Software Product Description

PRODUCT NAME: VMScluster Software Version 1.5 for OpenVMS AXP **SPD 42.18.00**

DESCRIPTION

VMScluster Software for OpenVMS AXP is an OpenVMS AXP System Integrated Product (SIP) that provides a highly integrated OpenVMS computing environment distributed over multiple Alpha AXP CPUs. This environment is called a VMScluster system, and can be extended to include VAX family CPUs running VAXcluster Software for OpenVMS VAX. Refer to the VAXcluster Software for OpenVMS VAX, Version 5.5-2 Software Product Description (SPD 29.78.06) for details of the VAXcluster software product.

Except where noted, all references to OpenVMS and VMScluster features apply equally to VAX systems and Alpha AXP systems when configured together in a single VMScluster.

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

Applications running on one or more CPUs in a VMScluster system access shared resources in a coordinated manner. VMScluster software components synchronize access to shared resources, allowing multiple processes on any CPU in the VMScluster to perform coordinated shared data updates.

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

All VMScluster systems have the following software features in common:

- The OpenVMS operating system and VMScluster software allow all CPUs to share read and write access to disk files in a fully coordinated environment. Application programs can specify the level of cluster-wide file sharing that is required; access is then coordinated by the OpenVMS Extended QIO Processor (XQP) and Record Management Services (RMS).
- Shared batch and print queues are accessible from any CPU in the VMScluster system. The OpenVMS queue manager controls cluster-wide batch and print queues, which can be accessed by any CPU. Batch jobs submitted to cluster-wide queues are routed to any available CPU so the batch load is shared.
- The OpenVMS Lock Manager System Services operate in a cluster-wide manner. These services enable reliable coordinated access to any resource and provide signaling mechanisms, at the system and process level, across the whole VMScluster system.
- All physical disks in a VMScluster system can be made accessible to all CPUs. Additionally, VAX systems can serve TMSCP-compliant tapes to all systems in a VMScluster.
- Process information and control services are available cluster-wide to application programs and system utilities.
- Configuration command procedures assist in adding and removing CPUs and in modifying their configuration characteristics.
- The dynamic Show Cluster Utility displays the status of VMScluster hardware components and communication links.
- Standard OpenVMS system and security features work in a cluster-wide manner such that the entire VMScluster system operates as a single security domain.
- The VMScluster software balances the interconnect I/O load in VMScluster configurations that include multiple interconnects.
- Multiple VMScluster systems can be configured on a single or extended Local Area Network (LAN).

Definitions

The following terms are used throughout this SPD.

- CPU (Central Processing Unit) — An Alpha AXP family or VAX family computer running the OpenVMS operating system. A CPU comprises one or more processors and operates as a VMScluster node. A VMScluster node can be referred to as VMScluster member.
- Disk server — A CPU that makes disks to which it has access available to other CPUs in the VMScluster system, using the OpenVMS MSCP Server.
- Maintenance Operations Protocol (MOP) server — A CPU that services satellite boot requests, using DECnet for OpenVMS AXP software, to provide the initial Local Area Network (LAN) down-line load sequence of the OpenVMS operating system and VMScluster software. At the end of the initial down-line load sequence, the satellite uses a disk server to perform the remainder of the OpenVMS booting process.
- Satellite — A CPU that is booted over a LAN using a MOP server and disk server.
- Boot server — A CPU that is both a MOP server and a Disk server. A Boot server can fully service Satellite boot requests.
- Mixed interconnect VMScluster system — A VMScluster system that uses more than one type of interconnect for VMScluster communication.
- Mixed architecture VMScluster system — A VMScluster system that is configured with both VAX and Alpha AXP CPUs.
- Hierarchical Storage Controller (HSC) — An intelligent mass storage controller subsystem that connects to the CI.
- Star Coupler — A common connection point for all CI-connected CPUs and HSCs.
- MSCP (Mass Storage Control Protocol) — A message-based protocol for controlling Digital Storage Architecture (DSA) disk storage subsystems. The protocol is implemented by the OpenVMS DUDRIVER device driver.
- TMSCP (Tape Mass Storage Control Protocol) — A message-based protocol for controlling DSA tape storage subsystems. The protocol is implemented by the OpenVMS TUDRIVER device driver.

Interconnects

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

- Computer Interconnect (CI)
- Ethernet
- Digital Storage System Interconnect (DSSI)

Ethernet is an industry-standard general purpose communications interconnect that can be used to implement a Local Area Network (LAN).

CI and DSSI are highly optimized special purpose interconnects for CPUs and Storage subsystems in VMScluster configurations.

Additionally, VAX CPUs may use FDDI as an interconnect in a VMScluster system. Refer to the VAXcluster Software, Version 5.5-2 Software Product Description (SPD 29.78.06) for further information.

Configuration Rules

The following rules apply to Alpha AXP CPUs in a VMScluster system:

- A maximum of three DEC 7000/10000 CPUs can be configured in a VMScluster.
- A maximum of two DEC 4000 CPUs can be configured in a VMScluster.
- A maximum of twelve DEC 3000 CPUs can be configured in a VMScluster system.
- A maximum of two CI adapters can be configured on a DEC 7000/10000 CPU.
- A maximum of two LAN adapters can be configured on an Alpha AXP CPU.
- DEC 7000/10000 CPUs cannot be connected to the DSSI.
- DEC 4000/7000/10000 CPUs require a system disk that is accessed via a local controller or through a local CI or DSSI connection. These CPUs cannot be configured to boot as VMScluster satellite nodes.

The following rules, and those defined in the VAXcluster Software, Version 5.5-2 Software Product Description (SPD 29.78.06), apply when VAX CPUs are configured in a mixed architecture VMScluster system. Where rules in the VAXcluster Software SPD conflict with rules in this SPD, this SPD takes precedence.

- When DEC 7000/10000 CPUs share a CI with VAX CPUs, the total number of CPUs (Alpha AXP plus VAX) on that interconnect cannot exceed eight.
- When DEC 4000 CPUs share a DSSI with VAX CPUs, the total number of CPUs (Alpha AXP plus VAX) on that interconnect cannot exceed three.
- DEC 4000 CPUs can coexist on a DSSI with VAX 4xxx, VAX 6xxx, VAX 7xxx, and VAX 10xxx CPUs.
- DEC 4000 CPUs cannot coexist on a DSSI with VAX 3xxx or VAX 9xxx CPUs.
- DEC 4000 CPUs cannot coexist on a DSSI with the KFQSA adapter.
- A VAX CPU and an Alpha AXP CPU cannot boot from a common system disk.
- Cross architecture satellite booting is not supported. Alpha AXP satellites must boot from an Alpha AXP boot server, and VAX satellites must boot from a VAX boot server.
- Alpha AXP CPUs cannot access shadow or stripe sets created on VAX CPUs by the Volume Shadowing for OpenVMS VAX or VAX Disk Striping products.

In addition, all VMScluster configurations must adhere to the following general rules.

- A maximum of 96 CPUs can be configured in a VMScluster. This limit is the sum of Alpha AXP CPUs plus VAX CPUs in the total VMScluster configuration.
- A maximum of two Star Couplers can be configured in a VMScluster.
- Every CPU in a VMScluster system must be connected to every other CPU via any of the supported VMScluster interconnects.
- All CPUs connected to a CI or DSSI must be configured as VMScluster members. VMScluster members configured on a CI or DSSI will become members of the same VMScluster (this is imposed automatically by the VMScluster software).
- 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 an OpenVMS 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 separate CPUs, and the CPUs must be of the same architecture.
- CPUs that use an Ethernet for VMScluster 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.
- All LAN paths used for VMScluster communication must operate, at a minimum, with approximately 10 megabits per second throughput and low latency.
- A DECnet communication path must exist between all nodes in a VMScluster system.
- A single time zone setting must be used by all CPUs in a VMScluster system.
- A VMScluster system can be configured with a maximum of one Quorum Disk. A Quorum Disk cannot be a member of an OpenVMS volume set, or of a shadow or stripe set created by the Volume Shadowing for OpenVMS or VAX Disk Striping products.

Recommendations

The optimal VMScluster 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
- Interconnect and adapter throughput and latency characteristics
- Disk and tape I/O capacity and access time
- Number of disks and tapes being served
- Interconnect utilization

Digital recommends VMScluster system configurations based on its experience with the VMScluster 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 VMScluster system, consider the following recommendations:

- VMScluster CPUs should be configured using interconnects that provide appropriate performance for the required system usage. In general, use the highest performance interconnect possible. CI and DSSI are the preferred interconnects between DEC 4000, 7000, and 10000 CPUs. For performance reasons,

configuring VMScluster systems containing multiple DEC 4000, 7000, and 10000 CPUs using Ethernet as the only interconnect should be avoided.

- While VMScluster systems can include any number of system disks, system performance and management overhead should be considered in determining their number and location. It is important to recognize that while the performance of multiple system disk configurations may be higher than with a single system disk, system management efforts increase in proportion to the number of system disks.
- Digital recommends that all CI connected CPUs be connected to at least one common Star Coupler. When CPUs are configured with multiple CIXCD CI adapters it is recommended that each CIXCD be connected to a different Star Coupler.
- Data availability and I/O performance are enhanced when multiple VMScluster nodes have direct access to shared storage; when possible, configure systems to allow direct access to shared storage in favor of OpenVMS MSCP Served access. Multi-access DSSI- and HSC-based storage provides higher data availability 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.
- VMScluster 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 VMScluster interconnects (CI, DSSI, Ethernet) are supported.
- To enhance availability, LAN-based and mixed interconnect VMScluster systems should include multiple boot servers. When a server fails in configurations that include multiple servers, satellite access to multipath disks will fail over to another path. Disk servers should be the most powerful CPUs in the VMScluster and should use the highest bandwidth LAN adapters available.
- When a LAN-based VMScluster 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 VMScluster configurations can be simplified with the aid of the OpenVMS LAVC$FAILURE_ANALYSIS program, which is available in the SYS$EXAMPLES directory.
- VMSclusters are sensitive to LAN traffic levels. The average LAN segment utilization should not exceed 60 percent for any 10-second interval. Nodes may leave the VMScluster if they cannot properly exchange control messages every three seconds. If the overall level of LAN network traffic becomes excessive, bridges can be used to localize VMScluster traffic, either by locating the VMScluster system behind a bridge, or by placing bridges between groups of VMScluster members.
- When under heavy network load, bridges are subject to packet loss and retransmission, due to congestion. In a VMScluster 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.

HARDWARE SUPPORT

Supported CPUs

Any Alpha AXP CPU, as listed in the OpenVMS AXP Operating System, Version 1.5 Software Product Description (SPD 41.87.01), can be used in a VMScluster.

Any Alpha AXP CPU, as listed in the OpenVMS AXP Operating System, Version 1.5 Software Product Description (SPD 41.87.01), can be configured as a VMScluster satellite node, with the exception of DEC 4000/7000/10000 CPUs.

Supported LAN Adapters

Table 1 shows the types of Local Area Network (LAN) adapters supported by VMScluster software.

Table 1

Bus	Ethernet
XMI	DEMNA
TURBOchannel	PMAD
Integral	LANCE,TGEC

Supported Peripheral Options

Alpha AXP CPUs in VMScluster systems can use all peripheral options supported by the OpenVMS AXP operating system. Refer to the OpenVMS AXP Operating System, Version 1.5 Software Product Description (SPD 41.87.01) for further information. VAX CPUs in VMScluster systems can use all peripheral options supported by the OpenVMS VAX operating system. Refer to the OpenVMS VAX Operating System, Version 5.5-2 Software Product Description (SPD 25.01.37) for further information.

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 eight, regardless of size.

HSC Subsystems

VMScluster software supports all models of the HSC family of intelligent mass storage controllers. The following rules apply for HSC subsystems:

- The following minimum software revisions are required for HSC subsystems:
  - — HSC Software V4.1 for the HSC50
  - — HSC High Performance Software V8.1 for the HSC40, HSC60, HSC65, HSC70, HSC90, and HSC95
- All HSC device ports can be used for disk storage. The maximum number of ports that can be used for tapes is twenty-four for the HSC70, HSC90 and HSC95; twelve for the HSC40, HSC60, and HSC65; eight for the HSC50.

SOFTWARE REQUIREMENTS

- OpenVMS AXP Operating System V1.5

  Refer to the OpenVMS AXP Operating System, Version 1.5 Software Product Description (SPD 41.87.01) for further information.

  Note: The OpenVMS AXP operating system, V1.5 can coexist in a VMScluster configured with VAX CPUs that are running the OpenVMS VAX operating system, V5.5-2. VAX systems running OpenVMS VAX operating system, A5.5-2 must be upgraded to V5.5-2 when included in a mixed architecture VMScluster system. Only one version of OpenVMS can exist on a system disk. Note that cross-architecture upgrades cannot be performed; a VAX system disk cannot be upgraded to an Alpha AXP system disk, and vice versa.
- DECnet V1.5 for OpenVMS AXP

  All VMScluster CPUs require a DECnet for OpenVMS AXP license. If the Cluster Alias feature is required, at least one VMScluster member must install an Extended Function license.

  Refer to the DECnet Version 1.5 for OpenVMS AXP Software Product Description (SPD 42.25.01) for further information.

OPTIONAL SOFTWARE

For information on VMScluster support for optional software products, refer to the *VMScluster Support* section of the Software Product Descriptions for those products.

Optional products that are particularly useful in VMScluster systems include:

- DECram for OpenVMS V1.1 (SPD 34.26.xx)
- VAXcluster Console System V1.4 (SPD 27.46.xx)

GROWTH CONSIDERATIONS

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

DISTRIBUTION MEDIA

CD-ROM

ORDERING INFORMATION

Software Licenses: QL-MUZA*-AA

Software Media & Documentation: QA-MT1AA-HA

Software Product Services: QT-MUZA*-**

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

DOCUMENTATION

OpenVMS AXP customers automatically receive the *VMScluster Systems for OpenVMS* manual, in on-line viewable format, on the OpenVMS AXP Operating System, Version 1.5 CD-ROM. In addition, the *VMScluster Systems for OpenVMS* manual is included in the OpenVMS AXP Operating System, Version 1.5 hardcopy documentation, as part of the Standard Documentation Set.

Refer to the OpenVMS AXP Operating System, Version 1.5 Software Product Description (SPD 41.87.01) for additional information on OpenVMS AXP documentation and ordering information.

SOFTWARE LICENSING

A VMScluster software license is required for each Alpha AXP CPU in a VMScluster system. A VAXcluster software license is required for each VAX CPU in a VMScluster 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.

License Management Facility Support

The VMScluster Software product supports the Digital 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 OpenVMS AXP Operating System Software Product Description (SPD 41.87.xx) or documentation set.

SOFTWARE PRODUCT SERVICES

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

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.

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

™ The DIGITAL Logo, Alpha AXP, AXP, CI, DEC, DECnet, Digital, HSC, HSC40, HSC50, HSC60, HSC70, HSC90, MSCP, OpenVMS, RA, TA, TMSCP, TURBOchannel, VAX, VAXcluster, and VMScluster are trademarks of Digital Equipment Corporation.