations based on the CI as the interconnect use the HSC controller, which offers several features for availability and performance. All disks connected to an HSC may participate as shadow set members in VMS Volume Shadowing. HSC subsystems can only connect to the CI.

4.2.3 Digital Storage Systems Interconnect

DSSI is a daisy-chained multidrop bus that connects up to eight nodes, four of which can be VMS nodes. The DSSI is physically different from the CI, but logically they are similar. CPUs on the DSSI communicate directly with storage devices.

The important features of the bus are:

- Single 8-bit parallel multidrop data path with both byte parity and packet checksum for error detection
- Peak throughput of 4 megabytes per second
- Greater than 3.75 megabytes per second of usable bandwidth
- Maximum length of the entire bus is 27 meters in a computer room, and 20 meters in an office environment
- A maximum of eight nodes, four of which can be VMS nodes

DSSI is a high-reliability bus. DSSI storage often resides in the same cabinet as the CPUs, which can help maximize computer-room space. However, it is important to note that maintainability differs for these DSSI configurations; the whole system may need to be shut down for service, as opposed to configurations and interconnects with separately housed CPUs and storage devices.

Processors ranging from the MicroVAX II to the VAX 6000 family can be connected directly to DSSI. DSSI supports up to seven Integrated Storage Elements (ISEs) daisy-chained through a single cable to an adapter in the host.

For more information on DSSI, see the *DSSI VAXcluster Installation and Troubleshooting Manual*.

DSSI Adapters

There are two different types of DSSI adapters — embedded and optional. The embedded adapter is part of the CPU. An optional adapter can be purchased separately to add to the system. Table 4–4 lists the available DSSI adapters and identifies the CPUs or system buses to which they can attach.

Table 4–4 DSSI Adapters for VAXcluster Systems

Device	Internal Bus	Throughput ¹	Data Rate ²	VAX CPU
EDA640	Embedded	360	1.5	MicroVAX 3300/3400
SHAC	Embedded	800	2.8	VAX 4000 family
SWIFT	Embedded	360	1.5	VAXft family
KFMSA	XMI	1600 ³	5.6 ⁴	VAX 6000 and 9000 families
KFQSA ⁵	Q-bus	190	1.5	MicroVAX 3300/3400 /3800, VAX 4000 family, MicroVAX II

¹I/Os per second.

²Megabytes per second.

³For both DSSI buses, 800 I/Os per second on each DSSI.

⁴For both DSSI buses, 2.6 megabytes per second for each DSSI.

⁵Can only be used for storage access, not VAXcluster communications.

DSSI-Connected Storage

DSSI configurations use ISEs connected to a DSSI bus. Each ISE contains a disk and a disk controller or a tape and tape controller (this removes the need for an HSC-like device separate from the storage unit). Each disk controller has a dedicated cache that can dramatically speed data reads. DSSI configurations provide multihost disk and tape access, flexibility, high reliability, and excellent price/performance. Multiple DSSI buses are supported for some CPUs. VMS Volume Shadowing supports shadowing on ISEs in a DSSI configuration.

4.2.4 FDDI Interconnect

FDDI is an ANSI standard 100-megabit per second interconnect that uses fiber-optic cable. FDDI supports VAXcluster functionality over greater distances than other interconnects. FDDI also augments the Ethernet by providing a high-speed interconnect for multiple Ethernet segments in a single VAXcluster system.

The FDDI standards define the following two types of nodes:

- Stations — The ANSI standard single attachment station (SAS) can be used as an interconnect to the FDDI ring. Digital recommends that stations be attached to concentrators, and concentrators be attached to the FDDI ring, making the ring more stable. The DEMFA (listed in Table 4–5) connects the XMI internal bus to the FDDI.
- Wiring concentrator — The wiring concentrator (CON) provides a connection for multiple SASs or CONs to the FDDI ring. A DECconcentrator 500 is an example of this device.

FDDI limits the total fiber path to 200 kilometers (125 miles). The maximum distance between adjacent devices in an FDDI VAXcluster is 40 kilometers with single mode fiber, but only 2 kilometers with multimode fiber. This maximum distance includes any devices that connect to the FDDI network, including connections to bridges and adapters.

FDDI supports transfer using large packets (up to 4468 bytes). Only FDDI nodes connected to the same ring can make use of large packets.

FDDI Adapter

Table 4–5 FDDI Adapter

Device	Internal Bus	Peak Throughput ¹	VAX CPU
DEMFA	XMI	12	9000 family 6000 family

¹Megabytes per second.

4.3 Mixed-Interconnect Configurations

VAXcluster systems can contain a combination of CI, DSSI, FDDI, and Ethernet interconnects. Mixed interconnects allow growth in VAXcluster systems. For example, an Ethernet VAXcluster system that requires more storage can expand with CI or DSSI connections.

When forming a new connection between two CPUs, the interconnect is selected by the VAXcluster Software in the following order of preference — CI or DSSI, then FDDI or Ethernet. In mixed-interconnect configurations, if a failure occurs on one interconnect, communications automatically move to another interconnect.

For more information on mixed-interconnect configurations, see the section titled Mixed-Interconnect VAXcluster Configurations.

4.4 Multiple Interconnects of the Same Type

VMS currently includes support for more than one CI, DSSI, FDDI, or Ethernet in a single VAXcluster system.

With more than one of the same interconnect in a VAXcluster system, you have higher availability, because VMS supports automatic failover of VAXcluster traffic between two interconnects. Depending on your configuration, performance improvements may be possible. Multiple paths from one CPU may transfer more information than a single path. Multiple interconnects enable load distribution and reduce the chance that an adapter could be bottleneck. Fault tolerance is also increased with multiple interconnects. For more information on using multiple interconnects, see Section 4.4.1.

4.4.1 Multiple CI Interconnects

A CI path between two nodes in a VAXcluster system consists of the adapters, cables, Star Coupler, and optionally the CISCE. Multiple CI adapters can be placed on some VMS nodes, and multiple Star Couplers and multiple CI cables can be used in the same VAXcluster system. Configuration rules and guidelines for using multiple CI interconnects are discussed in Section 6.2.1.

CI Load Sharing

With multiple CI adapters in a single CPU, the traffic load may be shared among the adapters. This avoids I/O bottlenecks and increases the total system I/O throughput.

4.4.2 Multiple DSSIs

You can use more than one DSSI adapter on some CPUs. Some CPUs can contain a mix of embedded and optional adapters.

The VMS operating system performs load sharing on systems configured with multiple DSSI adapters (excluding the KFQSA). The load-sharing algorithms are the same as those used for CI. DSSI load-sharing occurs when one VMS node can see another through multiple DSSI adapters.

4.4.3 Multiple Ethernets

Multiple Ethernet paths provide high availability and potentially increased performance. Ethernet segments can become congested and adapters can become loaded with traffic. Multiple Ethernet paths between VAXcluster nodes can alleviate congestion. If multiple Ethernet VAXcluster systems are configured according to the guidelines, server nodes can usually use some of the additional throughput provided by the added Ethernet adapters and increase the overall performance of the VAXcluster system. However, the performance increase depends on the configuration of the VAXcluster system and the applications it supports.

For more information on configuration rules and guidelines for multiple Ethernets, see Section 6.4.

Ethernet Segment Load Balancing

If only Ethernet paths are available, the choice between which path to use is based on latency. If delays are equal, either path can be used. Otherwise, the channel with the smallest latency (computed network delay) is chosen. The network delay across each segment is calculated every 3 seconds. Traffic is balanced across network segments, not adapters.

Ethernet Traffic Management

Two products are useful in managing congestion on the Ethernet: the LAN Bridge 200 and the LAN Traffic Monitor VMS. You can use the LAN Bridge 200 to limit VAXcluster traffic to a segment of the Ethernet in configurations where more than one VAXcluster system uses an extended Ethernet LAN. You can use the LAN Traffic Monitor in conjunction with a LAN Bridge 200 to monitor all traffic, of all protocol types, on an Ethernet segment and to monitor the load on the Ethernet and identify any serious bottlenecks.

The DECbridge 500 unit is a transparent FDDI-to-Ethernet translating bridge. It provides an interconnection between a 10-megabit per second Ethernet segment and a 100-megabit per second FDDI ring. DECbridge performs high-speed translation of network data packets between the FDDI and Ethernet frame formats.

4.4.4 Multiple FDDIs

Since FDDI is ideal for spanning great distances, you may want to supplement its high throughput with high availability, by ensuring that critical nodes are connected to multiple FDDI rings. Routing the FDDI links along separate physical paths helps ensure that the configuration is disaster tolerant. The VAXcluster Multi-Datacenter Facility (MDF) is installed using this and other guidelines that support disaster tolerance.

4.5 SCSI Bus

The Small Computer System Interface (SCSI) bus is not a VAXcluster interconnect, but, as a storage interconnect, it lets CPUs access storage. Based on an ANSI industry standard, a SCSI bus supports eight devices per bus, where one is the host. It allows a 1.5 megabyte asynchronous transfer rate and a 5.0 megabyte synchronous rate. It connects a limited number of CPUs to magnetic disks and tapes, printers, scanners, plotters, voice products, and manufacturing control devices. CPUs that can connect to SCSI devices include VAXstation 3520, 3540, and the 3100 family CPUs, and some VAX 4000 family CPUs. RZ-series, RRD-series, and TZ-series media can connect to the SCSI bus.

5

Designing Your Storage Subsystem

This chapter helps you design the storage subsystem of your VAXcluster system.

A distinguishing feature of VAXcluster systems is that they enable multiple central processing units (CPUs) to share data and storage devices. The performance characteristics, availability, and growth of the storage subsystem should be chosen and designed to match the rest of the VAXcluster system.

Capacity and input/output (I/O) rate are the key criteria in selecting storage devices, but they do not make the complete criteria set. Floor print, backup requirements, and storage management, for example, can play a role in storage subsystem design.

This chapter assists you in designing your storage hierarchy and selecting storage devices according to your desired interconnect. This includes storage devices connected to:

- CI
- Digital Storage Systems Interconnect (DSSI)
- Boot servers, file servers, and workstations
- Local adapters, for local storage

Note

Currently, FDDI cannot be connected directly to storage, but storage can be accessed through CPUs that have access to disks, either directly or through storage controllers. Ethernet can directly connect only to the InfoServer as a storage device.

All storage subsystems contain a hierarchy of storage devices. Often, there are storage devices that hold several megabytes with rapid access times, which the VAXcluster uses for the most frequent and repetitive data I/O (system functions, temporary storage, cache, and so on). The next level of the hierarchy is formed by disks that store information requiring intermittent VAXcluster access. Tape devices reside at the bottom of the storage hierarchy. They are primarily used for backup storage and contain seldom-accessed or archived data. Figure 5–1 contains a conceptual drawing of the storage hierarchy.

5.1 Storage Subsystem Design Preliminaries

Designing the storage subsystem requires an understanding of the applications, data, and environment under consideration. This top-down approach leads to knowledge of the following elements of the storage subsystem:

- Capacity
 - How much storage is required?

- Performance
 - What is the rate at which each application makes requests of the system during the peak period?
 - How much data is retrieved at each reference?
- Availability
 - How much data redundancy is needed to ensure data availability?
 - What backup strategy should be used?

With this knowledge, you can often discern what types of equipment you will need and design a preliminary storage subsystem configuration.

These details are often not available early in the planning stages. For this reason, this chapter also contains methods of approximating the I/O load for an initial storage subsystem design.

5.2 Description of a Storage Hierarchy

A storage hierarchy is a logical organization of storage devices whose placement within the organization is ordered, or ranked, according to certain attributes. *Price/performance* and *price/capacity* are typically the ordering attributes.

Figure 5–1 Storage Hierarchy

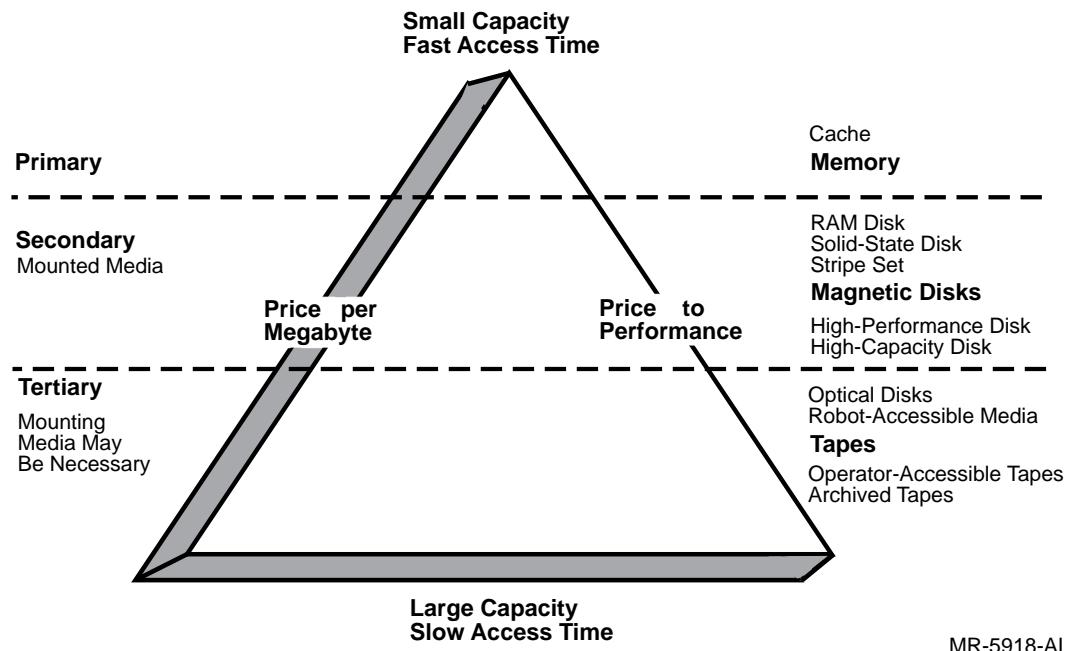
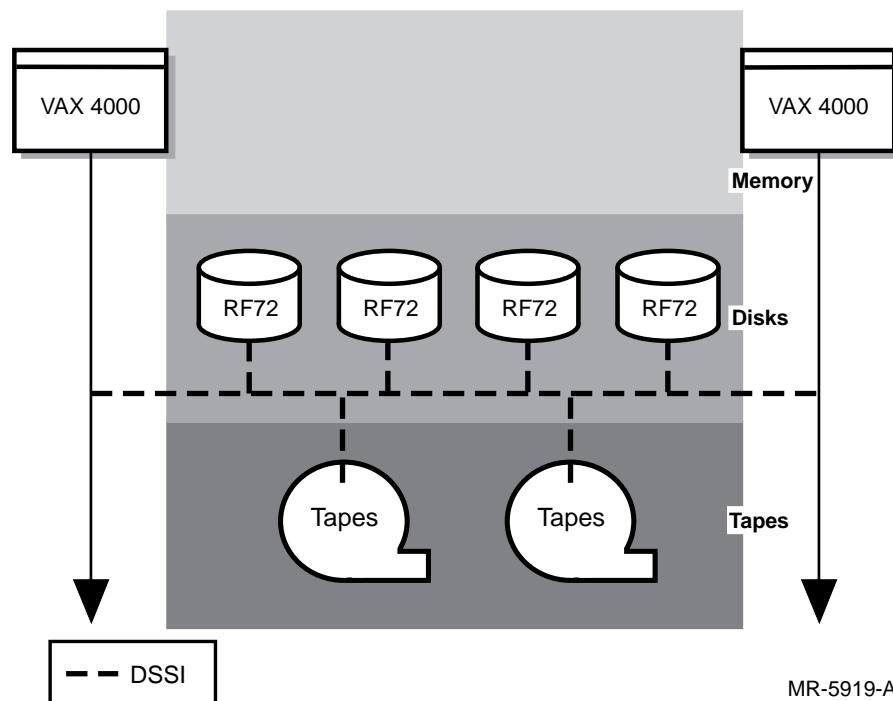


Figure 5–2 illustrates memory, disks, and tapes in a storage hierarchy. The basic levels in the hierarchy are:

1. Primary — Memory used for storage. All data must pass through memory before being used by the system's processors. Response times are measured in nanoseconds.

2. Secondary — Online storage, typically, magnetic disks, where data is continuously accessible at tens of milliseconds response time. Disks composed of solid-state memory, such as ESE-series solid-state disks, are still considered online storage because, to the operating system, they appear as disks.
3. Tertiary — Offline storage, usually tape, but also includes robot-accessible media, where data accessibility is limited by:
 - Drive and operator (or robot) availability
 - Ability to determine the media on which the file exists
 - Time to locate and mount the media, resulting in response times that are typically tens of seconds for robot-accessible media to minutes or even days for manual operation

Figure 5–2 Storage Hierarchy in a Simple Configuration



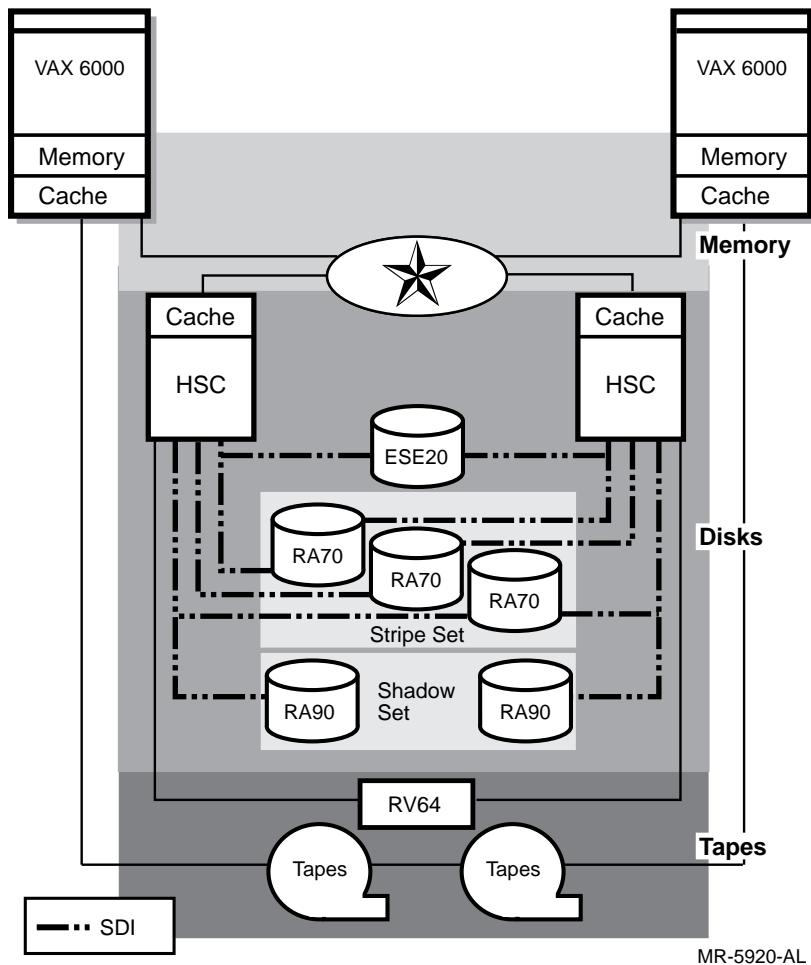
In most storage subsystems, old data and files expire from use and new data and files come into active use. The old data and files can be moved from online to backup storage to free online storage for new data and files. Often, to meet high-performance requirements, sites need many online storage devices with very high performance.

Managing a storage subsystem may require moving data and files from a storage device at one level to a storage device at another level. The storage subsystem of Figure 5–3 shows a more complex storage configuration.

The storage devices needed in your storage subsystem depend primarily on the following:

- Performance requirements of the applications

Figure 5–3 Complex Storage Subsystem



- Frequency of information retrieval from the storage subsystem
- Creation rate of new data and files
- Information backup policies of the computing environment

5.3 How to Design Your Storage Hierarchy

The process for designing an effective storage subsystem includes:

- Estimating the storage capacity required
- Designing an initial storage hierarchy
- Modifying the design for application performance requirements
- Refining the hierarchy for application availability requirements
- Considering storage management options

Each step is described in a separate section. By the end of this chapter, you should have a working knowledge of what your storage needs are and what products can best be used to meet them. You may want a VAXcluster design expert to assist you with a rigorous design of a storage subsystem. Your account representative can put you in touch with these resources.

5.3.1 Gather Application I/O Requirements

When designing a storage hierarchy, you must understand the applications that will use the storage subsystem. The storage capacity needed is the accumulation of the individual applications' capacity requirements to install and operate. The storage performance needed is determined from the I/O loading of the mix of active applications during a day's peak period.

Sometimes there are two peaks — one during interactive daytime hours (typically, 10:00 a.m.) and another during overnight batch processing. Other peaks might occur during shift change, system backup, or at end-of-month accounting. You should design the storage hierarchy to handle the work load of the peak period.

Understanding the work load can be made easier and more exact if the applications can be characterized on an existing system using DEC Performance Solution (DECps) software (formerly the VAX Performance Advisor (VPA)). The modeling part of DECps software can predict the performance of a VAXcluster design in handling a work load previously characterized by DECps. If there is no existing VAX system running the applications, examine known characteristics of similar industry applications.

5.3.2 Estimate Storage Capacity Requirements

This section provides guidelines for estimating the storage capacity requirements for each level of the storage subsystem.

Note

Storage capacity is measured in blocks. There are 512 bytes per block.

Memory Storage Capacity

The memory included with the CPU is designed to support the processing power of the CPU. Most applications find this amount of memory sufficient for their needs.

Online Storage (Disks) Capacity

To estimate capacity requirements for secondary storage, use Table 5–1 to help you estimate the required amount of online storage.

Follow these directions to complete the Online Storage Capacity Work Sheet in Table 5–1:

1. VMS operating system — Enter the number of blocks required by the VMS operating system. See Appendix A for this information.
2. Paging, swapping, and dump files — VMS installation software calculates paging, swapping, and dump file sizes and places the files on the target system disk. Place these files on local storage devices. You may choose to alter the file sizes, depending on application requirements. Sample block sizes for paging files are 40,000 blocks for workstations and 200,000 blocks for VAX 8XXX CPUs. The *Guide to Setting Up a VMS System* can provide you with more information on how to calculate and modify these file sizes.

Performance improves if you place paging and swapping files on disks other than the system disk. See Section 7.5 for a discussion of this.

3. Site-specific utilities and data — Enter an estimate of the disk storage requirements for site-specific utilities, command procedures, online documents, and associated data files.

4. Digital layered products — Enter the space required for each Digital layered product to be installed on your VAXcluster system. Consult the appropriate Software Product Description (SPD) or Appendix A in this book to estimate the space required for normal operation of any layered product you need to use.
5. Third-party application programs — Enter an estimate for the space required for third-party application programs and their associated databases, using information from the suppliers.
6. User data — Enter an estimate of users' space requirements.
 - The *single application* user utilizes only a specific application, for example, order entry, and needs space only for user initialization files. Allocate 100 blocks for each application user.
 - The *occasional* user only reads/writes/deletes VAXmail, has few, if any, programs, and has little need to keep files for any length of time. Allocate from 500 to 5000 blocks for each occasional user.
 - The *moderate* user employs the system extensively for electronic communications, keeps information on line, and has a few programs for private use. Allocate from 10,000 to 50,000 blocks for each moderate user.
 - The *programmer/developer* can require a significant amount of storage space for programs under development and data files, in addition to normal system use for electronic mail. This user may require more than 100,000 blocks of storage, perhaps several hundred thousand, depending on the number of projects and programs being developed or maintained.
7. Application-specific databases or other shared files — Enter an estimate of the size of each database. This information should be available in the documentation pertaining to the application-specific database.
8. Anticipated growth — Enter an estimate for anticipated growth for your storage needs. Consider how you deal with obsolete or archival data to ensure adequate storage capacity for the future.
9. Total — Add up the online capacity requirements. The result is the amount of disk storage needed for your VAXcluster system configuration.

Table 5–1 Online Storage Capacity Work Sheet

Software Component	Blocks of Storage
VMS operating system	
Paging, swapping, and dump files	
Site-specific utilities and data	
Digital layered products	
Third-party application programs	
Databases	
Libraries	
User data	No. of users Blocks per user
Single application	x 100 =
Occasional	x 3000 =
Moderate	x 30,000 =
Heavy	x 200,000 =
Anticipated growth	
Total	

Backup Storage (Tapes and Robot-Accessible Libraries) Capacity

Backup storage provides the least expensive storage medium. Tapes are the most common medium for tertiary storage and provide a range of capacities, cost, and shelf life. What distinguishes backup storage from the other storage levels is that its media are removable and generally off line.

Be sure that sufficient backup media are available to meet the initial backup needs of the site. This depends upon how much data will be backed up daily, weekly, and monthly. Will you conduct full or incremental backups? How often for each? Consult the documents of the backup and storage management utilities, such as the VAX Storage Library System (SLS), to assist in determining both the initial amount of backup media and its growth rate. Backup strategies and procedures are available in detail in the *VMS Backup Utility Manual*, *Using VMS BACKUP*, and the *TA90 and TA90E Configuration and Backup Performance Guidebook*.

5.3.3 Consider Additional Storage Subsystem Attributes

This section will assist you in evaluating the adequacy of your initial design.

5.3.3.1 How CPU Selections Affect Storage Subsystem Design

The more powerful VAX computers can have much of their computing potential wasted if the storage subsystem cannot meet the computer's appetite for I/O. This is also true for the aggregate compute power of a VAXcluster system. Because storage technology in the form of magnetic disks has not kept pace with performance gains in CPU technology, solid-state disks, disk striping, or controllers with disk cache (HSC60/90 and RFxx disks) should be considered when VAXcluster systems contain more than 25 VUPs or when applications present heavy I/O loads.

5.3.3.2 Select Storage Devices to Meet Your Requirements

See Tables 5–4 and 5–5 to select online storage devices; see Table 5–6 to select appropriate storage arrays, and see Tables 5–7 and 5–8 to select backup storage devices. In forming your initial storage hierarchy, select online storage devices that best meet price/capacity requirements, such as the RA92 for CI configurations and the RF73 for DSSI configurations. These selections can be modified to meet performance requirements, as discussed in Section 5.3.4).

5.3.4 Gather Additional Data

Use the work sheet provided in Table 5–2 to assist you in estimating your online storage performance and availability or I/O work load.

In many timesharing systems, up to half of the I/O requests can be demanding information from about 1 percent of the data stored online. ESE-series devices can yield the rapid access needed for this subset of the data.

DECps can collect data on I/O rates in a VAXcluster system, and documentation for some applications can also yield this information. In addition, using modeling techniques, DECps can do capacity planning. The gathered data becomes a work load, adjustable to the mix of applications planned for your VAXcluster system. Your storage subsystem design can be presented to the modeling portion of DECps to assess the design's performance characteristics. You can also assess additional theoretical designs.

Benchmarks may have data that characterize applications like those anticipated for your VAXcluster system. This data can be used to construct an I/O work load that can, in turn, be applied in forming a storage subsystem design. These I/O workload characterizations can be helpful in making a general storage subsystem design, but do not lead to accurate placement of data and files or distinguish hot files that need exceptionally high-performing storage.

DECcmds is a system management tool that monitors, identifies, and quickly resolves problems associated with down time. This package allows the system manager to investigate and diagnose problems with low memory, I/O, disks, paging files, and swapping files in real time.

5.3.4.1 Online Storage Performance and Availability Work Sheet

Table 5–2 will help you design your storage subsystem. Follow these directions to complete the Online Storage Performance and Availability Work Sheet:

1. Select a peak period of activity, for example 10:00 to 10:30 a.m.
2. Enter the number of *active* users during the peak period.
3. List the batch jobs and associated data areas (directories where files or databases reside) active during the peak period. Use a separate line for each data area.
4. List the applications from Digital's layered products and associated data areas or directories active during the peak period. Use a separate line for each data area.
5. List the applications from other vendors or user-written and associated data areas or directories that are active during the peak period. Use a separate line for each data area.
6. For each application data area pair, enter the three I/O fields. The I/Os per second can be estimated by summing the reads and writes to the data area by the application, multiplying that by the number of application executions per user expected during the peak period, and multiplying that by the number of active users. Estimate the average size in bytes or blocks (stay consistent) for each I/O and the ratio between reads and writes. I/O size is used in ascertaining the advantage of disk striping or other techniques. High read-to-write ratios indicate candidates for caches or random-access memory (RAM) disks.
7. Complete the performance fields of the work sheet by filling in the performance goal of the application. For batch jobs, this may be elapsed time. For interactive applications, this may be response time or throughput. Use this information to place the data areas of the applications with the more critical performance goals on those devices that best meet the performance needs.
8. Ignore the Mission Critical column for the timebeing. This is used in Section 5.3.5, which discusses availability requirements.

Table 5–2 Online Storage Performance and Availability Work Sheet

Software Component	Data Area Accessed	I/Os Per Second	Average I/O Size	Performance Goal	Mission Critical
Batch Jobs					
Digital layered products					
Third-party application programs					
Total					

5.3.4.2 Modify the Storage Hierarchy According to Performance Requirements

Group the data areas according to similar performance requirements of the applications that reference them. There are always files or database areas that are referenced more or less than other files. Put those that need the best performance at the top and continue ordering through all the files and databases. Use Table 5–2 to assist in this process.

5.3.4.3 Performance Considerations for HSC Controller Subsystems

For maximum performance and to minimize the possibility of data bus contention in the HSC, it is important to configure the devices on the HSC according to the order of their priority. The requester number assigned to the device defines what its priority is to the HSC and the VAXcluster. Follow these guidelines:

1. To use the data bus bandwidth of the HSC as economically as possible, configure devices in order of peak transfer speed. See Tables 5–5 and 5–8 for transfer speed information (measured in I/Os per second).

Note that the transfer speed is not equivalent to the I/O per second rate. They are different variables.

2. Place stripe set and shadow set members on separate requesters. Since only one drive on a requester can transfer data at a time, separating the members allows simultaneous reads and writes to both members.
3. Because only one device can transfer data at a time on a requester, devices that transfer data using large transfer sizes, such as a paging or swapping file disk, should be on a separate requester or on a requester with low-usage devices.

The HSC60 and HSC90 can have up to eight drives on a requester.

5.3.4.4 Backup Storage Performance

Tape Drive Performance

The TA91 offers the highest performance for STI tapes. Its magazine of IBM 3480 compatible cartridge tapes lets it back up 38 gigabytes unattended. Like the TA90, its formatter can transfer at 2.6 megabytes per second. To achieve this performance, connect a TA90 or TA91 through a KDM70 controller or have it reside in a configuration with multiple CI adapters, so that the path to the tape drives is separate from the path to the disk drives.

The TF867 offers the best tape performance for DSSI configurations. Its magazine of half-inch cartridge tapes can hold up to 42 gigabytes of data for unattended backup. Its transfer rate is 0.8 megabytes per second. The TF857 can read TK50 and TK70 tapes, and its magazine can hold up to 18 gigabytes of data.

The TSZ07 allows Small Computer System Interface (SCSI) configurations to access 9-track reel-to-reel tapes. Its capacity and performance are similar to the TA79, namely 140 megabytes per reel and 750 kilobytes per second transfer rate. The TZK10 offers a less expensive, but slower, performing tape solution for SCSI configurations. It uses a quarter-inch cartridge that holds 525 megabytes and can transfer at 200 kilobytes per second.

StorageTek 4400 ACS

You can attach the StorageTek 4400 ACS, a storage *silo*, to an HSC using the TC44 adapter, or directly to the XMI bus of a VAX 6000 using a KCM44 adapter. The StorageTek Silo automates access to a library of IBM 3480-compatible cartridge tapes. The library can contain up to 16 library storage modules. Each module can hold up to 1.2 terabytes of data (or more with ICRC data compaction) in 6,000 tape cartridges. A robotic arm can find and mount a requested tape within 45 to 90 seconds. Data movement for tape applications, such as BACKUP, is performed in the same way as with a TA90 tape drive.

InfoServer 150 Software

The InfoServer 150 supports a maximum of 100 unique nodes. A single InfoServer with 2 CD drives can serve hundreds of users. The InfoServer 150 provides access to compact disk read-only memory and shared read/write access to RZ (SCSI) disks. InfoServer 150 performance is enhanced by using most of its 4 megabytes of internal memory as a cache.

5.3.4.5 Performance Considerations When Including Newer Technology Storage

Recent advances in disk technologies have been in reliability and in decreasing the size of the disk while improving data density. The result is smaller disks with capacity similar to the previous generation of larger disks.

However, new storage technologies are using semiconductor memory to form disk caches. Accesses that can be satisfied from the cache can be done almost immediately and without any seek time or rotational latency. For these accesses, the two largest components of the I/O response time are eliminated. The HSC60 and HSC90 contain caches. The RF disks have a disk cache as part of the embedded controller contained in the Integrated Storage Element (ISE).

When replacing or upgrading mature HSC models (HSC40, HSC50, and HSC70 subsystems), configure the most active disks across the SDI channels of the newer subsystem to ensure the most effective use of disk cache.

VAXcluster systems with at least one VAX 6000, VAX 4000, or VAXft processor can accommodate DSSI disks and tapes. The TF800 family provides an unattended backup solution. A VAX 6000 system configured with RF disks can serve MSCP commands for the other CPUs in the VAXcluster system.

5.3.5 Meet Availability Requirements

For storage subsystems, there are two kinds of availability — availability of the device and availability of the data. A choice arises between the cost to protect the data and the cost of periods of unavailable data. Factor into this the probability of a failure that would render the data unavailable. Thus your storage subsystem design needs to evaluate both mean-time-between-failures (MTBF) and the mean-time-to-repair (MTTR) a failure.

One extreme is when the cost of unavailable data is negligible and data availability can be maintained by backups. The other extreme is when a business can lose tens of thousands of dollars for every minute the data is unavailable, when a production line needs to be stopped, or when controls for a critical resource may be unable to adjust, causing its destruction. In such environments, extreme attention to availability is easily justified.

It may also be critical to note whether or not your applications can tolerate data loss or data inaccessibility (memory cannot be reached).

Volume shadowing can be used to increase data availability, by making several sets of data available. By using the VAXcluster Multi-Datacenter Facility (MDF), these shadow sets can be available in two separate locations which combine to form one VAXcluster system.

More information on increasing availability can be found in *Building Dependable Systems: The VMS Approach*.

5.3.5.1 Complete Your Online Storage Performance and Availability Work Sheet

You have already filled out Table 5–2 in designing your storage subsystem. Follow these directions to complete the Online Storage Performance and Availability Work Sheet:

For each application data area pair, enter a Y in the Mission Critical column if the data area is crucial to the business. Such data areas are candidates for placement within a shadow set, placement on a dual-ported or dual-hosted disk, placement on a disk with multiple access paths to it, or some combination of these.

Note that system disks and page and swap disks are considered crucial if a particular processor is crucial or if one common system disk is used.

5.3.5.2 Device Availability

The methods for ensuring device availability employ multiple access paths to a storage device. These can take the following three forms:

1. Two HSC subsystems connected to the same Star Coupler and storage devices dual ported between the two HSC subsystems¹
2. SDI disks dual ported between controllers on two CPUs in the same VAXcluster system
3. DSSI disks connected to two CPUs

For the first two methods, in the event of a failure along the main memory access path, a failover to the other access path occurs, thereby maintaining access to the device where the data resides or is to be stored. For the third method, both CPUs have continuous access to each disk.

Configuring multiple access paths protects against hardware failures along the path to the device, but does not protect against failure of the device itself. VAXsimPLUS, in conjunction with volume shadowing, can detect most imminent device failures with sufficient lead time to move the data on the device to a spare. For mission critical data, this is not likely to provide sufficient protection.

5.3.5.3 Data Availability

High data availability can be maintained in a VAXcluster system in spite of hardware failures by designing the storage subsystem with multiple devices of the same kind and VMS Volume Shadowing software. Up to three devices are used to form a shadow set. VMS Volume Shadowing software maintains multiple copies of data to protect against data loss caused by disk or device failure. In addition, the disks of the shadow set can be placed on separate access paths, thereby accomplishing the protection against hardware failures along the path to the device, discussed in Section 5.3.5.2. With multiple access paths and VMS Volume Shadowing, data continues to be available, even when media deterioration or failure causes a device failure. Since disk failure occurs without data access interruption, data continues to be available. Any single disk failure is transparent to the VAXcluster.

All disks connected to an HSC may participate as shadow set members in VMS Volume Shadowing. All DSSI storage, SCSI storage, and local RA-series storage can also be combined into shadow sets. Shadow sets consist of two or three identical disk types. For more information on Volume Shadowing, see Section 5.3.5.6

Device reliability affects data availability. Typically, newer devices continue to improve their reliability and MTBF. Newer controllers also improve reliability by taking advantage of newer chip technologies. By using newer devices, you may increase the reliability and availability of your VAXcluster configuration.

Device reliability can be enhanced with the use of appropriate software tools. Use device failure prediction tools, such as VAXsimPLUS, where high availability is needed. The storage subsystem using VAXsimPLUS typically has a spare device that can be used to create a shadow set copy of a device whose increasing fault rate indicates a future failure. After the copy is made, the suspect device can be taken off the system for examination and repair without any loss of data availability.

¹ Dual porting storage devices between HSC subsystems that are connected to separate Star Couplers is not supported.

The following list provides availability features that the interconnects offer the storage subsystem:

- CI — HSC subsystems provide multiple access paths for both disks and tapes in a VAXcluster system. Disk and tape mount verification supply automatic failover and powerfail recovery for HSC controller disks and tapes. For information on HSC controller channel and drive support, see Table 5–3.

Table 5–3 HSC Controller Channel/Drive Support

HSC Model	Channel Modules per HSC ¹	Disk Drives per Channel Module ²	Drives per HSC ³
HSC40	3	4	12
HSC50	6	4	24
HSC60	3	4-8 ⁴	20
HSC70	8	4	32
HSC90	8	4-8 ⁵	48

¹Maximum number of data channel modules that can be supported by that HSC model.

²Maximum number of disk drives that can be supported by each data channel module on that HSC model.

³Maximum number of disk drives that can be supported on that HSC model with the full complement of channel modules installed.

⁴There can be up to two 8-port modules and one 4-port module, or three 4-port modules.

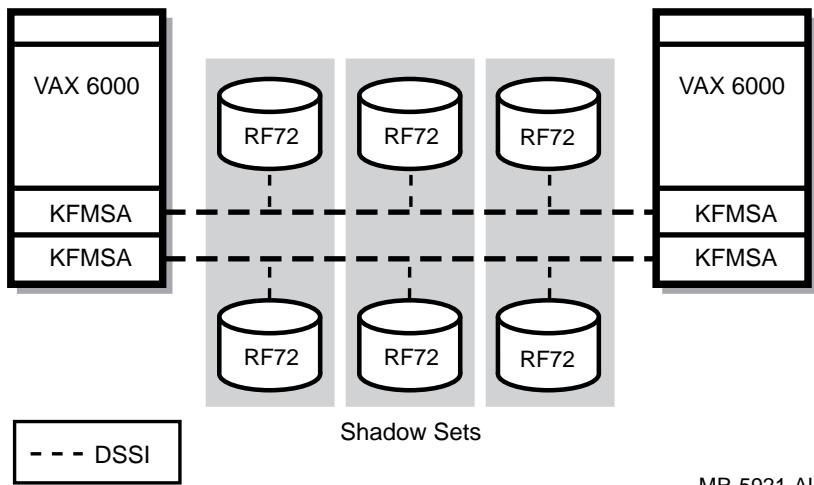
⁵There can be up to four modules that support eight drives, and the rest can accommodate 4-drive modules.

Disk drives that are dual ported between two HSC controller subsystems can provide both automatic failover and system manager-controlled, static load balancing. HSC subsystems with multihost access paths also allow offline diagnostics to be run on a failed HSC controller subsystem without bringing down all the disk drives on the system. With VMS Volume Shadowing, place shadow set members on separate controllers for optimal availability.

- DSSI — DSSI ISEs and KFMSA adapters are highly reliable because their MTBF is very high. In addition, to minimize the impact of CPU failure, DSSI disks can be dual *hosted*, where the DSSI bus has a CPU at each end. Trihost configurations are also supported. In these configurations, disks are simultaneously accessible to multiple CPUs. A DSSI bus supports up to eight nodes, so, for example, with two VAX 4000 hosts, you can attach up to six DSSI disks to that bus. Note that the KFMSA has two DSSI buses and that each bus can support seven disks, for a total of 14.

DSSI disks cannot be dual ported to another DSSI bus running between the same two hosts. Any of these disks can form a shadow set with a disk of the exact same type on this bus or a separate DSSI bus, as shown in Figure 5–4. Use of VMS Volume Shadowing software protects data availability despite hardware failures to device, interconnect, or CPU.

Figure 5–4 DSSI Shadow Sets



MR-5921-AL

- Local adapters — You can use local adapters, such as the UDA50, KDA50, KDB50, and KDM70, to connect each disk to two access paths (dual ports). Dual porting allows failover of disks between CPUs that support the UNIBUS, Q-bus, BI, or XMI. This sets up multiple access paths that can fail over to the other path automatically.

In DSSI, Ethernet, or mixed-interconnect VAXcluster systems, you can configure dual-disk servers with local system disks and allow sharing of the disk containing satellite system roots between the servers, as shown in Figure 5–5. Figure 5–6 shows how system disks can be dual-ported in a VAXcluster, and served out to an Ethernet. The two DSSI buses are supporting separate shadowed sets of disks. This type of configuration ensures that disk servers are highly available; even if one path or adapter fails, the VAXcluster system can still access the disk through the second server.

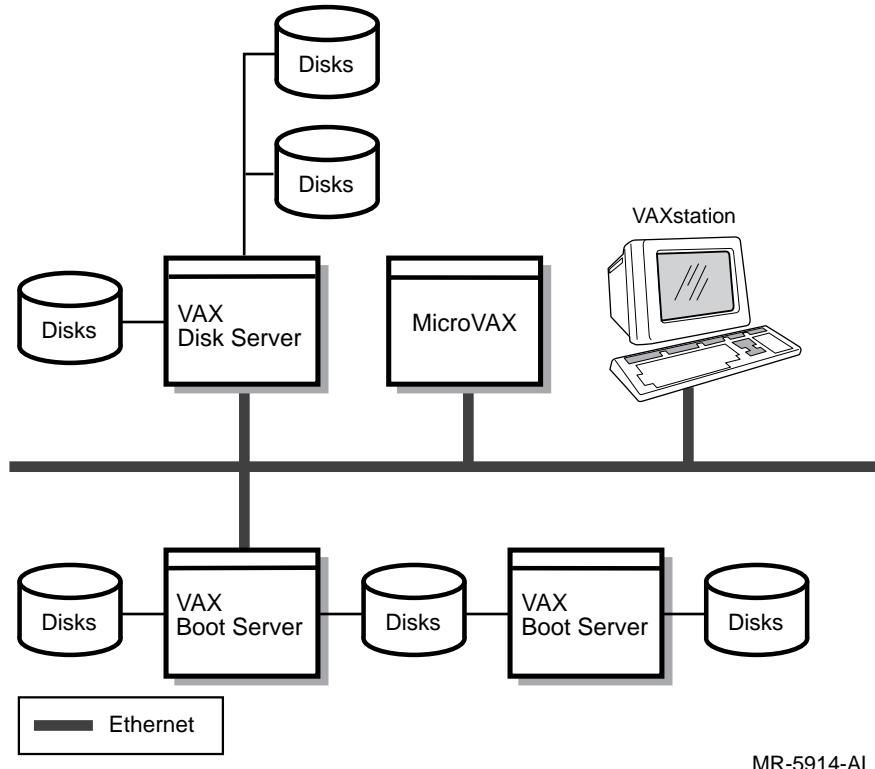
5.3.5.4 System Disk Redundancy

You can optimize availability and redundancy of system software by judicious placement of system files on disk drives with multiple access paths. You can also configure a VAXcluster system with any number of system disks; these system disks can contain identical software components for maximum redundancy. However, multiple system disks require increased system management. This problem is simplified when you can form a shadow set of the system disk (see Section 5.3.5.6 for a description of shadow sets).

5.3.5.5 Site Redundancy

High availability can be obtained with a VAXcluster MDF configuration. MDF VAXcluster configurations follow design guidelines that enhance availability with redundancy and that support great distances between two sites. If one site suffers down time because of some physical or power-based phenomenon, the other site is still available for use.

Figure 5–5 Dual-Hosted Mixed-interconnect Disks



MR-5914-AL

5.3.5.6 VMS Volume Shadowing Software

VMS Volume Shadowing lets shadow set members be anywhere within the storage subsystem of a VAXcluster system, subject to restrictions listed later in this section.

VMS Volume Shadowing software replicates data written to a virtual disk by writing the data to one or more physically identical disks that form a shadow set. With replicated data, users can access data even when one disk becomes unavailable. If one shadow set member fails, VMS Volume Shadowing software takes the drive out of the shadow set, and processing continues with the remaining drives. Shadowing is transparent to applications and enables data storage and delivery despite media, disk, controller, and interconnect failures.

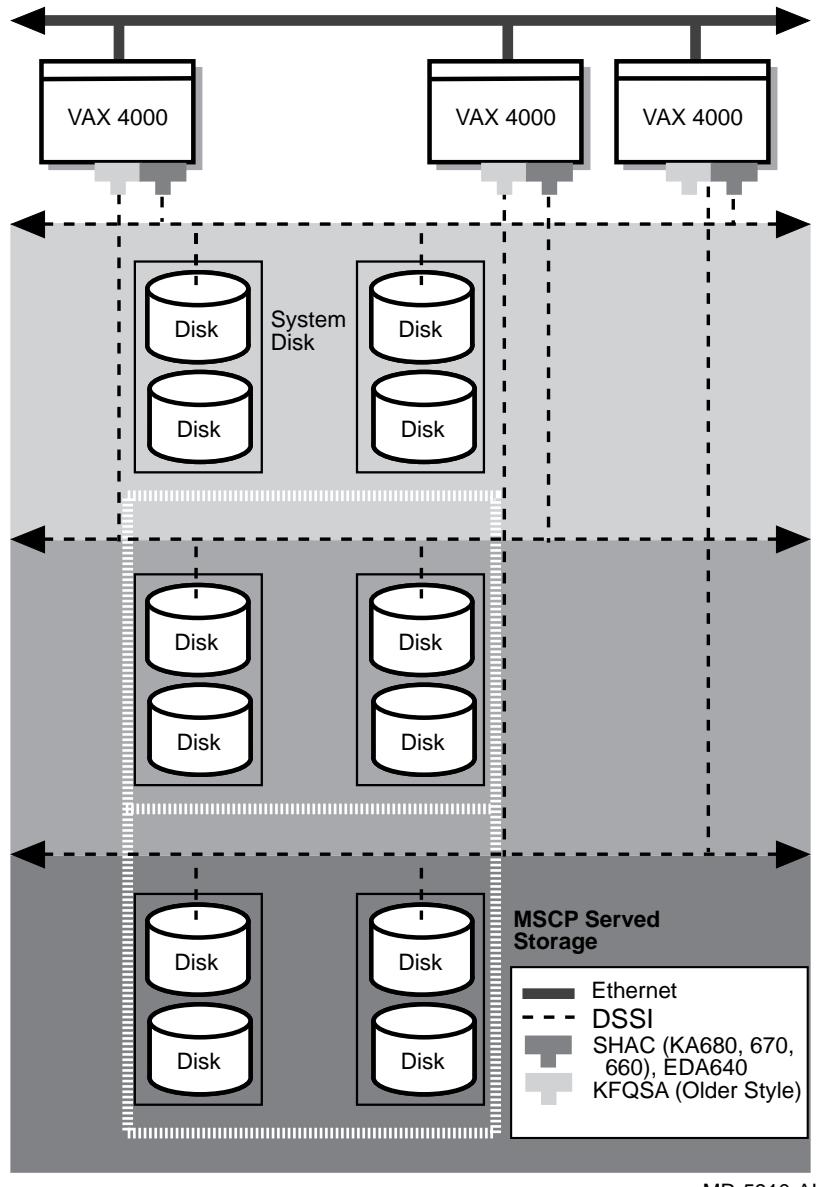
VMS Volume Shadowing software ensures that shadow set members maintain identical data. Each time that data is written to the virtual unit that represents the shadow set, the corresponding logical blocks of each member record the data. When a disk volume is temporarily removed from a shadow set and then remounted, the data on that volume is updated to reflect the data on the remaining volumes of the shadow set.

VMS Volume Shadowing software supports up to three members in any given shadow set. Often two-member shadow sets are sufficient. Highly critical applications may justify three-member shadow sets to reduce the risk of data loss to a minimum.

With VMS Volume Shadowing software, consider the following:

- All redundant disks of a given shadow set must have the same size and exact number of logical blocks. Use two or three of the same type of disk to form a shadow set.

Figure 5–6 Dual-Ported Mixed-interconnect Disks



MR-5910-AL

- Shadow sets can be user disks or system disks; a quorum disk cannot be a shadow set member.
- Shadow set members can be any RA, RF, or ESE drives and can be located anywhere within the VAXcluster system. SCSI disks that support READL/WRITEL can also be members of shadow sets.
- Each shadow set cannot have more than three members.
- There can be up to 75 shadow sets in a VAXcluster system.
- For additional performance and reliability, place shadow set members on separate data channel modules.

- If you are using the *hot spare* repair strategy, where a spare disk drive is kept available to replace a failing drive, exercise the spare periodically to ensure that it functions when needed. This applies to HSC controller subsystems as well. You cannot use the spare disk to store data because a copy of the failing disk is made to the spare disk by including it in a shadow set with the failing disk. At this point, all existing data on the spare disk is overwritten.
- Shadowing can also be used to accomplish a rapid disk backup. For example, when no disk write activity is in progress, a three-member shadow set can be reduced to two members by dismounting the shadow set, remounting the shadow set with two members, and copying the third disk to magnetic tape. After this, the third disk can be reincorporated in the shadow set. When using this method, try to ensure that all writes complete and no new writes are initiated before you remove the third disk from the shadow set. Note that the backup disk must be the same device type as the disk being replaced.

For more information on Volume Shadowing, see the *VMS Volume Shadowing Manual*.

Disk Striping

Disk striping can improve performance. This technique lets applications access an array of disk drives in parallel for higher throughput. Disk striping works by grouping several disks into a *stripe set* and dividing the application data into *chunks*, which are spread equally across the disks in the stripe set in a round-robin fashion. Then, any given I/O request is broken up by VMS software into several requests sent out to the disks of the stripe set at the same time. The data is not shadowed or stored redundantly, unless you use volume shadowing on individual stripe set members.

By reducing access time, disk striping may improve performance, especially if the application:

- Performs large data transfers in parallel
- Requires load balancing across drives

5.3.5.7 Redundancy Through Backup Strategy

You may want to maintain redundant copies of certain files or partitions of databases that are, for example, updated overnight by batch jobs. Rather than using shadow sets, which maintain a complete copy of the entire disk, it might be sufficient to maintain a BACKUP-produced copy on another disk or even on a hot standby tape of selected files or databases. See Section 5.3.6.2 to determine backup requirements that can affect your storage subsystem design.

5.3.6 Storage Management Considerations

At this point, you have designed a storage hierarchy to meet the application needs of the user. This satisfies the hardware parts, but is not yet a working solution. Storage management includes operating the storage hierarchy, maintaining optimal placement of data, ensuring data integrity, planning for future capacity and performance changes, and minimizing costs. To help manage these, you need to consider:

- Disk utilization and fragmentation
- Backup

5.3.6.1 Disk Utilization and Fragmentation

Typically, a system can use all the formatted space on a disk volume. However, as available disk space shrinks, *file fragmentation* usually increases. Disk fragmentation means that files are written in numerous, noncontiguous areas on the disk and can cause a degradation in system performance because more seeks to read a single file are required. One way to minimize disk fragmentation is to keep disk utilization between 50 and 70 percent.

In a highly active environment, files can be created and deleted on the same disk from more than one VAXcluster system CPU. Disk contention can take the form of cache invalidations and more frequent writes and reads of the portions of the bitmap associated with available space.

You can minimize disk fragmentation and provide more effective caching by estimating available disk storage at no more than 70 percent of disk capacity. If a disk contains only files that are created once and kept indefinitely, with no deletions or extensions, the disk can be kept almost full.

Digital offers the Disk File Optimizer (DFO), which performs disk defragmentation.

5.3.6.2 Backup

In any computer system, hardware and electrical failures and human error occur. All data that should not be lost must be backed up to limit the effects of these errors. There are a number of ways to do this, depending on the time and resources available.

First, the system or application manager must decide how much lost work is acceptable in the event of a failure. This determines how often the data needs to be backed up. Second, the system or application manager must decide how long the data can remain unavailable while it is being backed up. Finally, a backup schedule needs to be established, including the frequency and times of the day and week that backups will occur.

Determining Your Backup Strategy

The following are ways of providing a copy of data for backup:

- For static data, such as the sources of programs in production, documentation files, and distribution kits, you may be satisfied to have copies of the data archived on magnetic tape and exclude the online files from any other backup procedure.
- For databases that are continually changing and in which transactions cannot be lost, use a combination of backup of the database, at a time when it is known to be static, and journaling transactions to the database. See the following manuals for additional information:
 - *VAX RMS Journaling Manual*
 - *Guide to VMS File Applications*
 - *VAX Rdb/VMS Guide to Database Maintenance and Performance*
 - *VAX Rdb/VMS Guide to Database Design and Definition*
 - *VAX DBMS Database Design Guide*
 - *VAX DBMS Database Maintenance and Performance Guide*

- For data that must be accessible all the time (including nights and weekends), use Volume Shadowing software to create an extra copy of the data (see Section 5.3.5.3).
 - When a set of data can be unavailable for an extended period of time for backup, use BACKUP to make an image copy of a volume or a file-by-file copy of specified sets of files. BACKUP can make a copy on another disk (or set of disks) or on magnetic tape. Restoring from an image copy requires that the entire image be written to a disk. When you restore specific files, they are copied from the restored disk to the intended destination.
- On the other hand, an image copy is faster than a file-by-file copy. With a file-by-file copy, the files are copied one at a time. Restoring a single file from the backup copy is easy. Also, a file-by-file restore virtually eliminates fragmentation of the restored disk. See the *VMS Backup Utility Manual*, the *Introduction to VMS System Management*, or the *VMS System Manager's Manual* for more information.
- Some files, such as scratch files, are intermediate files and can be readily recreated from other files. You may choose not to provide any backup for these files.

With current tape drive technology, you can initiate a large backup operation that completes without operator intervention to change tapes. Such unattended backups can save significant time and reduce staffing costs. Cartridge tape loaders with tape magazines, such as the TF8X7 or the TA91, provide for unattended backups of up to nearly 20 gigabytes of online storage. Backups can also be done to robot-accessible media, such as the StorageTek 4400 ACS through the TC44 interconnect adapter, which provides terabyte capacity for backup archives.

5.3.6.3 VAX SLS

VAX SLS is a VMS layered software product that manages reel-to-reel magnetic tape, cartridge tape (TA90 and TF857 media loaders), and optical cartridges (RV20).

VAX SLS software maintains records of all files backed up or archived by using its services for quick retrieval by users, operators, and storage administrators.

5.3.6.4 Floor Space for your Storage Devices

Where the cost of floor space is high, you want to minimize the floor space used for storage devices in your design. The SA and SF disk storage arrays were designed for this purpose. In addition, trade-offs between higher performing disks and higher capacity disks may favor higher capacity when floor space costs are high. A business practice of regular upgrades to newer technology storage arrays or disks may need to be part of your capacity planning process. For example, replacing an SA482 with an SA800 increases capacity from 2.4 to 12.0 gigabytes without affecting floor space consumption. Several storage devices come in stackable cabinets for labs with higher ceilings.

Similarly, factors such as disaster tolerance, power, and air-conditioning may affect the design or selection of storage devices for the storage subsystem. These are beyond the scope of this book.

5.4 Storage Device Characteristics

The following tables let you compare the attributes of disks, tapes, and library systems.

Table 5–4 lists storage devices by the interconnect/bus they use. Table 5–5 provides basic information to help you choose disks for your applications.

Digital has packaged disk drives together into storage arrays. For example, an SAXXX disk option is a multidisk package made up of several RA disk drives. Total capacity is the sum of all drives in that storage array. Table 5–6 provides information on these arrays. Table 5–7 lists large capacity robot-served storage for libraries; Table 5–8 lists tape drives. Table 5–9 lists bus types and I/O rates for SDI controllers.

Table C–8 lists disk attributes for drives that can be used in a VAXcluster configuration, but are no longer shipped by Digital. Table C–9, Table C–10, Table C–11, Table C–12, and Table C–13 list information on other clusterable products no longer shipped by Digital.

Table 5–4 Storage Devices Listed by Interconnect

Interconnect	Storage Devices
SDI	ESE-series, RA-series, SA-series
ST506	RD-series
DSSI	RF-series, SF-series, TF-series
SCSI	RRD-series, TK-series, TS-series, TZ-series
SCSI-2	RZ-series
Q-bus ¹	RRD-series, TK-series, TS-series, TU-series, TZ-series
STI ²	TA-series, RV-series, TU-series
FIPS-60 ³	StorageTek Silo
Ethernet	InfoServer 150

¹Can accommodate RF-series, SF-series, and TF-series with the KFQSA Q-bus-to-DSSI adapter.

²STI includes HSC, KDM70, and STI interconnects.

³One KDM44 supports up to 16 drives.

Table 5–5 Disk Attributes

Disk Drive	Formatted Capacity MB	Formatted Capacity Blocks	Average Access Time ¹	Requests Per Second ²	Spiral Read Rate ³
ESE20	120-240	0.24 M	.23 ⁴	1300 ⁴	1909
ESE50	120-600	.23-1.9 M	.25	1200+	1909
ESE56	640	1.25 M	.25	1200+	1909
RA70	280	0.55 M	27	45	885
RA71	700	1.56 M	20.8	56	1254
RA72	1000	1.95 M	20.8	58	1306
RA90	1216	2.39 M	26.8	46	1722
RA92	1500	2.93 M	24.8	53	1723
RD31	20	0.04 M	73.3	13	625
RD32	42	0.08 M	48.3	20	625
RD54	159	0.31 M	310.5	32	625
RF31T	381	0.75 M	13.1	82	1200
RF35	850	1.66 M	15.1	73	2000
RF71 ⁵	400	0.78 M	29.3	35	750
RF72 ⁵	1024	3.19 M	21.7	44	1300
RF73	2000	3.9 M	21.2	47	2000
RRD42 ⁶	600	1.20 M	450	2	150
RZ23	104	0.20 M	33.4	26	1250
RZ23L	121	0.23 M	26.8	30	1500
RZ24L	245	0.48 M	23	31	1500
RZ25	426	0.83 M	20.8	31	1500
RZ26	1050	2.00 M	15.1	31	1500
RZ55	332	0.65 M	24.3	35	1250
RZ56	665	1.30 M	24.3	35	1875
RZ57	1024	2.10 M	22.8	35	2200
RZ58	1380	2.69 M	18.1	35	5000

¹Average seek plus average latency in milliseconds.

²Measured when average response time was less than 100 milliseconds.

³Measured in kilobytes per second.

⁴Values for HSC configurations using Version 6.0 or later. HSC configurations with Version 5.0A or earlier have an access time of 0.53 milliseconds and a request rate of 300 per second. When connected to the KDM70 controller, use 0.25 and 1200, respectively.

⁵Can be a removable drive.

⁶Only offered as a removable drive.

Table 5–6 Storage Array Attributes

Storage Array¹	Capacity (formatted)	Average Access Time²	I/O Requests Per Second³	Component Drives
SA482	2.4 GB	32 ms	124	RA82
SA600	9.6 GB	27 ms	368	RA90
SA650	9.5 GB	27 ms	640	RA90/RA70
SA705	4.4 GB	27 ms	720	RA70
SA800	12.0 GB	25 ms	424	RA92
SA850	11.3 GB	25/27 ms	640	RA92/RA70
SF35	10.2 GB	15.1 ms	188	RF73
SF72	4.0 GB	22 ms	176	RF72
SF73	10.2 GB	21.2 ms	876	RF73
SF200 ⁴	24 GB	22 ms	1056	RF72
SF210 ⁴	48 GB	21 ms	1128	RF73
SF220 ⁴	61.2 GB	15.1 ms	5256	RF35
SF300	61.2 GB	15.1 ms	5256	RF35
SF400 ⁵	102 GB	15.1 ms	8760	RF35

¹The storage arrays currently available incorporate only disk drives.

²Average seek plus average latency in milliseconds.

³Measured when average response time was less than 100 milliseconds.

⁴SF2XX family can hold any combination of up to six SF72s, SF73s, and SF35s.

⁵SF400 can hold any combination of up to 10 SF72s, SF73s, and SF35s.

Table 5–7 Library Attributes

Media Type	Capacity¹	Fetch Media Time²	Data Transfer³	No. Drives
StorageTek Silo	1200 ⁴	60-90	2300	1-152
InfoServer 150	Varies ⁵	N/A	1000	1-13

¹Total unit capacity in gigabytes.

²Robotic service time in seconds.

³Kilobytes per second, per drive.

⁴1.2 terabytes per library storage module. Up to 19.2 terabytes in 16 connected library storage modules.

⁵Accommodates 1 internal RZ23L with 121 MB, plus up to 12 RZ5X drives

Table 5–8 Tape Attributes

Tape Drive	Density ¹	Speed ²	Data Transfer ^{3, 5}	Recording Method	Media Size	Media Capacity
TA81	6250/1600	25/75 ⁴	468	Group, PE ⁶	190 MB	
TA79	6250/1600	125	781	Group, PE ⁶	190 MB	
TA90	38,000	140	2700	IBM 3480 ⁷	200 MB	9.6 GB
TA91	38,000	140	2700	IBM 3480 ⁷	200 MB	38 GB
TF85	42,500	10 ⁴	800	TTSP ⁸	2.6 GB	2.6 GB
TF85 ^{7,12}	42,500	10 ⁴	800	TTSP ⁸	2.6 GB	18 GB
TK50	6667	75 ⁴	45	SSP ⁹	95 MB	
TK70	10,000	100 ⁴	90	SSP ⁹	295 MB	
TSx05	1600	25 ⁴	40	PE ⁶	40 MB	
TSZ07	6250/1600	25 ⁴	40	Group, PE ⁶	150 MB	
TU81+	6250/1600	25/75 ⁴	468	Group, PE ⁶	150 MB	
TZ30	6667	104 ⁴	62.5	SSP ⁹	95 MB	
TZK10 ¹¹	320/525	120 ⁴	200	QIC ¹⁰	525 MB	525 MB
TZ85	42500	100		MFM	2.6 GB	
TZ857	42500	200		MFM	2.6 GB	

¹Bits per inch.²Inches per second.³Kilobytes per second.⁴Streaming.⁵User data.⁶Group code recording to ANSI Standard X3.54-1976 and Phase Encoded X3.39-1973.⁷IBM 3480-compatible cartridge tape format.⁸Two-track serpentine pattern.⁹Serial serpentine pattern.¹⁰Quarter-inch cartridges.¹¹Available on MicroVAX, VAXserver, VAXstation 3100 processors only.¹²TF857 has a media loader, TF85 does not.**Table 5–9 Bus Type and I/O Rates for SDI Controllers**

Controller	Bus Type	No. of Ports	No. of SDI Channels	Transfer Rate Per second	Request Rate Per second
HSC40	CI	12	3	4 MB	1150 ²
HSC50 ¹	CI	24	6	4 MB	550
HSC60	CI	20	3	4 MB	1300 ²
HSC65	CI	20	3	4 MB	2000
HSC70	CI	32	8	4 MB	1150 ²
HSC90	CI	48	8	4 MB	1300 ²
HSC95	CI	48	8	4 MB	2000
KDA50	Q-bus	4	1	880 KB	100
KDB50	BI	4	1	1 MB	105
KDM70	XMI	8	8	3.4 MB	700

¹Requires HSC controller Version 4.1 software.²Requires HSC controller Version 6.0 software.

6

VAXcluster Configuration Rules and Guidelines

Digital specifies configuration rules based on supported features and on its experience with VAXcluster systems. General rules apply for all VAXcluster systems. There are also specific rules for certain configurations that are given in this chapter. Some guidelines are also included. All configuration rules conform to the VAXcluster Version 5.5 Software Product Description (SPD).

In the case of mixed-interconnect VAXcluster systems, all rules pertinent to each interconnect apply.

6.1 General VAXcluster Configuration Rules

The following rules apply for all VAXcluster configurations:

- The maximum number of VAXcluster nodes supported in a VAXcluster system is 96.
- The following Central Processing Units (CPUs) are not supported in any VAXcluster configuration:
 - VAXstation I
 - MicroVAX I
 - VAX-11/725
 - VAX-11/730
 - VAX-11/782
 - VAXstation 8000
- No VAXcluster node, storage controller, or storage device may participate in more than one VAXcluster system at a time.
- The *rule of total connectivity* must be met. This rule is that in a VAXcluster system, every VMS node must be able to communicate directly with every other VMS node. VAXcluster nodes do not perform routing for VAXcluster messages.

6.2 Configuration Rules for CI VAXcluster Systems

The following rules apply for a VAXcluster system using a single CI. See Section 6.2.1 for configuration rules regarding multiple CI adapters.

- The maximum number of VAXcluster nodes that can be connected to a single CI is 16.
- Dual porting of devices between an HSC and a local controller cannot be used if automatic failover is desired.

- TA-series tape drives may be dual ported between pairs of HSC subsystems with HSC Microcode Version 3.9 or higher.

6.2.1 Configuration Rules for CPUs with Multiple CI Connections

If a VAX CPU can support more than one CI adapter, then it can communicate on more than one CI path, through one or more Star Couplers. The following rules apply to multiple CI adapters in a VAXcluster system. In addition, you must follow the general rules for VAXcluster systems, as well as the rules for CI VAXcluster systems.

- Some VAX CPUs can have multiple CI adapters.
- Different types of CI adapters cannot be mixed in the same CPU.
- There is no requirement for a single Star Coupler through which all systems attached to the CI communicate, as long as total connectivity is maintained.
- A CPU cannot be attached to two Star Couplers that are not in the same VAXcluster system.
- A Star Coupler cannot be attached to two CPUs that are in different VAXcluster systems.
- Each CI adapter connected to the same Star Coupler must have a unique node number.
- Adapters in the same CPU may have the same node number if they are connected to different Star Couplers.
- Storage devices cannot be dual ported between HSC subsystems that are located on different Star Couplers.

Table 6-1 shows the numbers of CI adapters allowed in specific types of VAXcluster nodes. A single node cannot have different types of CI adapters. Information on CPUs that Digital no longer ships can be found in Table C-6.

Table 6-1 Maximum CI Adapters Per VAXcluster CPU

CPU	CI750	CI780	CIBCI	CIBCA-A	CIBCA-B	CIXCD
VAX 6000 family				1	4 ¹	4
VAX 9000 family						10 ²

¹There should be only one CIBCA-B per BI bus.

²There can be no more than four CIXCDs per XMI bus. This information supercedes the VAXcluster SPD CIXCD count for the VAX 9000 family.

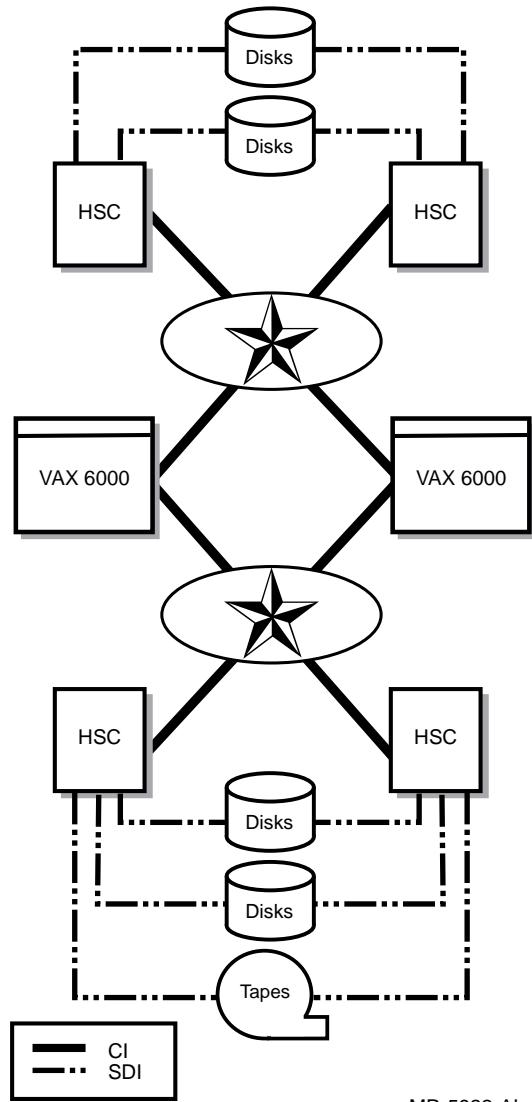
Figure 6-1 shows a multiple CI configuration.

6.2.2 Additional Guidelines for CI VAXcluster Systems

The following configuration guidelines are suggested for improved operation of your VAXcluster system:

- An Ethernet segment must be used for DECnet communication between all nodes connected to the CI.

Figure 6–1 Multi-CI VAXcluster System



MR-5922-AL

DECnet communications are required for VAXcluster operation. DECnet communications are automatically initiated on only one CI adapter (PAA0). DECnet communications are not supported over the CIXCD adapter.

- Each CI adapter in a single CPU does not have to be connected to a separate Star Coupler.

If additional CI adapters are installed for availability, connecting them to the same Star Coupler moves the single point of failure from the adapter to the Star Coupler. Since the Star Coupler is a passive device and is extremely reliable, this is often highly acceptable. Additional CI adapters and Star Couplers can be installed to improve system bandwidth.

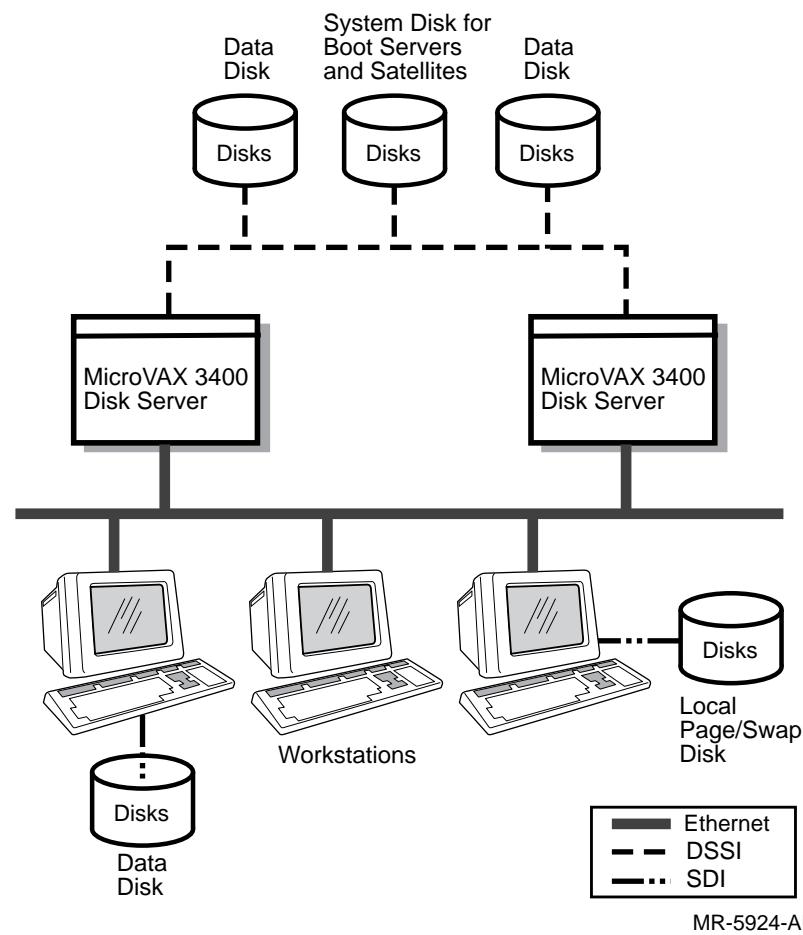
- Assign CI node numbers starting with zero, and continue numbering devices (CPUs and HSC subsystems) consecutively (without skipping numbers).

6.3 Configuration Rules for DSSI VAXcluster Systems

A DSSI VAXcluster node can access a common system disk and all data disks and tapes directly and serve them to satellites. Satellites (and users connected through terminal servers) can access any disk through either server. If one of the servers fails, applications running on satellites continue running, because disk access fails over to the other server.

In the dual-host configuration shown in Figure 6–2, the two servers and all satellites boot from a common system disk. If one server fails, the other server and satellites can still access both system and data disks.

Figure 6–2 Dual-Host MicroVAX 3400 Configuration



The following configuration rules apply to all DSSI VAXcluster configurations:

- All systems must be Q-bus VAX, MicroVAX, VAX 6000, or VAX 9000 systems.
- There can be no more than four VAXcluster systems on one DSSI bus, and no more than eight total nodes (including integrated storage elements (ISEs)).
- All systems connected to the same DSSI interconnect must be members of the same VAXcluster system.
- Each system must have Ethernet network hardware.

- Each DSSI interconnect must be terminated at both ends at all times, either by an adapter or by a DSSI terminator.
- Each node must have a unique DSSI ID number from 0 to 7 on each distinct DSSI segment (segments that do not connect to any of the same CPUs).
- If you use MSCP server nodes, the allocation class of the ISE must match the allocation class of the node that serves the ISE.
- A minimal ground offset voltage is required: 200 millivolts DC or 70 millivolts AC for up to 20 meters, 40 millivolts DC or 14 millivolts DC for 20 to 25 meters.

The following configuration rules apply to VAX 6000 DSSI VAXcluster systems:

- VMS Version 5.4-2 or later is required for VAX 6000 CPUs to support DSSI VAXcluster configurations. This version of VMS is also required for the KFMSA adapter and the TF857 tape ISEs.
- VMS Version 5.4-1 or later is required for RF72 disks.

6.3.1 Configuration Rules for CPUs with Multiple DSSI Connections

Multiple DSSI buses can be used between two CPUs to connect storage devices.

- Multiple DSSI adapters for each CPU are allowed (see Tables 6-2 and C-7).
- Different adapter types are allowed in the same CPU.
- Each DSSI adapter in a single CPU must be connected to a different DSSI segment.
- Up to three VAX 6000, VAX 4000, or VAX 3XXX CPUs are supported on any DSSI bus segment. Up to six KFMSA adapters are supported on each VAX 6000 CPU, for a maximum of 12 DSSI bus segments between two VAX 6000 CPUs.

Table 6-2 DSSI Adapters Per CPU

CPU	EDA640	SHAC	SWIFT	KFQSA	KFMSA
MicroVAX 3300/3400	1			2	
MicroVAX 3500/3600 /3800/3900				2	
VAX 4000-100/200		1		2	
VAX 4000-300/400/500 /600		2		2	
VAX 6000					6 ¹
VAX 9000					6 ¹
VAXft			4		

¹Each KFMSA contains two DSSI ports.

6.3.2 Additional Guidelines for DSSI VAXcluster Systems

For improved data availability, place your ISEs in different enclosures than the processors. If using VMS Volume Shadowing, shadow set members should reside in separate enclosures.

Including two VAXcluster nodes that can serve disks via MSCP in the DSSI VAXcluster configuration permits failover.

DSSI Node Numbering Guidelines

- DSSI bus IDs should be assigned as shown in Table 6–3:

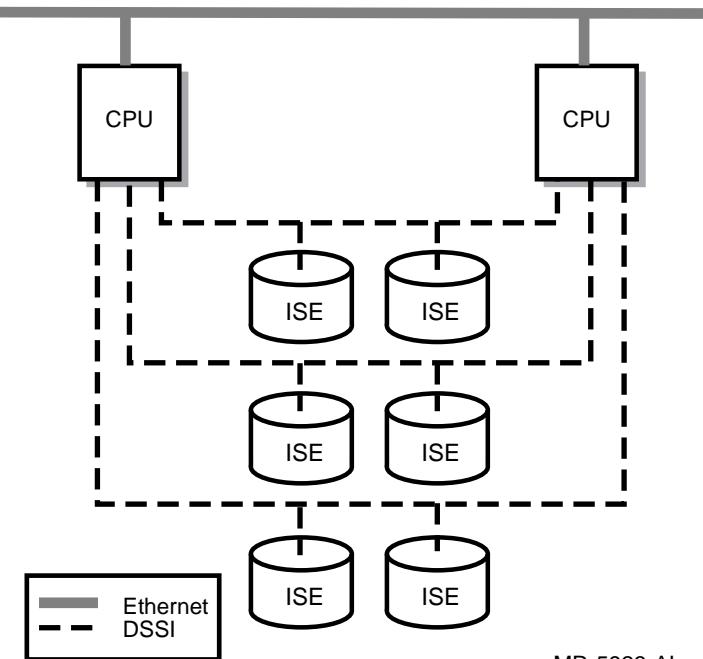
Table 6–3 DSSI Bus ID Assignments

Bus ID	Assignment
7	CPU host
6	CPU host ¹
5	CPU host ¹
4	CPU host ¹
3	RF and TF ISEs
2	RF and TF ISEs
1	RF and TF ISEs
0	RF and TF ISEs

¹Bus IDs 4, 5, and 6 can also be used for ISEs if they are not required for CPU hosts.

Figure 6–3 shows multiple DSSI buses connected between two CPUs. ISEs are uniquely numbered on each segment. The node numbers can be reused on different DSSI segments.

Figure 6–3 Multiple DSSI Segments in a VAXcluster System



6.4 Configuration Rules for Ethernet VAXcluster Systems

The following general rules apply for all Ethernet VAXcluster systems:

1. All VMS systems in a VAXcluster must be able to communicate with one another using DECnet. In a VAXcluster system, DECnet is commonly implemented on Ethernet and/or FDDI.
2. CPUs use the Ethernet for VAXcluster communications and may use it concurrently for other network protocols that conform to the applicable Ethernet standards, such as Ethernet Version 2.0, IEEE 802.2, and IEEE 802.3.
3. A low-latency data path providing approximately 10 megabits per second throughput must link all CPUs in a single VAXcluster system.
4. The extended LAN must be configured according to the guidelines in the *Telecommunications and Networks Buyer's Guide*.

6.5 Configuration Rules for FDDI VAXcluster Systems

Fiber Distributed Data Interface (FDDI) supports transfers using large packets (up to 4478 bytes). You can help maximize throughput in a VAXcluster that uses FDDI interconnects by setting the CPUs to permit the transfer of larger packet sizes.

- The target token rotation time (TTRT) in an FDDI token ring network must be set to 8 milliseconds, the default for Digital's FDDI product set.
- The ring latency when the FDDI ring is idle should be less than 400 microseconds.
- The total logical fiber-optic path length cannot exceed 200 kilometers (125 miles), which indicates a maximum physical ring circumference of 100 kilometers (62.5 miles). Nodes can be up to 40 kilometers (25 miles) apart.
- The maximum number of VAXcluster members that can connect to the FDDI via the DEMFA is 16.

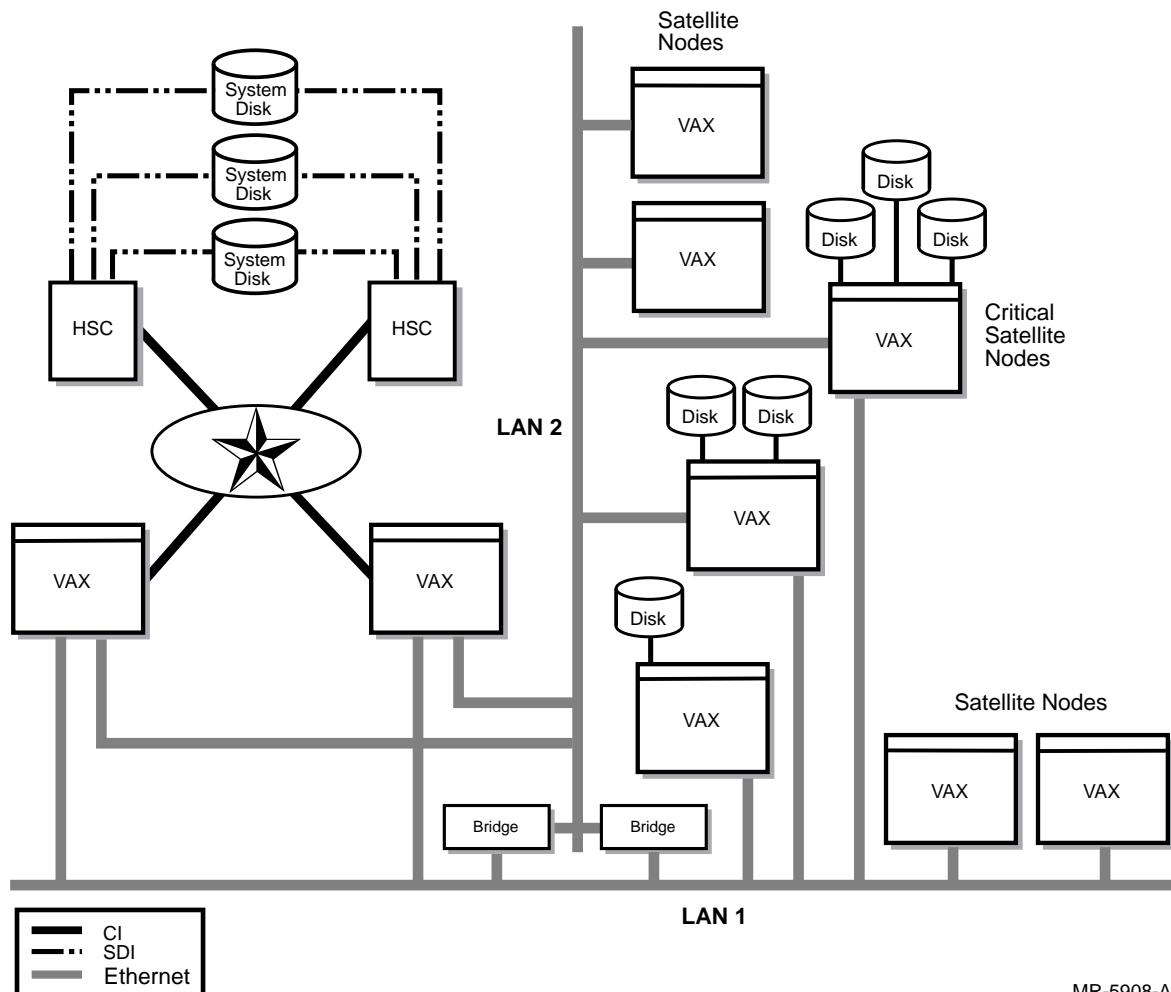
6.6 Configuration Rules for CPUs with Multiple LAN (Ethernet or FDDI) Adapters

Configurations for VAXcluster systems using multiple LAN adapters must meet the following requirements:

- Connect the MOP server and the system disk server for a given satellite to the same LAN segment.
- Provide a direct path from each node to all other nodes. A direct path can an unbridged LAN segment, or multiple bridged LAN segments.
- Distribute satellites equally among LAN segments. This helps distribute the VAXcluster load across all LAN segments.
- Distribute the LAN adapters that will downline load VMS to satellite CPUs among the LAN segments to ensure that LAN failures do not prevent satellite booting.

When configuring a VAXcluster using multiple LAN adapters and several segments, connect critical nodes to multiple segments or rings. This provides increased availability in the event of segment or adapter failure. Disk and tape servers can use some of the network bandwidth provided by the additional network connectivity. Critical satellites can boot using the other adapter if one adapter fails. Connecting disk servers to two or three LAN segments helps provide higher availability and better I/O throughput. Figure 6–4 is an example of a VAXcluster that uses multiple LANs.

Figure 6–4 Multiple LAN VAXcluster System



Optimizing VAXcluster System Design

The optimal VAXcluster system for any computing environment is based on trade-offs of cost, functionality, and performance. These trade-offs are influenced by the following factors:

- Applications in use
- Number and model of CPUs
- Growth potential
- Disk I/O capacity and access time
- Number of disks being served
- Interconnect (CI, DSSI, FDDI, or Ethernet) and adapter types
- Interconnect utilization

When evaluating your specific application dependencies and performance requirements, consider the guidelines and rules discussed in this chapter. Topics include the following:

- Increasing availability
- Guidelines for selecting disk servers and satellites
- Lock manager
- Backup strategy
- Configuring system disks
- Print services in your VAXcluster system
- Tools for managing your VAXcluster system

7.1 Increasing Availability

VAXcluster configurations achieve high availability through the following features:

- Redundancy of major system components
- Software support for failover between hardware components
- Provisions for maintaining a suitable environment for the VAXcluster hardware

Availability includes reliability, which indicates a component's ability to resist breaking. Recoverability, which indicates the ability to return to an unbroken state quickly, is also involved. Fault tolerance is a third part of availability; this is where a component or configuration can mimic full functionality, despite some loss of components or functionality.

For more information on availability, see Section 5.3.5, and *Building Dependable Systems: The VMS Approach*.

7.1.1 Hardware Redundancy Methods

You can configure a VAXcluster systems with no single point of failure. Redundancy is achieved by:

- Multiple CPU nodes — With more than one CPU node in a VAXcluster system, work can continue on the remaining nodes despite shutdown or failure of a CPU node.
 - Multiple paths to storage devices — Multiple paths can be provided to disk and tape drives with a dual-port or dual-host configuration, so that failure of a controller, adapter, or cable does not prevent access to a storage device.
- In addition, by using volume shadowing software, shadow set members can be distributed anywhere across a VAXcluster configuration. This helps avoid any disk being a single point of failure.
- Multiple interconnects — More than one interconnect can be connected to the nodes in a VAXcluster system, so a failure on one adapter or interconnect does not cause a loss of VAXcluster communications.

Each CI Star Coupler contains two separate paths, so failure in one path does not affect the other.

- Data redundancy — Multiple disk drives can be used with volume shadowing software to maintain multiple identical copies of data, so that loss of a disk drive does not cause loss of data.

Identical copies of data can also be kept on different disks.

Using multiple separate system disks in a VAXcluster system for additional performance and availability is one example of this technique.

- Site redundancy — Multiple datacenters or VAXcluster systems can be combined with the VAXcluster Multi-Datacenter Facility (MDF) to create disaster-tolerant configurations.
- To increase availability in a local area VAXcluster:
 - Bridge VAXcluster local area network (LAN) segments together to form a single extended LAN.
 - Provide redundant LAN segment bridges for failover support, preventing any one bridge from being a single point of failure.
 - Configure LAN bridges to pass the VAXcluster and satellite downline load requests. You can use the following in your configuration tasks:
 - * LAN bridge configuration documentation
 - * DEC Extended LAN Management Software (DECelms)
 - * Remote Bridge Management Software (RBMS)

7.1.2 Failover Mechanisms

The following VAXcluster mechanisms allow VAXcluster system operations to continue in spite of a failure in part of the VAXcluster system.

- DECnet VAXcluster alias — You can set up a DECnet *cluster alias* for some or all the nodes in the VAXcluster system. Connections directed to the cluster alias node name are distributed to the participating nodes in a round-robin fashion. If one of the nodes fails, new incoming networking connections are distributed among the CPUs that remain.

Network links that are operating when a CPU fails are terminated with an error status. Therefore, applications can be designed so that when they receive this error status they reinitiate a connection to the cluster alias node name and the new connection is made with one of the remaining nodes.

- VAXcluster LAT service — You can also set up a LAT *cluster service* to be offered by some or all the nodes in the VAXcluster system. This is done by setting up LAT service so that a service name is advertised from each of the participating VAXcluster nodes.

Interactive sessions are directed to the CPU with the lightest load in the set of those that are available for use.

Sessions on a failed node are terminated. The terminal server automatically makes a connection to one of the remaining nodes, but the user must log in again and restart the application.

- Generic batch queues — You can set up generic batch queues so that when a batch job is submitted, it is directed to and runs on one of the available CPUs. A job that is running on a CPU when it fails is terminated. If the batch job was submitted with the /RESTART qualifier, it is restarted on one of the remaining CPUs when startup occurs.
- Generic print queues — Similarly, when setting up a print queue to service a printer attached to a terminal server or to the network, you can create a generic print queue that directs print jobs to print queues on two or more nodes. These nodes all point to the same server port. The queue manager, which is part of the VMS job controller, directs jobs to a print queue on one of the systems that is operating. If a CPU fails while printing a job, the job is directed to another CPU to finish printing.
- Boot and file server failover — If two or more nodes are acting as MOP servers, satellite booting can occur despite shutdown or failure of one of these nodes. Likewise, if two or more nodes have disk serving enabled, satellite disk operations can continue despite failure of one of the disk servers. In both these cases, failover is automatic and no special action is needed.
- Multiple paths to storage devices — If there are multiple paths to disk drives and tape drives in a VAXcluster configuration, the VMS operating system automatically determines a working path and directs requests through it if the previous path it was using fails.
- Mixed interconnects — Mixed multiple interconnects can ensure there is no single point of interconnect communication failure in the VAXcluster system. VMS also frequently uses the fastest functional communication path to connect with other VAXcluster members.
- Volume shadowing — Volume shadowing software maintains duplicate copies of disk data so that when a disk fails, the data is still available to applications and systems that require it.

- MDF — To ensure that your entire datacenter continues to function even after a full site goes down, the VAXcluster Multi-Datacenter Facility permits applications to resume at a fully shadowed, fully functional remote site.

7.1.3 Environmental Protection

The one power cable that supplies the whole computer room can be a single point of failure for a VAXcluster system. Likewise, if the air-conditioning system fails, computer operations can halt because of the rise in temperature or humidity. The following steps can be taken to minimize the risk of environmental problems affecting VAXcluster system availability:

- A separate alternate utility feed line can be connected to provide protection from an outage on the primary connection.
- A generator or Uninterruptible Power System (UPS) may provide power to replace utility power for use during temporary outages.
- Extra air-conditioning equipment can be configured so that a failure of a single unit does not prevent use of the computer equipment.
- Battery backup for CPUs.

7.1.4 Quorum scheme

The availability of resources is a key feature of a VAXcluster system. An important aspect of VAXcluster availability is the synchronization of access to those resources. It is essential for successful VAXcluster operations that all members' access to disks, for example, be handled in a coordinated manner.

The Connection Manager uses an algorithm to prevent a VAXcluster system from partitioning into two or more independent VAXcluster systems that have uncoordinated access to the same set of shared resources. This strategy is based on the concept of a *quorum*, or the minimum number of voting members required to conduct business.

Typically, the main, important nodes in a VAXcluster system are each given one vote. Less important nodes and all satellite nodes normally have their VOTES set to 0; this way, they can enter and leave the VAXcluster system without affecting quorum.

A special case occurs when there are only two VAX systems with votes and they have an equal number of votes. Since the value of quorum is higher than the votes of either one alone, they must both be operational for the VAXcluster system to function.

There are two ways of dealing with this situation: using a *quorum disk* or a *quorum VAX processor*.

You may designate a shared disk to hold information that can be used to verify that only one set of systems is accessing the shared resources. This disk is called a quorum disk. In the simple case of two VAX systems, each with one vote, setting up a quorum disk and giving it one vote allows either of the systems to boot without the other and allows either one to continue operation after the other fails.

A quorum VAX cannot be a satellite and must boot from a directly accessible system disk. This VAX provides an additional vote to a two-node VAXcluster system.

The quorum disk can also be used to simplify VAXcluster system operation when it is possible that a certain number of VAX systems may be up or down at a time and yet VAXcluster system operation may continue. By weighting the votes given the VAXcluster nodes and quorum disk, you can ensure the most vital nodes in the VAXcluster can continue to operate even if other less vital nodes are down.

For more information on VAXcluster quorum, see *VMS VAXcluster Manual*.

7.1.5 VAXcluster State Transitions

State transitions occur when a computer joins or leaves a VAXcluster system. Connection Manager software controls these events to ensure the preservation of data integrity throughout the VAXcluster system. 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 shutdown of a current member. One of the remaining CPU nodes acts as coordinator and exchanges messages with all other VAXcluster members to determine the new configuration. Either transition typically occurs in a matter of seconds. Applications continue in the VAXcluster, but must be started or restarted on any node entering or reentering the configuration.

For more information on VAXcluster state transitions, see *VMS VAXcluster Manual*.

7.2 Guidelines for Selecting Disk Servers and Satellites

In selecting disk servers and satellites for VAXcluster systems, determine whether your configuration will provide reasonable support for anticipated disk-serving traffic. While there is no simple formula to make that determination, you can follow these general guidelines:

- A disk server is a VAXcluster member that serves a local disk to other members and, therefore, must actively handle that CPU's I/O requests for the disk.
- A disk client is a CPU that depends, for its normal operation, on disks served to it by a disk server. For example, an Ethernet satellite is always a client, because its system disk is served to it.

Consider each CPU as either a disk server or a service client. A configuration generally has substantially more clients than servers.

7.2.1 Disk Server I/O Capacity

The I/O capacity of disk-serving CPUs in VAXcluster configurations can affect satellite performance. The realized I/O capacity of a disk server is determined by its lowest capacity resource, whether CPU, memory, adapter, interconnect, or disk subsystem.

A special subset of disk servers are called Maintenance Operations Protocol (MOP) servers. MOP servers enable downline loading of VMS when satellites are booted.

The number of I/Os supported by a disk server's CPU, adapter, interconnect, and disk subsystem combination is the minimum of the number supported by any component. Sections 7.2.2 through 7.2.5 discuss disk server I/O capacity issues.

7.2.2 CPU I/O Capacity

Table 7–1 shows the I/O capacities of various VAX disk servers based on CPU model. The numbers in the table are for an average I/O size of 4 blocks, which is common for many applications. The numbers are also based on 80 percent CPU utilization, because unacceptable response times can occur at higher server utilizations.

The table also shows the I/O capacity of MicroVAX disk servers. While these servers are normally included in Ethernet VAXcluster systems, they can also serve their local disks in mixed-interconnect VAXcluster configurations. For VAX disk servers, the numbers apply whether the servers are configured in an Ethernet or mixed-interconnect VAXcluster system, because HSC disk serving is roughly equivalent to locally attached disk serving.

For information on CPUs no longer shipped by Digital, see Table C–14.

Table 7–1 Disk Server I/O Capacity Based on 80% CPU Utilization

CPU Type	Average (4-Block) I/Os Per Second
MicroVAX 3100	150
MicroVAX 3100–90	400
MicroVAX 3200	150
MicroVAX 3300/3400	130
MicroVAX 3500/3600	165
MicroVAX 3800/3900	165
VAX 4000–90	325
VAX 4000–100	400
VAX 4000–200	240
VAX 4000–300	325
VAX 4000–400	410
VAX 4000–500	500
VAX 4000–600	580
VAX 6000–210	165
VAX 6000–2x0 (x=2,3,4)	150
VAX 6000–310	234
VAX 6000–3x0 (x=2,3,4,5,6)	200
VAX 6000–410	410
VAX 6000–4x0 (x=2,3,4,5,6)	420
VAX 6000–510	850
VAX 6000–5x0 (x=2,3,4,5,6)	800
VAX 7000	800
VAX 9000–110	1700
VAX 9000–210	1700
VAX 9000–310	1700
VAX 9000–410	1700
VAX 9000–4x0 (x=2,3,4)	1600
VAX 10000	800

Note that symmetrical multiprocessing (SMP) disk servers (VAX 83X0, VAX 88XX, VAX 6000–2X0, 3X0, 4X0, and 5X0, and VAX 9000–4X0) process remote I/O requests on the primary CPU. Thus, an SMP system does not provide more disk serving I/O capacity than the equivalent uniprocessor system. Thus, an SMP system does not provide more disk serving I/O capacity than the equivalent uniprocessor system. SMP systems can provide additional computing power for local activity, because the secondary processors can be used for local interactive

and batch processes. Performance problems should not occur unless the server is saturated with remote disk I/Os and there is also a very heavy local I/O load.

7.2.3 Ethernet I/O Capacity

The I/O capacity of a disk server's Ethernet adapter can limit overall I/O performance of satellite nodes. Table 7-2 shows the number of I/Os per second supported for the average I/O size of 4 blocks. Compare these numbers with those shown in Table 7-1 for CPU capacity to size disk servers and evaluate potential disk server bottlenecks. If the disk server uses multiple Ethernet adapters, their I/O capacity is summed for a greater total throughput.

Table 7-2 Ethernet Adapter I/O Capacity

Ethernet Adapter	Average (4-Block) I/Os Per Second
DEUNA	45
DELUA	100
DELQA	120
DEQTA	120
DESQA	120
DEBNA	135
DEBNI	340
DESVA	130
DEMNA	400
SGEC	240

7.2.4 Disk Drive I/O Capacity

The limiting factor in VAXcluster system I/O performance can often be an individual disk drive. For a disk server that serves only its locally attached disks, the number of disks that can be configured on the disk controller and the disk controller itself can also be limiting factors. HSC disks can also be served by multiple VAX nodes.

A mixed-interconnect VAXcluster system can include many powerful CPUs acting as disk servers for satellites and also supporting local processing activity. Because these CPUs access common HSC or RF disks, the potential for overloading a single disk is greater than in an Ethernet VAXcluster system. Disk drives are the first place to look for potential clusterwide bottlenecks. In configuring a system and designing applications, it is important not to overload a single drive.

7.2.5 Summary of Disk Server I/O Capacity

In Ethernet VAXcluster configurations, overall disk server I/O capacity is determined by the minimum of the CPU or Ethernet adapter capacity.

Table 7-3 summarizes disk server I/O capacity and identifies limiting resources. Note that the I/O capacity of a MicroVAX disk server is determined not only by its CPU and Ethernet adapter, but also by its disk subsystem — the disk controller and number of shared disks. The results assume that there are enough disks to meet the demand and that the work load is well balanced over all the disks. For information on disk servers that Digital no longer ships, please see Table C-3.

Table 7–3 Disk Server Capacity — Average (4-Block) I/O Operations Per Second

CPU Type	Ethernet Adapter Type	Throughput ¹	Limiting Resource
VAX 8500, VAX 8530, VAX 8550, VAX 8800, VAX 8700, VAX 88x0	DEBNI or DEBNA	340 (DEBNI) 135 (DEBNA)	DEBNI (for VAX 88xx) DEBNA (for all others)
VAX 6000-2x0, VAX 6000-3x0, VAX 6000-4x0	DEBNI or DEBNA or DEMNA	340 (DEBNI) 135 (DEBNA) 400 (DEMNA)	DEMNA
VAX 9000-110, VAX 9000-210, VAX 9000-310, VAX 9000-4x0	DEMNA	400	DEMNA
MicroVAX 3100 MicroVAX 3100e	DESVA	130	DESVA
MicroVAX 3500	DELQA	120	DELQA
MicroVAX 3600	DELQA	120	DELQA, 4 RA82s
MicroVAX 3300/3400	On CPU module	120	CPU
MicroVAX 3800	DELQA DESQA	120	CPU
MicroVAX 3900	DELQA DESQA	165	CPU
VAX 4000-100	SGEC	400 (SGEC)	CPU
VAX 4000-200	SGEC	200 (SGEC)	CPU
VAX 4000-300	SGEC	325 (SGEC)	CPU
VAX 4000-400	SGEC	400 (SGEC)	CPU

¹4-block I/Os per second.

7.3 Lock Manager

Resources in a VAXcluster are utilized by a single entity in a VAXcluster at any point in time. This resource management is handled by locks which are granted on any given resource to help coordinate access and use of resources within a VAXcluster system. VMS software is designed to minimize the overhead of maintaining the distributed locks in a VAXcluster system. A resource master node keeps track of the locks that were granted. Each node also stores information on its own locks. Thus, access and processes that require resources can continue unhindered.

7.4 Backup Strategy

The Backup Utility (BACKUP) includes a method of scanning files on the input disk that results in rapid save and copy operations. Scanning does not improve BACKUP's performance during restore, compare, verify, or list operations, however.

Section 7.4.1 lists qualities of HSC BACKUP and VMS BACKUP. You can select which to use based on your requirements.

7.4.1 Using HSC BACKUP or VMS BACKUP

If a VAXcluster configuration contains HSC subsystems, disk backups can be done using VMS or HSC BACKUP. Table 7–4 describes the main advantages of both, and the differences between them.

Table 7–4 HSC BACKUP Versus VMS BACKUP

HSC BACKUP	VMS BACKUP
Operation may take less time. The timing is more predictable, which may ease backup scheduling.	Noncritical files can be marked /NOBACKUP and are not stored during backup operations.
May keep a streaming tape drive in streaming mode more easily, particularly with small files.	If the disk is partially full, saving a disk may take fewer tapes than to store the entire drive's contents
Data goes directly from disk to tape, rather than crossing the CI	Data crosses the CI twice: once from disk to memory and again from memory to tape.
Up to two backup operations can be underway at once.	No limit to the number of simultaneous backup operations.
Only the entire backup can be restored.	Individual files can be retrieved.
Requires exclusive access to the disk drive.	Does not require exclusive access to the disk drive.
Restoration does not result in contiguous files.	Restoration results in contiguous files.
A missing tape means a missing range of logical blocks, so pieces of files may be missing, rather than whole files.	If a tape is missing, files not on that tape can be retrieved from the remaining backup tapes.
Cannot do incremental backups.	Can do incremental backups.

7.5 Configuring System Disks

The system disk contains the VMS operating system, common utilities, and libraries. When any CPU is turned on or rebooted, it automatically boots from its designated system disk.

In large VAXcluster systems, a system disk normally has multiple access paths between two HSC subsystems for availability. If a single HSC subsystem or CI component fails, the disk remains accessible through the other path. System disk availability can be further enhanced with VMS Volume Shadowing software.

In *common-environment* VAXcluster systems, where all users require the same services and applications, you can generally use a single common system disk. Less time is required to maintain and upgrade the software on a single system disk than on multiple system disks. In some very large configurations, where many nodes typically need to access the system disk simultaneously, you may decide to have more than one system disk, even while maintaining a common environment. However, be aware of the increased system management work involved in maintaining multiple system disks.

In *multiple-environment* VAXcluster systems, where different users require different services and applications, you can plan a system disk for each environment. Again, multiple system disks involve increased system management.

If the I/O activity to the system disk is straining the disk's capabilities, it is a good idea to place certain potentially heavily accessed system files, such as paging and swapping files, on other less active disks. This arrangement can produce performance rewards by removing the I/O activity of paging and swapping from the system disk and placing it on another disk that can easily accommodate them.

In most environments, if heavy paging, and especially if heavy swapping is occurring, then SYSGEN parameters likely need to be reset or additional memory needs to be added to the system to optimize performance. Where high performance is a goal, memory and SYSGEN parameters should be set to eliminate swapping. DECperformance Solution (DECps) software can determine the paging and swapping rates and recommend changes.

7.5.1 Booting Activity

All VAXcluster workstations are simultaneously active during a VAXcluster system reboot (for example, after a power failure). All satellites are waiting to reload, and, as soon as a boot server is available, they begin to boot in parallel. Because this booting activity places a significant I/O load on the system disks, system disk I/O capacity can be the limiting factor for VAXcluster system reboot time.

Note that you can reduce overall VAXcluster system boot time by configuring multiple system disks and distributing system roots for VAXcluster nodes evenly across those disks. For VAXcluster systems with substantial system disk I/O requirements, you can use multiple system disks, each configured as a shadow set. Multiple MOP servers can also help reduce the system boot time.

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

- Cost of having workstations unavailable during a VAXcluster system 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

7.6 Print Services in Your VAXcluster System

An important consideration in designing your VAXcluster configuration is providing printing services for your users. Table 7-5 lists some of the printers available from Digital. Printer speed, connectability, and ease of normal maintenance can help you decide which printers most closely suit your requirements. Table C-15 lists information about older printers that Digital no longer ships.

Table 7–5 Selected Digital Printers

Model	Speed	Type	Interface	Protocol	Consumables Replacement	Volume
LA75 ³	250 cps ⁴	Dot matrix	Serial	ASCII	User	Specialized /low
LG31	300 lpm ⁶	Dot matrix	Serial	Sixel	User	Low
LP29	2000 lpm	Band	Parallel	ASCII	Operator	High
LPS40	40 ppm	Laser	Ethernet	PostScript	User	Medium
LPS20	20 ppm	Laser	Ethernet	PostScript	User	Medium
LJ250/ LJ252	167 cps	Ink jet	Serial	Sixel	User	Medium
LG01/ LG02	600 lpm	Dot matrix	Parallel or serial	Sixel	Operator	Medium

³LA75 Companion Printer.

⁴Characters per second.

⁶Lines per minute.

You can connect printers locally to a terminal controller, such as a DHQ11, or distribute them by connecting each printer to a DECserver on the Ethernet or, for an LPS40, directly to the Ethernet. Connecting printers to terminal servers or to the Ethernet is the preferred method because it allows sharing over the entire network.

7.7 Tools for Managing your VAXcluster System

Various tools are available to help you manage and maintain your VAXcluster system. Tools and utilities such as the CLUSTER_CONFIG.COM command procedure, the Show Cluster Utility (SHOW CLUSTER), and the System Management Utility (SYSMAN) are components of the VMS operating system. Other tools (see Section 7.7.5) are optionally available products.

7.7.1 CLUSTER_CONFIG.COM Command Procedure

With the VAXcluster configuration command procedure CLUSTER_CONFIG.COM, you can configure or reconfigure a VAXcluster system easily without invoking VMS utilities directly. You use the command procedure to perform the following functions:

- Add a CPU to the VAXcluster system
- Remove a CPU from the VAXcluster system
- Change a VAXcluster CPU's characteristics
- Create a duplicate system disk

For detailed information on the CLUSTER_CONFIG.COM command procedure, see the *VMS VAXcluster Manual*.

7.7.2 Show Cluster Utility

SHOW CLUSTER monitors nodes in a VAXcluster configuration and displays information about VAXcluster activity and performance. SHOW CLUSTER collects information from the system communication services (SCS) database, the Connection Manager database, and the port database. SHOW CLUSTER outputs the information to your terminal or to a specified device or file. You can use SHOW CLUSTER interactively or with command procedures and user-defined default settings. See the *VMS Show Cluster Utility Manual* for more information on this utility.

7.7.3 System Management Utility

SYSMAN centralizes the management of nodes and VAXcluster systems. Rather than logging into different nodes and repeating a set of management tasks, SYSMAN lets you define your management environment to be a particular VAXcluster system or node. With a defined management environment, you can perform system management tasks from your local node; SYSMAN executes the tasks on all selected nodes in the target environment.

SYSMAN uses standard VMS software procedures. It accepts DIGITAL Command Language (DCL) commands, such as SET, SHOW, MOUNT, DEFINE, and INITIALIZE. SYSMAN can also execute VMS system management utilities and command procedures, such as the Authorize Utility (AUTHORIZE), Install Utility (INSTALL), and Automatic Generation Utility (AUTOGEN). SYSMAN contains the DISKQUOTA command set and the parameter setting functions of SYSGEN. For more information on SYSMAN, refer to the *VMS SYSMAN Utility Manual*.

7.7.4 VMS Local Area VAXcluster Failure Analysis

VMS provides Network Failure Analysis, a subsystem that can help detect and isolate a failed network component. Local Area VAXcluster Network Failure Analysis provides notification of problems in the communication paths by directing OPCOM messages to CENTRAL, CLUSTER, DEVICE, and NETWORK destinations.

Each LAN adapter used for VAXcluster communications communicates regularly with remote cluster nodes. If this regular communication is not received within eight seconds, the channel is considered to have failed and is closed. If you have enabled Network Failure Analysis, it groups together channels that fail in this manner for analysis. Network Failure Analysis then looks for a common failure or a group of unrelated failures.

Local Area VAXcluster Network Failure Analysis takes as input a specific description of the physical network that is used for local area VAXcluster communications. Once the description is loaded, the failure analysis can be enabled. The Network Failure Analysis program collects future channel failures into groups and maps them onto the physical description to obtain a list of non-working network components related to the bad channels. Then the Network Failure Analysis displays the components in the list that have a probability of causing one or more channel failures.

7.7.5 Optional Tools and Products

This section lists several optional software tools and products you can obtain for your VAXcluster system.

- Data Center Monitor (DCM)

DCM automatically scans a set of specified nodes to detect selected events or problems. It then triggers corrective action or forwards information to another application for corrective action.

- DECamds

DECamds is a system management tool that lets you monitor, identify, and resolve areas of resource denial in a LAN. It monitors, investigates, diagnoses, and lets you correct system resource utilization problems for CPU usage, low memory, lock contention, hung or runaway processes, I/O, disks, paging files, and swapping files.

- DECalert

DECalert alerts key system management personnel of conditions that require attention. It can assist in implementing a lights-out computing environment.

DECalert monitors input from many sensing sources, such as DECmcc, VAXsimPLUS, VAXcluster Console System (VCS), or user-written sensors, and generates and distributes alarms when necessary. This notification can be made using pagers, electronic mail, telephone calls using DECTalk, or graphic displays.

- DECperformance Solution (DECps)

DECps is a tool that can help you evaluate and optimize current system resources in a VAXcluster system. It offers automated analysis and dynamic modeling manipulation. A data collector feeds information into three separate functions: a performance advisor (formerly VPA), a capacity planner, and accounting chargeback.

- DECscheduler

DECscheduler lets you automate scheduling, execution, and monitoring of repetitive tasks, such as file maintenance, backups, and production application jobs. DECscheduler features include scheduling with job and time dependencies, VAXcluster system load balancing, and VAXcluster system failover assistance. It has a DECwindows interface for job monitoring and control in addition to a menu-driven interface for character-cell terminals.

- LAN Traffic Monitor (LTM)

LTM displays real-time data about Ethernet LAN throughput and utilization to monitor Ethernet usage and collects data on multivendor protocols.

- VAXcluster Multi-Datacenter Facility (MDF)

The VAXcluster MDF package combines VAXcluster technology with software, services, training, and licensing to support automatic failover and predictable recovery in the event of a disaster. You can manage multiple sites from multiple locations, and consolidate two separate datacenters or VAXcluster systems into one single VAXcluster that can be managed from any member site.

- **VAX Remote Environmental Monitoring Software (REMS)**

REMS is a hardware and software package that is used to monitor the physical environment, including temperature, humidity, or the presence of water on the floor, of a datacenter. It can relay signals, send mail, and spawn preprogrammed subroutines automatically, based on input from the monitoring probes.

- **Remote System Manager (RSM)**

RSM lets you manage several computer systems connected with the DECnet network. RSM helps automate recurring system management tasks, such as distributing software, tracking software configurations, performing file backup and restore, and system administration.

- **VAXcluster Console System (VCS)**

VCS is a layered product that provides a central location for coordinating and managing up to 24 system console lines. These console lines are connected to either VAX or HSC console ports. With VCS, you can monitor current console data from VAXcluster nodes, output data to VAXcluster nodes, explore console data in any order, and review historical data from one or more nodes. You can also connect VCS to nodes that are not in a VAXcluster system.

- **VAX Distributed File Service (DFS)**

DFS provides VMS users with the ability to use remote VMS disks as if they were directly attached to their own VAXcluster system. DFS provides users and applications with transparent, high-performance file-read access while using fewer CPU resources than standard DECnet-VAX file access. The use and management of DFS is similar to the use and management of local disks. You can make directory structures available to other DFS nodes.

- **VAX Distributed Queuing Service (DQS)**

DQS uses the DECnet network to extend the standard VMS queue system to enable you to print jobs on printers connected to systems other than your own. On those other systems, you can show the status of your jobs, cancel your jobs, or change job specifications. These capabilities enable multiple VMS systems to share unique printers.

- **VAX Software Performance Monitor (SPM)**

SPM is a software performance management tool that collects, displays, reports, and graphs performance information useful in system tuning and capacity planning. The information includes resource utilization and load balance data. Data can be collected using a variety of user-specified parameters. You can start and stop data collection for all nodes in a VAXcluster system from a single terminal and store all performance data in a single file. Performance data is multikeyed for rapid retrieval by node name, disk name, and other access keys. Reports contain detail helpful to quantify system resource utilization (CPU, memory, and I/O) and identify system bottlenecks. Analysis of this data can reveal underutilized resources that can be used to alleviate a bottleneck.

- **VAX Storage Library System (SLS)**

SLS is a set of software tools that gives users the ability to manage collections of removable media, including magnetic tape, cartridge tape, and optical disks. Using SLS, you maintain a record of all information on a backup or archived media and retrieve the information quickly.

- DECmcc

DECmcc is a layered product that provides central management control and monitoring of DECnet nodes and network devices. These devices include LAN bridges, DECbridges, and DECconcentrators. DECmcc can also incorporate additional management modules to assist in managing other Digital and third-party devices.

A

SPD Disk Storage Requirements

Table A-1 is provided to help you estimate the amount of storage you need on your system disk to hold system or application software and any database the software uses. Note that the amount of storage shown is correct only for the software version listed; later versions can require a different amount of storage. (To determine the amount of storage needed for later versions, see the current Software Product Description (SPD) for that version; the current SPD also contains the part numbers you need for ordering.)

Note

Consider the numbers in these tables as guidelines only; they are provided only for your convenience, to help you in estimating your first pass storage requirements. In general, you need the amount of storage shown in the tables only once per system disk.

Table A-1 Space Required on VMS System Disk

Software Product	SPD Number	Blocks on System Disk for:	
		Installation	Operation
VMS operating system Version 5.5	25.01.35	82,000 ¹	82,000 ¹
VMS Workstation Software Version 4.3 ²	28.06.11	325 to 30,600	5,050 to 30,600
ALL-IN-1 Version 3.0	27.30.06	75,000	65,200
CDD/Repository for VMS, Version 5.0	25.53.20	33,000	28,500
DEC GKS for VMS Version 4.2	26.20.10	11,000/6,500	10,000/5,900
DEC PrintServer Supporting Host Software for VMS Version 4.0	27.68.06	9,800	9,500
DEC RALLY for VMS Version 3.0	27.03.08	20,100/6,000	18,100/3,000
DEC VTX Version 5.0	26.57.13	10,800/21,5000/ 41,000	4,700/8,700/ 22,000
DECintact Version 2.0	29.58.04	52,000	46,000
DECnet Router Server Version 1.2	30.34.06	2,200	3,224
DECnet-VAX Version 5.5 ⁴	25.03.30	2,400	2,400
DECnet/PCSA Client VAXmate Version 2.2	55.10.03	23,000	

¹Includes paging and swapping files; plus an additional 55,000 blocks with DECwindows; about 137,000 blocks for VMS Version 5.4.

²Required for VAXcluster VAXstation satellite.

⁴Required for VAXcluster System.

(continued on next page)

Table A-1 (Cont.) Space Required on VMS System Disk

Software Product	SPD Number	Blocks on System Disk for:	
		Installation	Operation
DECnet/SNA Data Transfer Facility Version 3.1	27.85.04	5,000 ³	2,950 ³
DECnet/SNA VMS APPC/LU6.2 Programming Interface Version 2.2	26.88.06	4,400	4,400
DECnet/SNA VMS Application Programming Interface Version 2.3	26.86.04	4,700	4,600
DECnet/SNA VMS DISOSS Document Exchange Facility Version 1.4	26.72.05	3,696	1,306
DECnet/SNA VMS Distributed Host Command Facility Version 1.2	26.71.03	650	450
DECnet/SNA VMS Printer Emulator Version 1.2	26.70.05	450	400
DECnet/SNA VMS Remote Job Entry Version 1.4	26.85.04	2,200	1,000
DECnet/SNA VMS 3270 Data Stream Programming Interface Version 1.4	26.87.05	1,300	1,100
DECnet/SNA VMS 3270 Terminal Emulator Version 1.5	26.84.06	450	300
DECpage Version 3.1	26.29.09	40,000	35,000
DECrouter 200 Version 1.1	27.72.02	1,320	600
DECserver 100 for VMS and MicroVMS Donwline Load Hosts Version 2.0	27.41.02	600	600
DECserver 200 for VMS and MicroVMS Version 3.1	27.53.05	900	764
DECtalk Mail Access Version 1.1	26.45.03	5,500	500
EDCS II Version 2.1	26.39.04	30,000/15,000	10,500/5,000
Ethernet Terminal Server for VMS and MicroVMS Version 3.0	27.39.02	2,300	2,300
External Document Exchange with IBM DISOSS Version 2.1	26.92.03	6,300	1,565/1,350
IEX-VMS-DRIVER Version 4.2C	26.30.08	1,300	1,000
LAN Traffic Monitor VMS Version 1.2	27.80.02	1,100	600
NMCC/DECnet Monitor Version 2.3	26.91.05	20,000	to 20,000
NMCC/VAX ETHERnim Version 2.3	26.96.05	4,000	5,500
Remote Bridge Management Software (RBMS) Version 2.0	27.12.03	1,800	1,600
Remote System Manager Version 2.3	29.59.03	28,000	19,000
Session Support Utility Version 1.3	27.79.02	400	400
Terminal Server Manager Version 1.5	27.64.06	3,700	1,700
VAX 2780/3780 Protocol Emulator Version 1.7	25.07.15	650	600
VAX 3271 Protocol Emulator Version 2.5	25.21.13	1,150	750

³Requires up to 50 tracks on IBM system (in IBM 3380 tracks).

(continued on next page)

Table A-1 (Cont.) Space Required on VMS System Disk

Software Product	SPD Number	Blocks on System Disk for:	
		Installation	Operation
VAX ACMS Version 3.2	25.50.10	to 46,000	to 39,800
VAX Ada Version 2.2	26.60.10	38,000	37,000
VAX ADE Version 2.5	25.76.09	2,200	2,000
VAX APL Version 4.0	25.31.11	7,000	to 5,300
VAX BASIC Version 3.5	25.36.22	6,800	6,600
VAX BLISS-32 Implementation Language Version 4.6	25.12.18	4,200	3,600
VAX C Version 3.2	25.38.18	12,500	9,200
VAX COBOL Version 4.4	25.04.23	7,000	4,000
VAX COBOL GENERATOR Version 1.3	27.16.04	5,500	4,600
VAX Data Distributor Version 2.2	27.76.05	6,000	3,000
VAX DATATRIEVE Version 5.1	25.44.23	17,000	12,000
VAX DBMS Version 4.3	25.48.22	to 29,500	to 8,000
VAX DECalc Version 4.0	25.79.12	4,000	4,000
VAX DECalc/DECgraph Package Version 4.0	27.51.06	4,000/5,000	4,000/1,600
VAX DECalc-PLUS Version 4.0	27.37.05	6,000	6,000
VAX DECgraph Version 1.6	26.07.11	3,500	1,600
VAX DECprom Version 1.1	26.49.03	1,128	393
VAX DECrad Version 4.0	25.77.08	195,000	300,000
VAX DECscan VMS Software Toolkit Version 2.1	26.98.02	6,600	3,300
VAX DECslide Version 1.4	26.11.09	3,500	1,000
VAX DECspell Verifier/Corrector Version 1.1	26.34.04	3,000/1,100	1,800/1,000
VAX DEC/CMS Version 3.4	25.52.16	7,000	3,900
VAX DEC/MMS Version 2.6	26.03.14	2,300	800
VAX DEC/Shell Version 2.2	26.69.08	4,400	3,800
VAX DEC/Test Manager Version 3.2	26.68.10	8,500	6,000
VAX DIBOL Version 4.2	25.49.15	7,500	6,500
VAX Distributed File Service Version 1.2	28.78.02	1,800	1,600
VAX Distributed Queuing Service Version 1.2	28.80.02	1,400	700
VAX DOCUMENT Version 2.1	27.55.06	to 40,000	to 30,000
VAX DOCUMENT/LN03 Font Package Version 1.1	28.84.02	32,500	16,250
VAX DSM Version 6.0	25.08.19	30,000	27,000
VAX DT07 Version 3.0	25.88.05	662	272
VAX Encryption Version 1.2	26.74.03	3,600	2,100

(continued on next page)

Table A-1 (Cont.) Space Required on VMS System Disk

Software Product	SPD Number	Blocks on System Disk for:	
		Installation	Operation
VAX FMS Version 2.4	26.10.12	3,000/1,850/350	2,650/1,300/300
VAX FORTRAN Version 5.7	25.16.38	9,000	4,900
VAX KMS11-BD/BE HDLC/BSC Framing Software Version 2.1	26.55.05	555	288
VAX KMS11-BD/BE X.25 Link Level Software Version 2.1	25.80.07	582	270
VAX Language-Sensitive Editor/Source Code Analyzer Version 3.1	26.59.12	10,000/13,025	6,700/9,725
VAX LIMS/SM Laboratory Information Version 1.5	26.18.06	60,000	to 23,600
VAX LISP/VMS Version 3.1	25.82.09	to 93,627	to 92,985
VAX MAILGATE for MCI Mail Version 2.0	27.34.01	7,000	4,000
VAX Notes Version 2.2	27.06.06	4,800	3,800
VAX OPS5 Version 3.0	27.04.05	4,550	2,450
VAX OSI Application Kernel Version 1.1	27.47.01	4,000	4,000
VAX Packet System Interface (P.S.I.) Version 4.3	25.40.15	6,000	8,000
VAX Packetnet System Interface (P.S.I.) Access Version 4.3	27.78.02	4,000	2,500
VAX Pascal Version 4.3	25.11.32	9,500/7,000	3,000/6,000
VAX Performance Advisor Version 2.1	27.71.04	14,307	8,133
VAX Performance and Coverage Analyzer Version 3.0	26.76.07	11,000	5,500
VAX PL/I Version 3.4	25.30.19	7,000	6,500
VAX PrintServer Client Software Version 3.1	27.67.04	3,500	2,700
VAX Public Access Communications Version 1.3	28.51.03	2,400	1,000
VAX Rdb/ELN Version 2.3	28.03.09	16,062	8,600
VAX Rdb/VMS Version 4.0	25.59.13	to 45,000	to 40,000
VAX RMS Journaling Version 5.5	27.58.06	—	—
VAX RPG II Version 2.1	26.05.06	2,610	1,200
VAX SCAN Version 1.2	26.93.03	4,400	1,600
VAX ScriptPrinter Software Version 2.1	27.84.03	5,500	2,800
VAX Software Performance Monitor Version 3.4	27.56.09	18,000	11,000
VAX Software Project Manager Version 1.2	27.52.03	5,300	5,300
VAX Source Code Analyzer Version 2.0	27.63.05	7,500	5,100
VAX SQL Version 1.1	27.70.01	16,970/6,200/ 5,400	11,270/4,000/ 2,800

(continued on next page)

Table A–1 (Cont.) Space Required on VMS System Disk

Software Product	SPD Number	Blocks on System Disk for:	
		Installation	Operation
VAX Storage Library System Version 2.1	29.67.04	4,100/3,800	3,500/2,900
VAX TDMS Version 1.9	25.71.15	12,000/2,500	5,000/3,000
VAX TEAMDATA Version 1.4	27.02.04	13,700	12,900
VAX VALU Version 2.1	26.94.03	3,000	1,600
VAX Xway Version 1.2	27.36.03	2,200	1,800
VAXcluster Console System Version 1.3	27.46.04	7,000	3,500
VAXcluster Software Version 5.5	29.78.05 ⁵	TBD ⁵	TBD ⁵
VAXELN Toolkit Version 4.3	28.02.13	42,000/72,000	34,000/68,000
VAXinfo I Version 1.4	27.19.04	67,375	48,750
VAXinfo II Version 1.4	27.20.04	100,550/72,600	73,425/53,755
VAXinfo III Version 1.4	27.21.04	113,250/84,500/ 13,250	76,225/57,175/ 7,775
VAXset Package Release 10	27.07.09	33,700/36,725	28,200/31,225
VIDA with IDMS/R Version 2.2	27.25.04	3,000	1,900
VMS/SNA Version 2.1	27.01.06	3,400	2,300
VMS Volume Shadowing Version 5.5	27.29.08	5	5
WPS–PLUS for VMS Version 4.1	26.27.11	27,000	17,500

⁵Part of the VMS operating system. Storage requirements included with VMS system blocks.

Most software products require additional space for user data, source and compiled code, database and various internal reasons. In many cases, this information cannot be readily characterized or tabulated.

B

VAXcluster Software Product Description

This is the Software Product Description (SPD) for Version 5.5 of the VAXcluster Software. The official SPD number is 29.78.05.

B.1 Description

VAXcluster Software is a VMS System Integrated Product (SIP). It provides a highly integrated VMS computing environment distributed over multiple VAX, VAX Workstation, and MicroVAX CPUs. This environment is called a VAXcluster system.

CPUs in a VAXcluster system can share processing, mass storage, and other resources under a single VMS security and management domain. Within this highly integrated environment, CPUs retain their independence because they use local, memory-resident copies of the VMS Operating System. Thus, VAXcluster CPUs can boot and shut down independently while benefiting from common resources.

Applications running on one or more CPUs in a VAXcluster system access shared resources in a coordinated manner. VAXcluster Software components synchronize access to shared resources, preventing multiple processes on any CPU in the VAXcluster from interfering with each other when updating data. This coordination ensures data integrity during multiple concurrent update transactions.

Because resources are shared, VAXcluster systems offer higher availability than standalone CPUs. Properly configured VAXcluster systems can withstand the shut down or failure of various components. For example, if one CPU in a VAXcluster is shut down, users can log on to another CPU to create a new process and continue working; since mass storage is shared VAXcluster-wide, the new process is able to access the original data. Applications can be designed to survive these events automatically.

All VAXcluster systems have the following software features in common:

- Shared file system — The VMS Operating System and VAXcluster software allows all CPUs to share read and write access to disk files in a fully coordinated environment. Application programs can specify the level of VAXcluster-wide file sharing that is required; access is then coordinated by the VMS Extended QIO Processor (XQP) and Record Management Services (RMS).
- Shared batch and print queues are accessible from any CPU in the VAXcluster system. The VMS queue manager controls VAXcluster-wide batch and print queues, which can be accessed by any CPU. Batch jobs submitted to VAXcluster-wide queues are routed to any available CPU so the batch load is shared.

- The VMS Lock Manager System Services operate in a VAXcluster-wide manner. These services enable reliable coordinated access to any resource and provide signaling mechanisms, at the system and process level, across the whole VAXcluster system.
- All disks and TMSCP tapes in a VAXcluster system can be made accessible to all CPUs.
- Process information and control services are available VAXcluster-wide to application programs and system utilities.
- An automated configuration command procedure assists in adding and removing CPUs and in modifying their configuration characteristics.
- The dynamic Show Cluster Utility displays the status of VAXcluster hardware components and communication links.
- Standard VMS system and security features work in a VAXcluster-wide manner such that the entire VAXcluster system operates as a single security domain.
- The VAXcluster software dynamically balances the interconnect I/O load in VAXcluster configurations that include multiple interconnects.
- Multiple VAXcluster systems can be configured on a single Local Area Network (LAN).

Definitions:

The following terms are used throughout this SPD:

- CPU (Central Processing Unit) — A VAX-family computer that includes one or more processors. A CPU operates as a VAXcluster node. A VAXcluster node can be referred to as VAXcluster member.
- Disk server — A CPU that makes disks to which it has direct access available to other CPUs in the VAXcluster system, using the VMS MSCP Server.
- Maintenance Operations Protocol (MOP) server — A CPU that services satellite boot requests, using DECnet-VAX software, to provide the initial Local Area Network (LAN) down-line load sequence of the VMS Operating System and VAXcluster software. At the end of the initial down-line load sequence, the satellite uses a disk server to perform the remainder of the VMS booting process.
- Satellite — A CPU that is booted over a LAN using a MOP server and disk server.
- Tape server — A CPU that makes TMSCP tapes to which it has direct access available to other CPUs in the VAXcluster system, using the VMS TMSCP Server.
- Mixed Interconnect VAXcluster System — A VAXcluster system that uses more than one type of interconnect for VAXcluster communication.

Interconnects:

VAXcluster systems are configured by connecting multiple CPUs with a communication media, referred to as an interconnect. VAXcluster nodes communicate with each other using the most appropriate interconnect available. Whenever possible, in the event of interconnect failure, VAXcluster software will automatically use an alternate interconnect. VAXcluster Software supports any combination of the following interconnects:

- Computer Interconnect (CI)
- Ethernet (NI)
- Digital Storage System Interconnect (DSSI)*
- Fiber Distributed Data Interface (FDDI)

Ethernet and FDDI are industry-standard general purpose communications interconnects that can be used to implement a Local Area Network (LAN). Except where noted, VAXcluster support for both of these LAN types is identical.

Configuration Rules:

The following configuration rules apply to VAXcluster systems:

- The maximum number of CPUs supported in a VAXcluster system is 96.
- Every VAXcluster node must have a direct communication path to every other VAXcluster node via any of the supported interconnects.
- VAX 11/7xx, 6000, 8xxx, and 9000-series CPUs require a system disk that is accessed via a local controller or through a local CI or DSSI connection. VAXcluster satellite booting is not supported for these systems.
- A Star Coupler is a common connection point for CI connected CPUs and HSC subsystems. All CPUs connected to a Star Coupler must be configured as VAXcluster members. A VAXcluster system can include any number of Star Couplers. The number of CI adapters supported by different CPUs can be found in Table 2 in this SPD; the number of Star Couplers that a CPU can be connected to is limited by the number of adapters it is configured with.
- The maximum number of CPUs that can be connected to a Star Coupler is 16, regardless of Star Coupler size.
- The RA-series disks and TA-series tapes can be dual pathed between pairs of HSC subsystems on the *same* Star Coupler, or between two local controllers. Such dual pathing provides enhanced data availability using a VMS automatic recovery capability called failover. Failover is the ability to use an alternate hardware path from a CPU to a storage device when a failure occurs on the current path. The failover process is transparent to applications. Dual pathing between an HSC and a local controller is not permitted. When two local controllers are used for dual pathing, each controller must be located on a separate CPU.
- When multiple CPUs are connected to a common DSSI they must all be configured as VAXcluster members. Since the KFQSA Q-bus-to-DSSI adapter does not support VAXcluster communication to other CPUs on the DSSI, CPUs using this adapter must include another interconnect for VAXcluster communication.
- VAX 6000-series CPUs can be connected to a DSSI bus using the KFMSA XMI-DSSI adapter. Any mix of VAX 6000-series and VAX 4000-series systems (excluding the VAX 4000 Model 200) can be configured on a common DSSI bus, up to a maximum of three CPUs.
- A maximum of three VAX 4000 series, Q-bus-based MicroVAX 3000 series, and MicroVAX II systems can be configured on a common DSSI bus. In triple CPU configurations, the middle CPU must be a VAX 4000 Model 300, or higher, system.

* The DSSI is not used as a VAXcluster interconnect when accessed via a KFQSA Q-bus adapter. The KFQSA adapter only supports access to DSSI mass storage devices.

- VAXcluster systems support 4 LAN adapters per CPU for VAXcluster communications. LAN segments can be bridged to form an extended LAN.
- CPUs that use an Ethernet for VAXcluster communications can concurrently use it for other network protocols that conform to the applicable Ethernet standards, such as Ethernet V2.0, IEEE 802.2, and IEEE 802.3.
- CPUs that use an FDDI for VAXcluster communications can concurrently use it for other network protocols that conform to the applicable FDDI standards, such as ANSI X3.139-1987, ANSI X3.148-1988, and ANSI X3.166-1990
- All LAN bridges must provide a low-latency data path, with approximately 10 megabits per second throughput for Ethernet and 100 megabits per second throughput for FDDI. Translating bridges must be used when connecting VAXcluster nodes on an Ethernet to those on an FDDI.
- The maximum number of VAXcluster members that can be directly connected to the FDDI, via the DEC FDDIconroller 400 (DEMFA), is 16.
- A DECnet–VAX communication path must exist between all nodes in a VAXcluster system.
- A single time zone setting must be used by all CPUs in a VAXcluster system.
- A VAXcluster system can be configured with a maximum of one Quorum Disk. A Quorum Disk cannot be a member of a shadow, volume, or stripe set.

Recommendations:

The optimal VAXcluster system configuration for any computing environment is based on requirements of cost, functionality, performance, capacity, and availability. Factors that impact these requirements include:

- Applications in use
- Number of users
- Number and model of CPUs
- Interconnects and adapter types
- Disk and tape I/O capacity and access time
- Number of disks and tapes being served
- Interconnect utilization

Digital recommends VAXcluster system configurations based on its experience with the VAXcluster Software Product. The customer should evaluate specific application dependencies and performance requirements to determine an appropriate configuration for the desired computing environment.

When planning a VAXcluster system, consider the following recommendations:

- While VAXcluster systems can include any number of system disks, performance and disk space should be considered in determining their number and location. It is important to recognize that system management efforts increase in proportion to the number of system disks.
- VAXcluster CPUs should be configured using interconnects that provide appropriate performance for the required system usage. In general, use the highest performance interconnect possible. CI, DSSI, and FDDI are the preferred interconnects between powerful VAX CPUs.

- Data availability and I/O performance is enhanced when multiple VAXcluster nodes have direct access to shared storage; whenever possible, configure systems to allow direct access to shared storage in favor of VMS MSCP Served access. Multi-access DSSI- and HSC-based storage provide higher data availability and I/O performance than singly accessed, local controller-based storage. Additionally, dual pathing of DSA disks between local or HSC storage controllers enhances data availability in the event of controller failure.
- VAXcluster systems can provide enhanced availability by utilizing redundant components. For example, additional CPUs, storage controllers, and disks and tapes can be configured. Extra peripheral options such as printers and terminals can be included to further enhance availability. Multiple instances of all the VAXcluster interconnects (CI, DSSI, Ethernet, and FDDI) are supported.
- If possible, LAN-based and Mixed Interconnect VAXcluster systems should include multiple MOP and disk servers to enhance availability. When a server fails in configurations that include multiple servers, satellite access to disks fails over to another server. Disk servers should be the most powerful CPUs in the VAXcluster and should use the highest bandwidth LAN adapters available.
- When a LAN-based VAXcluster system is configured with high performance nodes, multiple LAN adapters and interconnects can be used to increase total communication bandwidth.
- Maintenance of complex LAN-based VAXcluster configurations can be simplified with the aid of the VMS LAVC\$FAILURE_ANALYSIS program, which is available in the SYS\$EXAMPLES directory.
- VAXclusters are sensitive to the LAN traffic levels. The average LAN segment utilization should not exceed 60 percent for any 10 second interval. Nodes can leave the cluster if they cannot properly exchange the HELLO messages every three seconds. LAN bridges can be used to localize VAXcluster system traffic should the overall level of network traffic be a concern. Also, it is possible for VAXcluster nodes to exist on both sides of a LAN bridge.
- The performance of an FDDI LAN will vary with each configuration. When an FDDI is used for VAXcluster communications, the ring latency when the FDDI ring is idle should not exceed 400 microseconds.
- When under heavy network load, bridges are subject to packet loss and retransmission, due to congestion. This is especially true of Ethernet to FDDI bridges. In a VAXcluster environment, heavy network loads can result when many satellite nodes are booted simultaneously. It may be necessary to minimize simultaneous booting, or limit the number of nodes that utilize these LAN bridges.
- The VAXcluster Multi-Datacenter Facility is specifically designed to allow successful implementation and management of disaster-tolerant configurations and to deliver predictable recovery from site failures.
For more information, refer to the VAXcluster Multi-Datacenter Facility Software Product Description (SPD 35.05.xx).
- The optional VMS Volume Shadowing System Integrated Product provides the following advantages:
 - Enhanced data availability in the event of disk failure

- Enhanced read performance with multiple system and data disks
- For more information, refer to the VMS Volume Shadowing Software Product Description (SPD 27.29.xx.).

B.2 Hardware Support

Supported CPUs:

Any VAX, VAXstation, or MicroVAX CPU, as documented in the VMS SPD, can be used in a VAXcluster, with the exception of the VAX-11/730, VAX-11/782, and VAXstation 8000 CPUs.

Any CPU can be configured as a VAXcluster satellite node, with the exception of VAX 11/7xx, 6000, 8xxx, and 9000-series CPUs.

For MicroVAX 3500 and MicroVAX 3600 CPUs configured with KFQSA DSSI adapters, the console ROMs must be at Revision Level V5.1, at a minimum.

Supported CI Adapters:

VAXcluster nodes can be configured with multiple CI adapters. Table B-1 shows the types of adapters that are supported by each CPU. There can only be one type of adapter configured on a CPU; the maximum quantity of each type is noted in the table. The CI adapters in a CPU can connect to the same, or different, Star Couplers.

Note

The CIBCA-A and CIBCA-B are different.

Table B-1 Number and Type of Adapters Supported

CPU Type	CI750	CI780	CIBCI	CIBCA-A	CIBCA-B	CIXCD
11/750	1	–	–	–	–	–
11/780	–	1	–	–	–	–
11/785	–	1	–	–	–	–
6000-xxx	–	–	–	1	4	4
82xx	–	–	1	1	1	–
83xx	–	–	1	1	1	–
85xx	–	–	1	1	2	–
86xx	–	2	–	–	–	–
8700	–	–	1	1	2	–
88xx	–	–	1	1	2	–
9000-xxx	–	–	–	–	–	6

Supported LAN Adapters:

Table B-2 shows the types of Local Area Network (LAN) adapters supported by VAXcluster software.

Table B–2 LAN Adapters Supported

Bus	Ethernet	FDDI
XMI	DEMNA	DEMFA
BI	DEBNI,DEBNA	
Q–bus	DELQA,DESQA	
Q–bus	DEQTA (DELQA-YM)	
UNIBUS	DEUNA,DELUA	
Integral	LANCE,SGEC	

Supported Peripheral Options:

VAXcluster systems can use all peripheral options supported by the VMS SPD. Refer to the VMS Software Product Description (SPD 25.01.xx) for further information.

Memory Requirements:

All VAXcluster CPUs must have a minimum of 4 megabytes of physical memory.

Star Coupler Expander:

A Computer Interconnect Star Coupler Expander (CISCE) can be added to any Star Coupler to increase its connection capacity to 32 ports. The maximum number of CPUs that can be connected to a Star Coupler is 16, regardless of size.

HSC Subsystems:

VAXcluster software supports all models of the HSC family of intelligent mass storage controllers. These controllers include many features:

- The ability to provide high data throughput and I/O rates
- Implementation of many mass storage performance optimization techniques
- Multi-CPU access to disk and tape units
- The ability to configure multiple disk and tape units
- Optional HSC-based disk caching (for the HSC60 and HSC90)
- HSC resident maintenance and backup facilities

The following rules apply for HSC subsystems:

- HSC Software, Version 6.0, at a minimum, is required for the HSC40, HSC60, HSC70, and HSC90. HSC Software, Version 4.1, at a minimum, is required for the HSC50.
- Each HSC40 supports a maximum of 12 ports.
- Each HSC50 supports a maximum of 24 ports.
- Each HSC60 supports a maximum of 20 ports.
- Each HSC70 supports a maximum of 32 ports.
- Each HSC90 supports a maximum of 48 ports.
- All ports can be used for disk storage. The maximum number of ports that can be used for tapes is 24 for the HSC70 and HSC90, and 12 for the HSC40, HSC50, and HSC60.

B.3 Software Requirements

- VMS Operating System

VAXcluster Software, Version 5.5 is a VMS System Integrated Product that requires VMS, Version 5.5.

Refer to the VMS Software Product Description (SPD 25.01.xx) for further information.

VMS, Version 5.4, and all its sub-versions (for example V5.4-1, and V5.4-2), can coexist in a VAXcluster with VMS, Version 5.5 (and all its sub-versions). Only one version of VMS can exist on each system disk. In configurations with multiple system disks, a rolling upgrade can be performed so that continuous VAXcluster system operation is maintained during the upgrade process. During a rolling upgrade, a separate system disk is required for each version. Rolling upgrades occur in a series of phases during which all VAXcluster nodes are brought up to the latest VMS version.

During a rolling upgrade from Version 5.4 to Version 5.5, V5.4 Batch and Print functionality is maintained. Once the VAXcluster system is fully upgraded to Version 5.5, the new V5.5 Batch and Print functionality becomes available. Once the new V5.5 Batch and Print facility is operational, booting V5.4 CPUs into the VAXcluster system is only permitted if they do not use any Batch and Print operations (i.e., START/QUEUE/MANAGER). Coexistence of the V5.4 and V5.5 Batch and Print facilities is not supported.

Digital recommends that all VAX systems in a VAXcluster run the latest version of VMS.

- DECnet–VAX Software

All VAXcluster CPUs require either an End Node or Full Function DECnet–VAX license.

Refer to the DECnet–VAX Software Product Description (SPD 25.03.xx) for further information.

B.4 Optional Software

For information on VAXcluster support for optional software products, refer to the *VAXcluster Support* section of the Software Product Description (SPD) documents for those products.

Optional products that are particularly useful in VAXcluster systems include:

- VMS Volume Shadowing (SPD 27.29.xx)
- VAX Performance Advisor (SPD 27.71.xx)
- VAXcluster Console System (SPD 27.46.xx)
- VAXcluster Multi-Datacenter Facility (SPD 35.05.xx)
- VAX Disk Striping (SPD 31.66.xx)

B.5 Growth Considerations

The minimum hardware/software requirements for any future version of this product may be different from the requirements for the current version.

B.6 Ordering Information

Software Licenses: QL-VBRA*-AA

Software Product Services: QT-VBRA*-**

* Denotes variant fields. For additional information on available licenses, services, and media, refer to the appropriate price book.

The above information is valid at time of release. Please contact your local Digital office for the most up-to-date information.

B.7 Software Licensing

A VAXcluster Software license is required for each CPU in a VAXcluster system.

This software is furnished under the licensing provisions of Digital Equipment Corporation's Standard Terms and Conditions. For more information about Digital's licensing terms and policies, contact your local Digital office.

A VAXcluster Multi-Datacenter Facility license is required when using VAXcluster software for implementing disaster tolerance. Disaster tolerance is the ability to recover from major site failure within a brief recovery period when using a single VAXcluster system that spans multiple buildings.

License Management Facility Support:

The VAXcluster Software product supports the VMS License Management Facility (LMF).

License units for this product are allocated on an Unlimited System Use basis.

For more information about the License Management Facility, refer to the VMS Operating System Software Product Description (SPD 25.01.xx) or the *License Management Facility Manual* of the VMS Operating System documentation set.

For more information about Digital's licensing terms and policies, contact your local Digital office.

B.8 Software Product Services

A variety of service options are available from Digital. For more information, contact your local Digital office.

B.9 Software Warranty

Warranty for this software product is provided by Digital with the purchase of a license for the product as defined in the Software Warranty Addendum of this SPD. Table B-3 contains trademark information pertinent to the software warranty.

Table B-3 Trademark Information

™ The DIGITAL Logo, BI, CI, DECnet-VAX, DELUA, DEUNA, HSC, HSC40, HSC50, HSC60, HSC70, HSC90, MicroVAX, Q-bus, RA, TA, UNIBUS, VAX, VAXstation, VAXcluster, VMS, and XMI are trademarks of Digital Equipment Corporation.

C

Specifications for Mature Products

This appendix contains information on products that can be used in a VAXcluster system, but that Digital no longer ships. The specifications for these products are provided so you can design these older products into your VAXcluster system, if you have them.

C.1 CPU Information

Table C-1 contains information related to Table 3-1.

Table C-1 VAX System CPU Performance Characteristics

CPU Name	VUPs	Maximum Memory ¹	Packaging	Upgrade Potential	Performance Option
2000	0.9	14	Desktop		4 or 8 plane
II	0.9	16	Pedestal compact cabinet	To VAXstation II/GPX	
II/GPX	0.9	16	Pedestal compact cabinet		To 8 plane
VAXstation 8000	1.2	32	Pedestal compact cabinet		58 plane
VAX-11/750	0.65	14	1 cabinet		
VAX-11/780	1.0	64 ²	Dual cabinet	780 ³ to 785	
VAX-11/785	1.5 to 1.7	64 ²	Dual cabinet		
8200/8250	1.0/1.2	128	Cabinet 12/24 slot	To 8350	
8350	2.0/2.3	128	Cabinet 12/24 slot		
8530/8550	4.0/6.0	320	1 cabinet, single bay	8530 to 8550	
8600/8650	8600: 4.2 8650: 6.0	256	1 cabinet, dual bay plus front end	8600 to 8650	
8700	6.0	256	2 cabinets, triple bay	To 8800	
8810/8820	6.0/11.4	512	3 cabinets, triple bay	8810 to 8820 8820 to 8830	
8830/8840	16.8/22.0	512	3 cabinets, triple bay	8830 to 8840	

¹In megabytes.

²To 128 megabytes with expansion cabinet.

³A VAX-11/782 system can be reconfigured into two VAX-11/780 systems for use in a VAXcluster system.

Table C–2 contains information related to Table 3–3.

Table C–2 VAX System I/O Performance Characteristics

CPU Name	Internal Bus		Storage Bus Options	VAXcluster Communications Options
	Type	Throughput ¹		
2000	Integral	>3.3	ST506	Ethernet
II	Q–bus	>3.3	Q–bus, SDI, DSSI	Ethernet
II/GPX	Integral	3.3	Q–bus, ST506	Ethernet
VAXstation 8000	VAXBI	13.3	VAXBI	Ethernet, CI
8200/8250	VAXBI	13.3	UNIBUS VAXBI, SDI	Ethernet, CI
8300/8350	VAXBI	13.3	UNIBUS VAXBI, SDI	Ethernet, CI
8500/8530 8550	NMI	16.0	UNIBUS ² VAXBI ³ ,SDI	Ethernet, CI
8600/8650	SBI	13.3	SDI UNIBUS ² MASSBUS	Ethernet, CI
8700/8800	NMI	30.0	UNIBUS ² VAXBI, SDI	Ethernet, CI
8810/8820 8830/8840	NMI	30.0	VAXBI ⁴ UNIBUS ² , SDI	Ethernet, CI
VAX–11/750	CMI	5.0	UNIBUS MASSBUS	Ethernet, CI
VAX–11/780	SBI	13.3	UNIBUS MASSBUS	Ethernet, CI
VAX–11/785	SBI	13.3	UNIBUS MASSBUS	Ethernet, CI

¹Maximum bus throughput in megabytes per second.

²Limited UNIBUS support.

³Theoretical I/O bus bandwidth per system in megabytes per second.

⁴Using up to three H9657 VAXBI expansion cabinets.

Table C–3 contains information related to Table 7–3.

Table C-3 Disk Server Capacity — Average (4-Block) I/O Operations Per Second

CPU Type	Ethernet Adapter Type	Throughput ¹	Limiting Resource
VAX-11/750	DEUNA	45	CPU or DEUNA
VAX 8300	DEBNA	50	CPU
VAX 8200	DEBNA	55	CPU
VAX 8350	DEBNA	60	CPU
VAX 8250	DEBNA	65	CPU
VAX-11/780	DELUA	70	CPU
VAX-11/785, VAX 8600, VAX 8650	DELUA	100	DELUA
MicroVAX 2000	DESVA	20	ST506 Controller
MicroVAX II/RD disks	DEQNA DELQA ²	45	RQDX3 Controller
MicroVAX II/RA disks	DEQNA DELQA ²	80	CPU

¹4-block I/Os per second.²The DEQNA must be Revision K or higher. The DELQA is the preferred device.

C.2 Adapter Information

Table C-4 contains information related to Table 4-2.

Table C-4 Ethernet Adapters

Device	Internal Bus	VAX CPU
DESVA	Embedded	VAXstation 2000, MicroVAX 2000
DEQNA ¹	Q-bus	VAXstation II, VAXstation II/GPX, MicroVAX II
DEUNA	UNIBUS	VAX-11/7XX
DELUA	UNIBUS	VAX-11/7XX, 86XX

¹The DEQNA must be Revision K or higher. The DELQA is the preferred device.

Table C-5 contains information related to Table 4-3.

Table C-5 CI Adapters

Device	Internal Bus	Data Rate ¹	VAX CPU
CI750	CMI	Limited by CPU	VAX-11/750
CI780	SBI	0.4 to 1.8	VAX-11/780, VAX-11/785, 8600, 8650
CIBCI	VAXBI	0.4 to 1.2	8840, 8800, 8700, 85x0, 8350, 8250
CIBCA-A	VAXBI	0.4 to 1.4	8840, 8800, 8700, 85x0, 8350, 8250

¹Megabytes per second. Data rate capacity varies with speed of CPU and message size.

Table C–6 contains information related to Table 6–1.

Table C–6 Maximum CI Adapters Per VAXcluster CPU

CPU	CI750	CI780	CIBCI	CIBCA-A	CIBCA-B	CIXCD
VAX–11/750	1					
VAX–11/780		1				
VAX–11/785	1					
VAX 82XX, 83XX		1	1		1	
VAX 85XX, 87XX, 88XX		1	1		2	
VAX 86XX	2					

Table C–7 contains information related to Table 6–2.

Table C–7 DSSI Adapters Per CPU

CPU	EDA640	SHAC	SWIFT	KFQSA	KFMSA
MicroVAX II				2	

C.3 Storage Information

Table C–8 contains information related to Table 5–5.

Table C–8 Disk Attributes for Mature Drives

Disk Drive	Formatted Capacity MB	Formatted Capacity Blocks	Average Access Time ¹	Requests Per Second ²	Spiral Read Rate ³	Interface	Fixed Removable
RZ22	52	102.4 K	33.4	26	1.25	SCSI-2	F
RA60	205	400 K	50	20	860	SDI	R
RA80	121	236 K	33.3	30	919	SDI	F
RA81	456	891 K	36.3	27	1219	SDI	F
RA82	625	1.2 M	32.3	31	1459	SDI	F
RD52	31	0.06 M	57.5	18	625	ST506	F
RD53	71	0.14 M	38.3	27	625	ST506	F
RF30	150	293 K	29.3	34	.75	DSSI	F
RF31	381	0.74 M	23.6	40	1200	DSSI	F/R
RF31F	200	0.39 M	20.5	46	1200	DSSI	F
RRD40	600	1.2 M	500	2	.15	SCSI, Q-bus	R
RRD50	600	1.2 M	2000	.5	.15	SCSI, Q-bus	R

¹Average seek plus average latency in milliseconds.

²Measured when average response time was less than 100 milliseconds.

³Measured in megabytes per second.

Table C-9 contains information related to Table 5-7.

Table C-9 Library Attributes

Media Type	Capacity ¹	Fetch Media Time ²	Data Transfer ³	No. Drives
RV64	128	10-15	262	1-4

¹Total unit capacity in gigabytes.
²Robotic service time in seconds.
³Kilobytes per second, per drive.

Table C-10 contains information related to Table 5-8.

Table C-10 Tape Attributes

Tape Drive	Density ¹	Speed ²	Data Transfer ³	Recording Method	Media Size	Media Capacity	Interface
TF837	10,000	100	90	TTSP ⁴	295 MB	2 GB	DSSI
TU79	6250/1600	125	781	Group, PE ⁵	150 M		UNIBUS
TU80	1600	25/100	160	PE ⁵	40 MB		UNIBUS

¹Bits per inch.

²Inches per second.

³Kilobytes per second. User data.

⁴Two-track serpentine pattern.

⁵Group code recording to ANSI Standard X3.54-1976 and Phase Encoded X3.39-1973.

Table C-11 contains information related to Table 5-9.

Table C-11 Bus Type and I/O Rates for SDI Controllers

Controller	Bus Type	No. of Ports	No. of SDI Channels	Transfer Rate Per second	Request Rate Per second
UDA50	UNIBUS	4	1	750 KB	90

Table C-12 contains information on UNIBUS devices.

Table C-12 UNIBUS Storage Devices

Drive Type	Medium	Formatted Capacity	
		Bytes	Blocks
RC25 ¹	Fixed/removable cartridge	26 MB	50.7 K
RK06 ²	Removable pack	14 MB	27.3 K
RK07 ²	Removable pack	28 MB	54.6 K
RL02 ³	Removable cartridge	10 MB	19.5 K

¹RUC25/RQC25 UNIBUS/Q-bus.

²RK611 UNIBUS.

³RL11 UNIBUS.

Table C-13 contains information on MASSBUS devices.

Table C-13 MASSBUS Storage Devices

Drive Type ¹	Medium	Formatted Capacity	
		Bytes	Blocks
RM03	Removable	67 MB	122.0 M
RM05	Removable	256 MB	488.0 K
RM80	Fixed	124 MB	242.0 K
RP04	Removable	88 MB	170.0 K
RP05	Removable	88 MB	170.0 K
RP06	Removable	176 MB	340.0 K
RP07	Fixed	516 MB	1.0 M

¹MASSBUS disk drives can be attached to RH780 or RH750 controllers.

Table C-14 contains information related to Table 7-1.

Table C-14 Disk Server I/O Capacity Based on 80 Percent CPU Utilization

CPU Type	Average (4-Block) I/Os Per Second
VAX-11/750	45
VAX 8300	50
VAX 8200	55
VAX 8350	60
VAX 8250	65
MicroVAX 2000	65
VAX-11/780	70
MicroVAX II	80
VAX-11/785	105
VAX 8500	210
VAX 8530	220
VAX 8600	105
VAX 8650	105
VAX 8700	370
VAX 88XX	340

C.4 Printer Information

Table C-15 contains information related to Table 7-5.

Table C-15 Selected Digital Printers

Model	Speed	Type	Interface	Protocol	Consumables Replacement	Volume
LN03	8 ppm ¹	Laser	Serial ¹	ASCII	User	Low
LN03R	8 ppm	Laser	Serial	PostScript	User	Low
LN03S	8 ppm	Laser	Serial	ASCII /Graphics	User	Low
LA210	40 cps	Dot matrix	Parallel ²	ASCII	User	Low
LP25	300 lpm	Band	Parallel	ASCII	User	Low
LP27	1200 lpm	Band	Parallel	ASCII	Operator	High
LQP03	25 cps	Print wheel	Serial	ASCII	User	Low
LQP45	45 cps	Print wheel	Serial	ASCII	User	Low

¹Connect a serial printer to a DHQ11 or DECserver (EIA-232 or DEC 423) port.

²Connect parallel printers directly to your I/O interface.

Glossary

availability

The amount of time a computer system is actually working compared to the total time it is expected to be operational, expressed as a percent. The perceived importance of availability is proportional to the perceived cost of down time. *See also* data availability.

boot server

A VAX CPU in a local area or mixed-interconnect VAXcluster system responsible for booting and providing system disk service to one or more satellite nodes.

CI

A high-speed bus with dual data paths that connect all nodes in a CI VAXcluster system. The bus bandwidth is 70 megabytes per second per path.

CI VAXcluster system

A VAXcluster system in which all nodes are connected to CI cables.

common system disk

A VMS disk that supports the booting of two or more VAX processors.

CPU

Central processing unit. An electronic unit that includes one or more processors and acts as a central controlling body. For example, the VAX 6000-620 CPU includes two processors. *See also* processor.

data availability

The amount of time data is actually accessible compared to the total time it is expected to be accessible, expressed as a percent. *See also* availability.

data disk

A disk that is strictly used for application (nonsystem) data. No system roots exist on it; no VAX CPU boots from it.

Digital building block

Any major hardware component you can use to configure a VAXcluster system.

DSA

DIGITAL Storage Architecture. An architecture for connecting disks and tapes to VAX processors.

disk fragmentation

Data stored on disk is stored most efficiently in contiguous blocks. When a disk has been used extensively, new data stored on the disk must often be subdivided into fragments that fit into the remaining space. Encountering several such fragments when retrieving data can reduce performance.

disk server node

A VAX CPU that provides disk service to other VAXcluster nodes.

DSA Disk

A disk designed according to the DIGITAL Storage Architecture (DSA), which connects using the SDI or DSSI.

DSSI

Digital Storage Systems Interconnect. A daisy-chained multidrop interconnect bus that attaches CPUs to integrated storage elements.

dual-ported disk drive

A disk drive in which each port is attached to a different controller or CPU.

Ethernet

Both a local area network protocol using a carrier sense multiple access with collision detection (CSMA/CD) scheme to arbitrate the use of a 10-megabit per second baseband coaxial cable and the coaxial cable itself. Similar to ANSI/IEEE Standard 802.3/.3a,.3b,.3c,.3d,.3e/.4.

Ethernet-based VAXcluster system

See local area VAXcluster system.

failover

The automatic or manual action that sustains access to a CPU, storage device, or queue even when one access path is broken. Automatic failover is transparent to the user; manual failover must be induced by operator intervention.

FDDI

Fiber Distributed Data Interface. An ANSI standard 100-megabit per second interconnect that uses fiber optic cable.

HSC

An intelligent mass storage subsystem, a non-CPU node on the CI, that provides disk and tape resources to users in the VAXcluster system. HSC subsystems are connected to Star Couplers with CI cables.

host

A VAX CPU in a VAXcluster system.

ISE

Integrated Storage Element. A DSSI storage device that has a storage controller built into it.

local area VAXcluster system

A VAXcluster system in which all nodes are connected using Ethernet or FDDI. Disks in a local area VAXcluster system are connected to adapters on local area VAXcluster system CPUs.

local system disk

A disk used as a system disk by a VAX CPU connected directly to it by a local adapter. If it is dual-ported, a local disk cannot be a local system disk.

MASSBUS

A mature mass storage interconnect that supports older storage systems.

MSCP

A standardized protocol for communicating between hosts and DSA storage devices.

MSCP server

An MSCP server makes disks available to VAXcluster systems over an interconnect.

mixed-interconnect VAXcluster system

A VAXcluster system in which nodes may be connected with multiple interconnect types, or even several interconnects simultaneously via multiple adapters.

multihost controller

An I/O controller attached to more than one CPU.

multiple access path

A path from user to data stored on disk and accessible over more than one route.

multiple access path disks

Disks that are connected to two storage controllers or adapters. A multiple access path to a disk creates a redundant configuration.

multiprocessing

The ability of a CPU to handle multiple jobs simultaneously.

multiprocessor

A CPU with more than one processor, for example, a VAX 6000-240.

multistream

The characteristic of a system that enables it to execute more than one job at a time, for example, multistream batch processing.

processor

An electronic unit that acts upon instructions to compute a result. Examples of processors include arithmetic processors, I/O processors, floating-point processors, vector processors, and communications processors. *See also* CPU.

quorum disk

A disk drive given a vote for purposes of maintaining VAXcluster availability. A quorum disk is generally used in a two CPU VAXcluster system where it helps to maintain quorum if one CPU leaves the VAXcluster system. You can use HSC disks or DSA disks that are connected to two local adapters as quorum disks.

response time

The time it takes a system to answer or react to a query from a terminal. More specifically, the elapsed time between generation of the last character of a message at a terminal and receipt of the first character of the reply. Response time can include terminal delay, transmission delay, service node delay, and processing delay.

satellite node

A workstation, MicroVAX, or VAX CPU, running VMS, connected to the Ethernet, and booted by a boot server.

satellite-only system disk

A disk that is used as a system disk only by VAX CPUs booting over the Ethernet. No VAX CPU performs a local disk boot from this disk.

scalability

The ability to grow a VAXcluster system while maintaining and fully using the initial configuration equipment.

SDI

The disk interconnection scheme used by RA-series disks that is part of the DIGITAL Storage Architecture (DSA).

server node

A VAXcluster CPU node that performs a specific function for the other members of the configuration.

shadow sets

All the disks to which duplicate data is written when volume shadowing is used.

SI

Storage Interconnect. The overall interconnection scheme used by RA-series disk drives, TA-series tape drives, and the ESE20 solid-state disk that is part of the DIGITAL Storage Architecture (DSA).

single-access path

A path from user to data that is stored on disk, achievable by one route only.

single-host controller

A disk drive controller accessible from only one VAX CPU.

single-ported disk drive

A disk drive that is connected to only one controller.

single stream

The characteristic of an application that restricts it to executing one instruction at a time.

SMP

Symmetrical multiprocessing. A system in which the processors share a common memory and execute a single memory-resident copy of VMS.

spare repair

A compatible and correctly configured spare disk drive that can be used to replace a failed member of a disk farm.

Star Coupler

The common connection point for all CPUs and HSC subsystems connected to the CI bus.

STI

The tape interconnection scheme that is used by TAxx tape drives and is part of the DSA.

system disk

A disk on which the VMS operating system is located.

virtual circuit

A logical data path between two nodes in a VAXcluster system. The port driver establishes a virtual circuit to cooperating port drivers on other nodes.

VMS Volume Shadowing

A process by which identical data is written to multiple disk volumes to increase data availability in the event of disk failures. The VMS Volume Shadowing product provides shadowing across a wide range of disks and controllers anywhere in a VAXcluster system.

VUP

VAX Unit of Processing. A measure of the relative performance of a processor. The performance of the VAX-11/780 is defined as 1 VUP.

work group

A group of system users who share common needs and use the same software. Work groups may benefit from sharing VAXcluster system files and public directories.

Index

A

- Access path
 - See also* Redundancy
 - dual-boot servers, 5–15
 - HSC, 5–14
 - local adapter, 5–14
 - system disk, 5–15
- Adapter
 - CI, 4–4, C–3
 - DSSI, 4–5
 - Ethernet, 4–2, 4–3, C–3
- Application requirements, 2–1
 - availability, 3–2
 - determining, 2–1, 3–1
 - gathering, 5–5
 - growth, 3–2
 - I/O, 3–2, 5–5
- Archive media, 5–7
- Automatic failover, 5–14
 - dual-ported disks, 5–14
 - HSC, 5–14
- Availability
 - See also* Volume shadowing
 - CI, 4–3, 5–14
 - CPU, 3–3
 - data, 5–13, 5–16
 - device, 5–13
 - disk, 2–3
 - disk striping, 5–18
 - HSC, 5–14
 - increasing in an LAVc, 7–2
 - interconnect, 5–13
 - printer, 2–3
 - quorum, 7–4
 - storage requirements, 5–12
 - system disk, 5–15, 7–9
 - VAXcluster system, 2–2
 - volume shadowing, 5–13

B

- Backup
 - database, 5–19
 - media, 5–7
 - static data, 5–19
 - strategy, 5–19

Backup (cont'd)

unattended, 5–20

Backup Utility

- disk fragmentation, 5–19
- file-by-file copy, 5–20
- HSC, 7–9
- image copy, 5–20
- optimizing performance, 7–8
- redundancy, 5–18
- VMS, 7–9

Bandwidth

See Throughput

Baseband network

See Ethernet

Booting

Bottleneck

- disk I/O, 7–7
- disk server I/O capacity, 7–7

Bus

SCSI, 4–8

C

Capacity planning, 2–3

- CI, 1–5
 - adapter, 4–4, C–3
 - adapter throughput, 4–4
 - advantages, 4–3
 - availability, 5–14
 - components, 4–3
 - load sharing, 4–7
 - multiple, 4–7
 - preferred path, 4–7
 - redundancy, 4–3
 - throughput, 4–4
- VAXcluster configuration, 6–1, 6–2

CI VAXcluster configuration

See VAXcluster configuration
(fig), 1–5

CISCE (Star Coupler Expander), 4–3

Clusterwide Process Services, 1–2

CLUSTER_CONFIG.COM command procedure,
7–11

Computing environment, 1–1

compute-intensive, 2–1

defining, 2–1

I/O-intensive, 2–1

Computing environment (cont'd)

- timesharing, 2-1
- Computing style, 1-2
 - batch, 3-2
 - interactive, 3-2
 - SMP, 1-2
 - uniprocessor, 1-2
- Configuration
 - See also* VAXcluster configuration
 - dual-host (*fig*), 6-4
 - multiple adapter (*fig*), 6-8
 - multiple CI (*fig*), 6-2
 - multiple DSSI (*fig*), 6-6
- Configuration rules, 6-1
 - CI VAXcluster system, 6-1, 6-2
 - DSSI VAXcluster system, 6-4, 6-5
 - Ethernet VAXcluster system, 6-7
 - FDDI VAXcluster system, 6-7
 - multiple CI connections, 6-2
 - multiple DSSI connections, 6-5
 - multiple LAN adapters, 6-7
- Connection Manager, 1-2
 - quorum, 7-4
 - state transition, 7-5
- Controller
 - See* HSC
- CPU, 3-1
 - adding to VAXcluster system, 7-11
 - availability, 3-3
 - changing characteristics, 7-11
 - characteristics, 3-4 to 3-8
 - client/server guidelines, 7-5
 - disk server I/O capacity, 7-5, 7-6
 - disk servers, 1-3
 - Ethernet adapter I/O capacity, 7-7
 - expanding power, 3-2
 - failure, 3-3
 - I/O characteristics, 3-7 to 3-8
 - interconnect options, 3-8
 - maximum CI adapters, 6-2, C-4
 - maximum DSSI adapters, 6-5, C-4
 - number in a VAXcluster system, 3-3
 - overhead, 4-1
 - packaging information, 3-6
 - performance characteristics, 3-4 to 3-7
 - performance interaction, 3-3
 - removing from VAXcluster system, 7-11
 - selection guidelines, 3-1
 - storage bus options, 3-8
 - supported in a VAXcluster system, 1-2
- CPU configuration, 1-1, 3-3
 - fault-tolerant, 3-3
 - selecting, 3-3
- VAXcluster system, 3-3

D

Data

- availability, 5-13, 5-16
- backup, 5-19
- redundancy, 2-3, 5-13, 7-2
- shared access, 1-2
- Database
 - backup, 5-19
 - storage requirements, 5-6
- DCM (Data Center Monitor), 7-13
- DECalert, 7-13
- DECcmds, 7-13
- DECmcc, 7-15
- DECnet-VAX, 1-2
- DECperformance Solution, 7-13
- DECscheduler, 7-13
- DESNC (Digital Ethernet Secure Network Controller), 2-3
- Device availability
 - See* Availability
- Disk
 - See also* Storage device
 - See also* System disk
 - attributes, 5-21, C-4
 - availability, 2-3
 - configuring system disk, 7-9
 - ESE-series, 5-21
 - estimating requirements, 5-5
 - ISE, 5-21
 - limiting I/O performance, 7-7
 - MASSBUS, C-6
 - quorum, 7-4
 - RA-series, 5-21, C-4
 - RD-series, 5-21
 - RF-series, 4-6, 5-21, C-4
 - RRD-series, 5-21, C-4
 - RZ-series, 5-21, C-4
 - SA-series, 5-22
 - server configuration guidelines, 7-5
 - UNIBUS, C-5
- Disk fragmentation, 5-19
- Disk server, 1-3, 7-5
 - dual-host, 6-4
 - Ethernet adapter I/O capacity, 7-7
 - failover, 7-3
 - I/O capacity, 7-5, 7-6
 - SMP CPU, 7-6
- Disk striping, 5-18
 - availability, 5-18
- Distributed Job Controller, 1-2
- Distributed Lock Manager, 1-2
 - resource master, 7-8
 - shared resources, 7-8

DSS (DECnet System Services), 1–2

DSSI, 1–5

adapters, 4–5

features, 4–5

load sharing, 4–7

multiple, 4–7

multiple access paths, 5–14

multiple DSSI VAXcluster (*fig*), 6–6

shadow sets (*fig*), 5–14

VAXcluster (*fig*), 1–5, 1–6

VAXcluster system, 1–5, 6–4

DSSI VAXcluster configuration

See VAXcluster configuration

See VAXcluster system

Dual-host configuration

(*fig*), 6–4

Dual-host MicroVAX system

See VAXcluster system

Dual-ported disk

Ethernet VAXcluster (*fig*), 5–15

Mixed-interconnect VAXcluster (*fig*), 5–15

E

Environmental protection, 7–4

Ethernet, 1–4

adapter, 4–3, C–3

advantages, 4–2

LAN Bridge 200, 4–8

LAN Traffic Monitor VMS, 4–8

load balancing, 4–8

multiple, 4–7

multiple adapters, 4–2

throughput, 4–2

traffic considerations, 4–2

traffic management, 4–8

VAXcluster system, 6–7

Ethernet VAXcluster configuration

See also VAXcluster configuration

See VAXcluster configuration

dual-ported disk (*fig*), 5–15

(*fig*), 1–4

F

Failover, 7–3

See also Automatic failover

between interconnects, 4–7

DECnet VAXcluster alias, 7–3

dual-host MicroVAX, 6–4

interconnects, 7–3

multiple access paths, 5–13

multiple access paths to disks, 2–3

print and batch queues, 7–3

storage devices, 7–3

VAXcluster LAT service, 7–3

volume shadowing, 7–3

Failure analysis

multiple adapter, 7–12

FDDI

advantages, 1–7

throughput, 4–6

VAXcluster system, 6–7

FDDI VAXcluster configuration

(See) VAXcluster configuration

File

read/write access, 1–2

shared access, 7–8

storage requirements, 5–5

H

HSC

attributes, 5–24, C–5

availability, 5–14

backup, 7–9

configuring subsystems, 5–14

performance considerations, 5–10

terminal, 7–11

I

I/O

application requirements, 5–5

CPU characteristics, 3–7

entry-level CPU performance, 3–7

fault-tolerant CPU performance, 3–7

mid-range and high-end CPU performance, 3–8

scaling performance, 3–2

SDI controller rates, 5–24, C–5

VAXstation performance, 3–7

I/O capacity

disk server, 7–5

InfoServer 150 Software, 5–11

Interconnect, 1–3

availability, 5–13

characteristics, 4–1

choosing, 4–1

components, 4–1

DSSI, 4–5

Ethernet, 4–2

failover, 7–3

FDDI, 4–6

multiple CIs, 4–7

multiple DSSIs, 4–7

multiple Ethernet, 4–7

multiples of the same type, 4–7

options for CPUs, 3–8

redundancy, 4–7

supported in a VAXcluster system, 1–2

throughput, 4–1

types, 4–2

ISE (Integrated Storage Element), 1–5, 5–21

L

LAN Bridge 200, 4–8
LAN Traffic Monitor, 4–8, 7–13
Load balancing
 Ethernet, 4–8
Load sharing
 CI, 4–7
 DSSI, 4–7

M

Management tools, 7–11
 CLUSTER_CONFIG.COM, 7–11
 Data Center Monitor, 7–13
 DEC Capacity Planner, 7–13
 DECalert, 7–13
 DECcmds, 7–13
 DECmcc, 7–15
 DECperformance Solution, 7–13
 DECscheduler, 7–13
 LAN Traffic Monitor, 7–13
 optional products, 7–13
 Remote System Manager, 7–14
 Show Cluster Utility, 7–12
 System Management Utility, 7–12
 VAX Distributed File Service, 7–14
 VAX Distributed Queuing Service, 7–14
 VAX Remote Environmental Monitoring
 Software, 7–14
 VAX Software Performance Monitor, 7–14
 VAX Storage Library System, 7–14
 VAXcluster Console System, 7–14
 VAXcluster MDF, 7–13
MASSBUS disk
 See Disk
Memory
 determining requirements, 3–4
 for storage, 5–2, 5–5
 sharing with SMP, 3–2
Mixed interconnect, 1–8, 4–6
 advantages, 1–8
Mixed-interconnect VAXcluster configuration
 See also VAXcluster configuration
 dual-ported disks (*fig*), 5–15
MOP server, 1–3
MSCP server, 1–2
Multiple adapter
 failure analysis, 7–12
Multiple adapter VAXcluster configuration
 (*fig*), 6–8
Multistream work load, 1–2

N

Node numbering
 DSSI, 6–6

O

Overhead
 CPU, 4–1

P

Performance
 characteristics of CPUs, 3–4 to 3–7
 CPU I/O, 3–7 to 3–8
 disk fragmentation, 5–19
 disk utilization, 5–19
 entry-level CPU, 3–5
 fault-tolerant CPU, 3–5
 HSC, 5–10
 I/O, 3–2
 interaction of CPUs, 3–3
 mid-range and high-end CPU, 3–5
 new disk technology, 5–11
 tape, 5–11
 tertiary storage, 5–11
 VAX system, 3–5, C–1
 VAXstation CPU, 3–5
Print and batch queue
 failover, 7–3
 sharing clusterwide, 1–2
Print service, 7–10
Printer, 7–10, C–7
 availability and redundancy, 2–3
 band, 7–10, C–7
 connecting, 7–11
 dot-matrix, 7–10, C–7
 ink jet, 7–10
 laser, 7–10
 PostScript, 7–10, C–7
 print wheel, 7–10, C–7
Processor, 3–1

Q

Quorum, 7–4
Quorum availability
 See Availability

R

Rebooting, 7–10
Redundancy
 BACKUP copies, 5–18
 CI, 4–3
 CPU, 7–2
 data, 2–3, 5–13, 7–2

Redundancy (cont'd)
disks, 5-13, 5-16
interconnect, 4-7, 7-2
multiple access paths, 7-2
printer, 2-3
site, 7-2
star coupler, 7-2
system disk, 5-15
volume shadowing, 7-2
Remote System Manager, 7-14
RMS (Record Management Services), 1-2

S

Satellite
function in a VAXcluster configuration, 1-3
Satellite node, 1-3
SCS (system communication services), 1-2
SCSI bus, 4-8
SDI controller, 5-24, C-5
Security, 2-3
VAXcluster system, 2-3
Server
function in a VAXcluster configuration, 1-3
Show Cluster Utility, 7-12
Site redundancy, 5-15, 7-2
SMP (symmetrical multiprocessing) system, 1-2
disk server, 7-6
Space requirements, 2-3
SPD (Software Product Description), A-1, B-1
Specifications for mature products, C-1
Star Coupler, 4-3
State transition, 7-5
phases, 7-5
Storage
See also Disks, Tapes
calculating growth, 5-6
capacity work sheet, 5-6
complex configuration (*fig*), 5-3
disk fragmentation, 5-19
estimating requirements, 5-5
hierarchy, 5-1
See also Storage hierarchy
hierarchy (*fig*), 5-2
performance and availability work sheet, 5-9,
5-12
requirements, 2-2
selecting for capacity requirements, 5-8
simple configuration (*fig*), 5-3
subsystem design, 5-1
Storage bus
CPU options, 3-8
Storage device
arrays, 5-22
disks, 5-21, C-4
failover, 7-3
library, 5-23
MASSBUS, C-6

Storage device (cont'd)
supported in a VAXcluster system, 1-2
tapes, 5-23, C-5
UNIBUS, C-5
Storage hierarchy, 5-2
CPU selection, 5-8
designing, 5-4
estimating capacity requirements, 5-5
factors affecting design, 5-20
management, 5-18
memory, 5-2, 5-5
performance requirements, 5-8, 5-10
primary, 5-2, 5-5
secondary, 5-2, 5-5
tertiary, 5-3, 5-7
Storage requirements, A-1
availability, 5-12
databases, 5-6
growth, 5-6
layered products, 5-5
paging, swapping, dump files, 5-5
third-party products, 5-6
user data, 5-6
utilities and data, 5-5
VMS operating system, 5-5
StorageTek 4400 ACS, 5-11
System disk
availability, 5-15, 7-9
booting load, 7-10
configuring, 7-9
contents, 7-9
creating duplicate, 7-11
redundancy, 5-15
System Management Utility, 7-12
System requirements, 2-2

T

Tape
See also Storage device
attributes, 5-23, C-5
capacity, 5-7
performance, 5-11
TA-series, 5-23
TF-series, 5-23, C-5
TK-series, 5-23
TS-series, 5-23
TU-series, 5-23, C-5
TZ-series, 5-23
Terminal server, 1-3
Throughput
CI, 4-4
definition, 4-1
Ethernet, 4-2
FDDI, 4-6
TMSCP server, 1-2

Troubleshooting
multiple adapter, 7-12

U

Utility
Backup, 7-8
Show Cluster, 7-12
storage requirements, 5-5
System Management, 7-12

V

VAX Distributed File Service, 7-14
VAX Distributed Queuing Service, 7-14
VAX Remote Environmental Monitoring Software,
 7-14
VAX Software Performance Monitor, 7-14
VAX Storage Library System, 5-20, 7-14
VAX Unit of Processing, 4-4
VAXcluster configuration, 1-3
 CI, 6-1
 CI (*fig*), 1-5
 comparison, 1-4
 DSSI, 1-5, 6-4, 6-5
 DSSI (*fig*), 1-5
 DSSI rules for VAX 6000, 6-5
 Ethernet, 1-4, 6-7
 Ethernet (*fig*), 1-4
 FDDI, 1-7, 6-7
 FDDI (*fig*), 1-7
 mixed-interconnect, 1-8, 4-6
 mixed-interconnect (*fig*), 1-9
 multiple DSSI (*fig*), 1-6
 rules, 6-1
 types, 1-3
VAXcluster Console System, 7-14
VAXcluster multi-datacenter facility, 7-13
VAXcluster SPD
 See SPD
VAXcluster system
 See also Management tools
 See also VAXcluster configuration
 adding a CPU to, 7-11
 availability, 2-2, 7-1
 booting, 7-10
 changing CPU characteristics, 7-11
 clusterwide system management, 1-2
 common-environment, 7-9
 communications, 1-2
 components, 1-2
 cost and growth, 2-3
 creating duplicate system disk, 7-11
 definition, 1-1
 design considerations, 7-1
 determining requirements of, 2-1
 dual-host MicroVAX, 6-4

VAXcluster system (cont'd)
environment overview, 1-1
features, 1-1
I/O performance, 7-7
membership, 1-2
multiple-environment, 7-9
number of CPUs in, 3-3
overall requirements, 2-2
print services, 7-10
redundancy, 7-2
removing a CPU from, 7-11
security, 2-3
sharing resources in, 1-2
space requirements, 2-3
state transition, 7-5
trade-offs, 7-1
VAXcluster MDF configuration, 7-13
VMS operating system, 1-2
 BACKUP, 7-9
 online storage requirements, 5-5
VMS tools and products, 7-13
Volume shadowing, 5-16
 availability, 5-13
failover, 7-3
rapid backup, 5-18
redundancy, 7-2
shadow set rules, 5-16
shadow sets, 5-16
spare repair, 5-17

VUP

See VAX Unit of Processing