CHARON-VAX application note

AN-022 Building VMS disk cluster systems with CHARON-VAX

Author:	Software Resources International	Date:	Revised 6 September 2004

Applies to:	CHARON-VAX/XM and CHARON-VAX/XL for Windows build 3.0.3 or later and VAX/VMS version 5.5-2H4 or later
Host OS:	Windows 2000 / XP Professional, Windows 2003 Server

VAX/VMS clusters can work in different node architectures. The "poor-man's" version is an LAVC cluster, where the interconnecting Ethernet handles all cluster traffic (management and data) between the nodes. The advantage of an LAVC cluster is simplicity (no shared disk storage hardware is required); the disadvantage is low data throughput, limited by the 10/100 Mbps Ethernet interconnect.

Much higher data access rates in a VMS cluster are achieved by direct access from the VAX/VMS nodes to shared storage devices, with Ethernet handling only the distributed VMS lock manager (access management) traffic. The traditional VAX hardware solution for such a cluster was the use of DSSI or CI disk controllers and storage shelves. Unfortunately, DSSI or CI hardware is hard to get, not supported by Windows host systems, and is limited in throughput (several Mbytes/sec) compared to current technology.

The ability of several VAX emulator models in CHARON-VAX/XM and CHARON-VAX/XL (notably the 4000-106 and the 4000-108 emulations) to map emulated MSCP drives to host-based disk images, host SCSI disks or iSCSI devices permits the emulation of low cost, high performance VMS clusters with direct disk sharing. The necessary VMS system configurations and the related host hardware requirements are described in this application note.

Note: Direct disk sharing does not work with emulated VAX SCSI disks as used for instance in the emulation of a VAX 3100 system. The emulated SCSI controllers in those systems do not support Tagged Command Queuing (TCQ) that is required to guarantee order of execution of SCSI commands.

The following description assumes the use of the VAX 4000-106 or 4000-108 models available in CHARON-VAX/XM and CHARON-VAX/XL (Plus) version 3.0.3 or later. These models provide an implementation of an MSCP disk controller that can work with either disk container files, physical SCSI drives (seen by the VAX environment as MSCP drives) or virtual disks created by the Microsoft iSCSI Initiator.

CHARON-VAX application note

The emulated MSCP disks can be shared and accessed directly as in a hardware VAX DSSI cluster, but with an access speed reflecting modern disk hardware. The two practical alternatives are a physical SCSI connection or the use of iSCSI.

Physical SCSI connections to the Windows hosts require a multi-port SCSI storage shelf and is most practical for a two-node cluster.

Note: When you use a shared SCSI storage shelf, avoid Windows-based RAID. Depending on the implementation of the RAID synchronization, VMS access to logical RAID drives can be slow and cause VMS disk offline errors. Use the standard VMS redundancy mechanisms like shadowing instead.

The use of iSCSI is an interesting alternative that offers a wide price / performance range due to cost-effective GB Ethernet and a choice of iSCSI hardware or software target implementations. An iSCSI storage subsystem (target) can support VMS clusters with many nodes.

As long as simultaneous hardware connectivity between the host systems and the shared disks is available, the number of clustered CHARON-VAX nodes is only limited by VMS cluster limitations, but VMS clusters of more than three nodes have not yet been tested.

VAX/VMS installation and configuration

Note that VAX/VMS version 5.5-2H4 is the oldest version that can be used for clustering with the VAX 4000-106 or 4000-108 models in CHARON-VAX/XM and CHARON-VAX//XL. In the tested configurations, each node had its own - not shared - system disk on its own host system configured as a Windows disk image or physical disk drive (i.e. \\.\PhysicalDriveX).

The installation of the VAX/VMS operating system for each of the CHARON-VAX/XM and CHARON-VAX/XL instances follows the same process as for a standalone CHARON-VAX installation, while observing the naming requirements for the shared disks (see below). After installation of each instance of CHARON-VAX, edit the configuration file to identify the VMS system disk (image, or physical disk drive, not shared), other not-shared disks /disk images, other peripherals and the shared MSCP disks.

Each of the cluster instances has its own dedicated system disk containing a copy of the VAX/VMS system. Since all those system disks can be accessible from all CHARON-VAX/XM and CHARON-VAX/XL instances **they must have different names**. Note that the disk volume label must also be unique. For example DUA0 boots the first instance and DUA1 boots the second instance.

Configuration of shared disks

By following these guidelines, each instance of VAX/VMS should boot properly as a standalone system. Make sure you boot each instance from the proper location (i.e. its own

configuration data) by using clear configuration file naming. Since the emulator console port can be defined as a Telnet session, you can boot all nodes from a single system using a suitable terminal emulator. In the configuration files of all CHARON-VAX instances, for all shared MSCP disks, assign identical device names and identical allocation classes so that all VAX instances see the same device with the same name and attributes (e.g. DUA0 points to the same storage container, physical disk or iSCSI target for all instances).

Note: DO NOT run any instances simultaneously until they are configured as VAX/VMS cluster nodes. Violating this rule causes shared data loss which can be difficult to detect.

It is recommended to set proper boot flags and default boot device in each CHARON-VAX instance. This will keep the booting process consistent while doing automatic startup and reboot.

Setting up the VAX/VMS Cluster

After successful installation of VAX/VMS on each of the CHARON-VAX/XM and CHARON-VAX/XL instances, they must be configured to run as VAX/VMS Cluster nodes. The VAX/VMS Cluster configuration must take place on each (!) instance running VAX/VMS.

Note: Make sure that each instance has the appropriate VMS LMF licenses available for the VAX CPU you emulate, including the LMF keys required to operate as a VAX/VMS cluster.

As the VAX/VMS instances cannot co-exist yet, the following steps must be repeated for each instance while the other instance(s) are not running:

1. Boot the instance of the CHARON-VAX/XM or CHARON-VAX/XL from the appropriate location (disk and system root).

2. Verify that the DECnet addresses on all systems are different (in particular when VMS system images are obtained from the same hardware VAX). DECnet addresses translate into physical NIC addresses. Two identical NIC addresses will cause the Ethernet segment to hang.

3. See the document **"OpenVMSCluster Systems"** for advice on how to set up cluster parameters and authorizations. See the paragraphs on cluster parameters below for some tips.

4. Load the appropriate VAX/VMS Cluster licenses.

5. Shut down (do not reboot!) the instance of CHARON-VAX/XM or CHARON-VAX/XL.

The steps listed above shall be performed on each CHARON-VAX/XM and CHARON-VAX/XL instance, while the other instances are NOT running.

CHARON-VAX application note

Cluster parameters some tips.

Read the OpenVMSCluter documentation for full details but here we include a few tips to help getting started.

Ideally set all parameters using the DCL script SYS$MANAGER:CLUSTER_CONFIG.COM

Set up cluster authorizations using
$ RUN SYS$SYSTEM:SYSMAN

SYSMAN> CONFIGURATION SET CLUSTER_AUTHORIZATION

Updates the cluster authorization file, CLUSTER_AUTHORIZE.DAT, in the directory SYS$COMMON:[SYSEXE]. (The SET command creates this file if it does not already exist.) You can include the following qualifiers on this command: /GROUP_NUMBER—Specifies a cluster group number. Group number must be in the range from 1 to 4095 or 61440 to 65535. /PASSWORD—Specifies a cluster password. Password may be from 1 to 31 characters in length and may include alphanumeric characters, dollar signs ($), and underscores (_).

SYSMAN> CONFIGURATION SHOW CLUSTER_AUTHORIZATION

       Displays the cluster group number.

SYSMAN> HELP CONFIGURATION SET CLUSTER_AUTHORIZATION

As an alternative to using the full cluster_config.com you may modify the SYS$SYSTEM:MODPARAMS.DAT file so that the VAX/VMS Cluster software is properly configured.

The modifications include the specification of the proper values for the following system parameters (the list below is an example for a two node cluster):

```
VOTES=1
EXPECTED_VOTES=2
VAXCLUSTER=1
DISK_QUORUM=""
NISCS_LOAD_PEA0=1
MSCP_LOAD=1
MSCP_SERVE_ALL=2
ALLOCLASS=1
INTERCONNECT="NI"
BOOTNODE="N"
```

ALLOCLASS may be any non-zero value but must be the same on all nodes.
If you want have a cluster with any (arbitrary) number of nodes (1, 2 … n) set
EXPECTED_VOTES to 1.

If your cluster MUST HAVE not less then N-nodes, then use value N for EXPECTED_VOTES and VOTES=1 for all nodes (without quorum disk).

Run the AUTOGEN utility to apply the changes made to the SYS$SYSTEM:MODPARAMS.DAT.

Warning: Just modifying the MODPARAMS.DAT you lose such cluster specific parameters like authorization – cluster group number and password. This information is in the binary file sys$system:CLUSTER_AUTHORIZE.DAT. As a workaround this file can be copied to all cluster nodes to have the same authorization parameters (group/password).

Please refer to the available VAX/VMS documentation for a detailed description of each of the parameters listed above, if necessary. After the five steps are finished for each of the CHARON-VAX/XM and CHARON-VAX/XL instances, all the instances can be booted simultaneously to form a VAX/VMS shared disk cluster.

Host system hardware

Three different ways to create a multi-node CHARON-VAX/XM and CHARON-VAX/XL direct access shared disk cluster have been tested. In each case each host was a dual CPU Windows 2000 or Windows XP Professional host system (or a Hyper-threading P4 system providing two logical CPUs under Windows XP) running an instance of CHARON-VAX/XM or CHARON-VAX/XL. The shared storage, represented as MSCP disks by CHARON-VAX, could be represented in the following alternate ways:

1. By disk container files on a Windows remote share on an additional Windows system, connected via 1GB Ethernet. This is a low-cost test solution with the disadvantage that if the network connection is broken the shared disk in a CHARON-VAX instance goes offline. Re-establishing the connection does not bring the disk back online. Locating a shared disk on one of the CHARON-VAX hosts leads in general to failure as the heavy system load creates a significant difference in access time between the 'local' and the remote VAX instances, causing timeouts.

2. By connecting two CHARON-VAX nodes to a dual port SCSI storage shelf. The shelf must allow true simultaneous disk access, something that not all multi-port SCSI shelves implement correctly. A storage unit that is qualified for VMS clusters (e.g. the HP MSA 1000) will work correctly. This solution is difficult to implement for a cluster with more than two CHARON-VAX nodes.

3. The use of iSCSI permits an easy implementation of multiple cluster nodes; the virtual SCSI drives created by the standard Windows iSCSI initiator can be configured to work as MSCP drives in CHARON-VAX/XM and CHARON-VAX/XL. The selection of an iSCSI target depends on the required I/O performance. While dedicated iSCSI storage hardware provides the highest performance, even a software implementation like WinTarget provides a performance exceeding that of DSSI hardware. With WinTarget on a 2.8 GHz P4 system and 1GB Ethernet, CHARON-VAX disk I/O transfer rates >12 Mbytes/sec have been measured for single file copy to a shared drive, and >2 Mbytes/sec for simultaneous copies from two CHARON-VAX nodes.

CHARON-VAX application note

In all cases, the CHARON-VAX nodes were interconnected with 100 Mbps Ethernet to handle the VMS cluster management and other network information. If concurrent database updates generate heavy lock manager traffic, a separate LAN for this traffic can be configured in VAX/VMS running on CHARON-VAX node.

Note: None of the three solutions permit the addition of VAX hardware systems to the VAX/VMS cluster. While hardware VAX systems can connect to physical SCSI drives (the only way to directly access the same disks as the CHARON-VAX cluster), those disks will not be considered by VMS running on the hardware VAX as sharable MSCP drives.

[30-18-022]