Software Resources International

CHARON-VAX application note

AN-040 CHARON-VAX clustering using the Open-e iSCSI target

Author: Robert Boers Date: 4 January 2006

Applies to: CHARON-VAX products for the Windows platform

Many VAX systems in operation are clustered to improve operational reliability or performance. The traditional VAX hardware solution for a cluster is the use of MSCP, DSSI or CI disk controllers and storage shelves. However, MSCP, DSSI or CI hardware is obsolete, not supported by the CHARON-VAX Windows host systems, and is limited in throughput compared to current technology.

The VAX emulator models that provide MSCP disk emulation (notably the 4000-106/108 in CHARON-VAX/XM/XK/XL and the CHARON-VAX/6000 product family) can map the emulated MSCP drives to host SCSI devices. The standard Windows iSCSI initiator can create local (virtual) SCSI disks that can be located on a common server and can be shared between multiple systems.

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

A disk cluster of multiple emulated VAX systems can be created by mapping their MSCP disks to these virtual SCSI disks on each Windows host system. The number of VAX nodes clustered this way is practically unlimited, only constrained by the aggregate throughput requirements.

There are many dedicated hardware iSCSI targets (servers) available that can act as a central storage and backup server for such VAX clusters, but they come at a significant cost. An alternative is to use iSCSI target emulators, but they require a host OS, setup procedures and virus protection.

An interesting alternative is provided by a new product, the Open-e (www.open-e.com) iSCSI target modules. For less than the cost of a host OS and iSCSI target software, these products (between 200 and 800 €) are solid state disk modules that contain the full iSCSI target functionality using an embedded Linux operating system. These modules are plugged into an IDE or RAID socket of a hardware PC (no operating system required) and automatically boot and configure the iSCSI target. Depending on the module type, the functionality supports multi-CPU hosts, hardware RAID, management console, fiber channel, 10 GB Ethernet, adaptive load balancing, adapter fault tolerance, automatic snapshots and more. The manufacturer can remotely update these modules for new functionality.

The configuration of the CHARON-VAX systems:

Node RDB717:	Windows Server 2003, AMD Opteron 875 (dual core 2.2 Ghz), 4 GB RAM
Network adapters:	Broadcom NetXtreme Gigabit Ethernet
	Intel(R) PRO/1000 Server Adapter
Node RDB718:	Windows XP, 2 x Xeon 2.8 Ghz, 4 GB RAM
Network adapters:	Intel(R) PRO/1000 MT Network Connection
	Intel(R) PRO/100 Network Connection

Both nodes ran CHARON-VAX/6610.

©2006 Software Resources International. This document is provided for information only and is not a legally binding offer. Software Resources International reserves the right to change the product specifications without prior notice or retire the product. The CHARON name and its logo are a registered trademark of Software Resources International.

The iSCSI target was configured as follows:

ISCSI server	AMD Athlon 64 X2 3800+ (dual core, 2 Ghz), 3 GB
Network adapters:	nVidia Corp. SK804 Ethernet Controller (rev a3)
	Intel Corp. PRO/1000 MT Desktop Adapter
Storage:	SATA II drive 200.0 GB (WD2000JS)

The Open-E iSCSI Enterprise module was used to load the iSCSI target software. The iSCSI server was connected to the Windows host computers via its two GB Ethernet network adapters (to node RDB717 with the Broadcom NetXtreme Gigabit and to node RDB718 with the Intel(R) PRO/1000 MT).

The VAX 6610 systems emulated by CHARON-VAX were connected via a separate 100 Mbit Ethernet network. On both cluster nodes OpenVMS/VAX V7.2 was used. Both VAX nodes were configured with MSCP disks and had a local system disk (1DUA0 for RDB718 and 1DUA3 for RDB717).

The shared disks were a disk volume on the iSCSI target (1DUA1), and a physical SCSI disk (1DUA2) on a SCSI bus connected to both CHARON-VAX nodes (the latter solution is for comparison, not for operational use. A host system power up will reset the SCSI bus and disconnect the shared volume from the other node).

After configuring the OpenVMS/VAX cluster (see AN-022 for the configuration procedure) the disks are visible as follows:

```
Device                  Device          Error    Volume         Free  Trans Mnt
 Name                   Status          Count     Label        Blocks Count Cnt
$1$DUA0:       (RDB718) Mounted             0  OVMSVAXSYS       735480     1   2
$1$DUA1:       (RDB717) Mounted             0  ISCSI          11096868     1   2
$1$DUA2:       (RDB717) Mounted             0  RSCSI          35114485     1   2
$1$DUA3:       (RDB717) Mounted             0  OVMS717         1851453   183   2
```

Performance measurements

CHARON-VAX performance:

	VUPs	Dhryst/sec	MIPS
CHARON-VAX/6610 2 x Xeon 2.8 GHz	54.2	126582	33.3
CHARON-VAX/6610 AMD Opteron 875	73.2	181818	40

Disk read performance test

This test reads directly logical blocks from disk (not RMS file records), using the VMS IO$_READBLK function. With increasing block size, the performance depends mostly on the IO Channel data transfer rate, not on the disk performance itself. With a 512 byte block, the difference in performance of the host systems in accessing the iSCSI volume (1DUA1) is due to the overhead for small blocks in the software iSCSI initiator. The difference is less for the physical SCSI drive (1$1DUA2). Obviously local access to the two system disks is much faster than remote access via the 100 Mbps VAX – VAX link.

Results in Kbytes/sec	1DUA0 718 Local	1DUA1 iSCSI vol	1DUA2 SCSI	1DUA3 717 Local
Read in 512 byte blocks				
From node RDB717	338	1694	3125	3080
From node RDB718	2182	893	2385	335

©2006 Software Resources International.

With an increasing block size, the iSCSI initiator overhead is reduced, and with a 50 KB block the access speed of the iSCSI volume on the iSCSI server approaches that of the directly connected physical SCSI drive (1DUA2), a 10K RPM, 160U 18 GB Fujitsu MAN3184.

Results in Kbytes/sec	1DUA0 718 Local	1DUA1 iSCSI vol	1DUA2 SCSI	1DUA3 717 Local
Read in 5120 byte blocks				
From node RDB717	834	11428	21512	22456
From node RDB718	16953	8737	18686	867
Read in 51200 byte blocks				
From node RDB717	1028	45714	53333	56888
From node RDB718	40000	42666	54468	1030

For comparison, reading a RD154 disk on a hardware MicroVAX 3600 with 51200 byte blocks results in a transfer speed of 903 Kbytes/sec.

Shared append write performance test

The shared / append write tests uses RMS records with commit after adding 100 80-byte records. It is a good measurement of the performance of the Files11/RMS system. In shared mode, the write lock of the first cluster member writing must be removed before the second cluster member can commit its data. The data shows that write locking is removed faster from the iSCSI volume than from the physical SCSI drive or the local shared volumes. This is a good database performance test.

Results in bytes/sec	1DUA0 718 Local	1DUA1 iSCSI vol	1DUA2 SCSI	1DUA3 717 Local
100 appends by 100 80-byte records				
Single access from node RDB717	20703	30030	10875	57142
Single access from node RDB718	36297	22573	10985	18735
Shared access from node RDB717 and RDB718	6135	15098	5675	13684

For comparison, **100 appends by 100 80-byte records** on a RD154 disk in a hardware MicroVAX 3600 result in a write speed of 2000 bytes/sec.

Note: It is not possible to add a hardware VAX system to the CHARON-VAX/VMS cluster described here. 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.

[30-18-040]

©2006 Software Resources International.