

Virtual Machine Migration Acceleration using Mellanox ConnectX[®]-2 EN 40Gb/s IO Adapter

Introduction	1
Results Summary	1
Test Goals	2
Test Description	2
Test Results	3
Summary	4

Introduction

Large virtualized environments including those found in cloud computing increase I/O performance requirements for applications such as VM migration. With high-performance and efficient networking, data center managers can migrate VMs to different servers faster, allowing them to meet stringent service level agreements (SLA) and lower total cost of ownership (TCO). Faster VM migration enables more jobs per second to run and therefore reduces infrastructure capital and operational expenses.

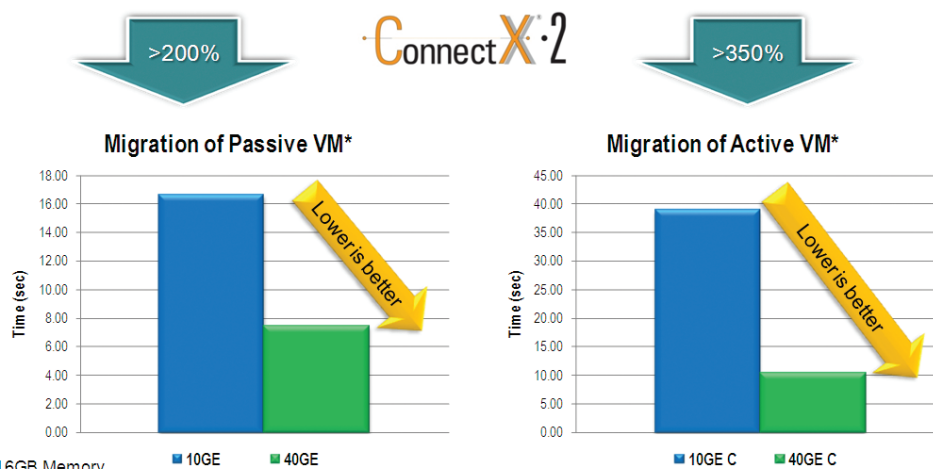
Mellanox is the first company to demonstrate VM migration at 40GigE. ConnectX[®]-2 EN 40GigE adapters are backward-compatible with 10GigE fabrics and are interoperable with available 10GigE or 1GigE and with the upcoming 40GigE switches from major switch OEMs.

Results Summary

Mellanox ConnectX-2 Ethernet adapters installed in VMware vSphere servers utilize a new 40Gb/s I/O bandwidth capacity that accelerates the migration of a single large VM (with a great amount of memory) or a large set of VMs simultaneously. The result is a greater than 200% reduction in migration time, boosting service levels by 80% while reducing capital and operational costs by more than 70%.

Mellanox ConnectX-2 EN 40Gb Ethernet adapters demonstrate record vMotion performance benchmarks of VM migration over VMware's ESX4.1.

One VM Migration with VMware vMotion over ConnectX-2 EN 10/40GigE NIC



2X to 3X Faster Execution Time over 40GigE

Test Goals

The main objective of this test is to compare between 10GbE and 40GbE vMotion over VMware ESX 4.1.

System under test

VMware® vMotion	
Software Version	Esx Server 4.1.0-258902 vSphere vCenter on Windows Server 2008 64bit
Machines	
Vendor Model	HP® DL380-G6
Processors	2xNehalem intel® Xeon X5550 (8M Cache, 2.66GHz)
Memory	Several tests performed with 24GB and several tests with 72GB
PCIe Bus	Gen2.0
40GE Adapter	Mellanox ConnectX2 VPI MHQH29B-XYR revB0 Firmware version: 2.7.700 Driver version: mlx4_en (OFED-1.5.2)
10GE Adapter	Mellanox ConnectX2 VPI MHQH29B-XYR revB0 Firmware version: 2.7.700 Driver version: mlx4_en (OFED-1.5.2)
FC Adapter	QLE2562 Driver version: vmware-esx-drivers-scsi-qla2xx_400.832.k1.27.1-1vmw
Disk	2xHP 146GB-10K-SAS (HP Model: DG0146FARVU)
OS	ESX Server 4.1.0-258902 Windows Server 2008 64bit
Operating System on VM	
Version	Linux RH5.4 64bit
Storage	
Vendor Model	EMC Clariion CX3-40, 4G-FC
Network Elements	
FC Switch	Brocade300-8G with 4G modules

- Hyper threading was disabled (no difference shown when enabling hyper threading).

Test Description

The test setup consists of two ESX4.1 machines connected back-to-back on the 40GigE and the 10GigE migration subnets, and a third machine with VMware-vCenter on top of MS-Windows-Server-2008 64bit. All machines are connected via 4Gb/s FC connection to EMC (shared storage for the ESX servers) and with 1GigE management subnet.

The test is a simple migration of one or more (8 concurrent) VMs between the two esx4.1 machines, over the 40GigE and 10GigE migration subnets.

The esx4.1 machines had two memory configurations:

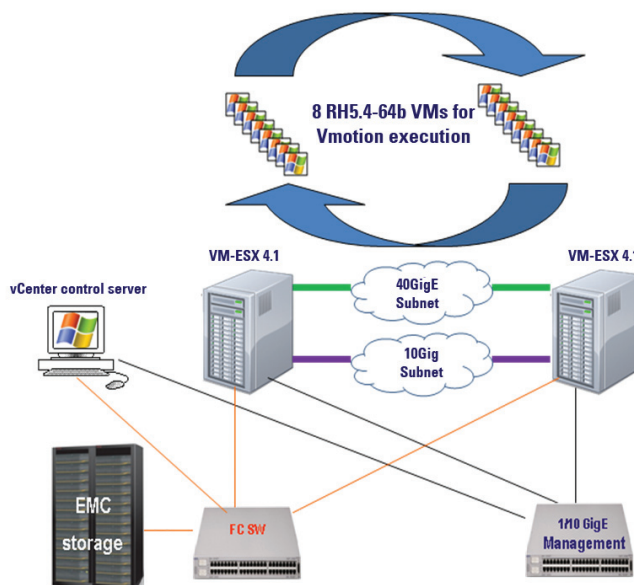
- Memory (DIMMs) total 24GB.
 - 1 VM configured as 16GB.
 - 8 VMs configured as 2GB each.
- Memory (DIMMs) total 72GB.
 - 1 VM configured as 64GB.
 - 8 VMs configured as 8GB each.

For each of the above memory configurations, several Vmotion tests were performed, using three different VM memory allocation methods:

1. Idle - all the VM's memory was not allocated (free).
2. Once - all the VM's memory was allocated once. This case simulates a situation where read-only data is loaded into the memory and is read very often.

3. Continuous - all the VM's memory was allocated repeatedly (loop of dirtying the blocks memory continuously before, during and after migration). In real life, this case imitates, for example, a situation where real time received records of data are written continuously into the memory.

Each test performed several iterations in order to verify results' stability.

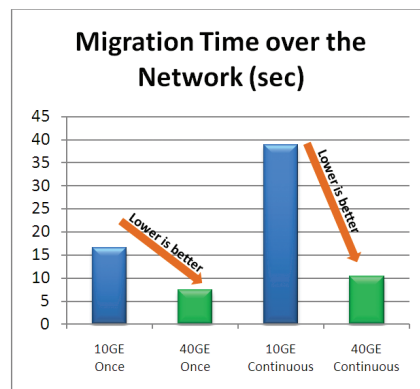
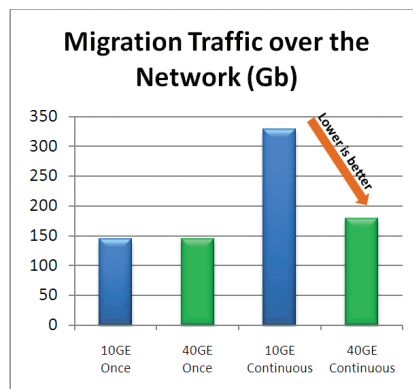


Test Results

Note: The example below mirrors the results of multiple tests conducted on several configurations.

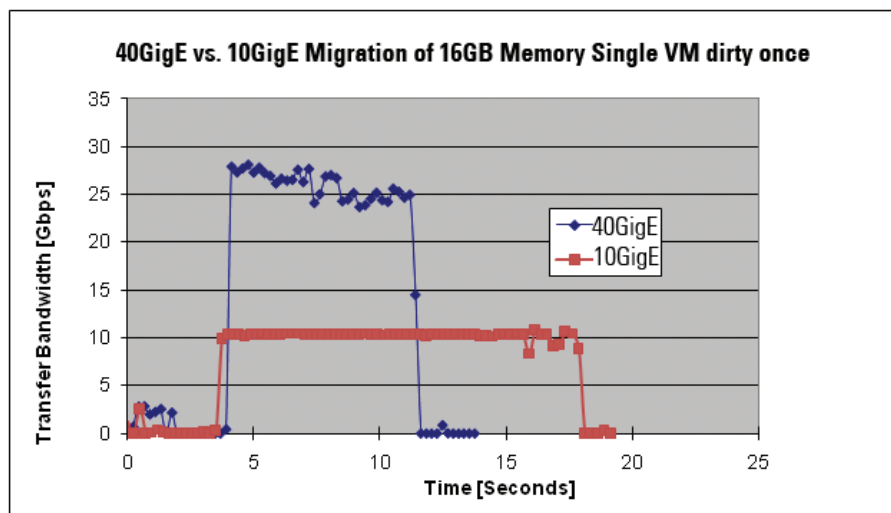
The following table displays the results of single 16GB VM migration on 24GB memory servers' setup. (Similar results were seen while migrating eight 2GB VMs on the same 24GB memory servers setup).

Memory Allocation Method	BW (Gb/s)	Time (Sec)	Data Size (Gb)
10GE Once	8.78	16.65	145.45
40GE Once	19.87	7.465	145.17
10GE Continuous	8.43	38.93	328.58
40GE Continuous	17.67	10.44	179.15



Average migration time comparison:

- In "Once" case - 40GigE is faster ~2.5X than 10GigE.
The acceleration is approximately 2.5X as expected by the PCIe 2.0 bus speed limitation (~27Gbps).
- In "Continuous" case - 40GigE is faster ~4X than 10GigE.
The acceleration is approximately 4X since more memory needs to be migrated.



The graph above shows that on machines with 24GB RAM, migration of a single 16GB memory VM allocated once reached the maximum PCIe Gen2 periodic bandwidth of ~27Gbps. Therefore, applications that frequently access the memory can benefit the most from the faster migration over 40GigE connection, resulting in smoother application operation and end user experience.

Summary

Mellanox ConnectX-2 EN 40GigE performance results show more than 200% saving on migration time achieved due to higher I/O bandwidth delivery in comparison to a 10GigE NIC. When PCIe Gen 3.0 will be available, the performance improvement will be higher using the same Mellanox adapter architecture.

Increasing I/O demands in the cloud and Web 2.0 scale computing can benefit from using Mellanox ConnectX-2 EN 40GigE NIC solution as it improves the speed at which cloud service providers can provision new users with new VM and application requirements, and meet their SLA without impacting service to existing users. The paradigm changes from bulk live migration of a VM across physical servers for disaster recovery to scheduled migration for maintenance or load balancing, thus enabling higher SLAs achievements due to the fast migration.

Data center managers can move VMs around available servers more easily through automated service level-driven policies, helping improving ROI.

Using Mellanox ConnectX-2 10/40GigE NIC VPI (Virtual Protocol Interconnect) installed on a virtualized server can interoperate with other virtualized servers that use any 10GigE NIC product. While the 40Gb/s bandwidth in the product can be dedicated to the VM migration network (such as the dedicated vMotion network), the product can also work in standard 10GigE mode for handling traffic to the LAN from VMs. In either case, the product is interoperable with available 10GigE and upcoming 40GigE switches from major switch OEMs.

ConnectX-2 excels in CPU offload, by dealing with the interconnect leaving much more CPU resources to the applications.



350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com