

Accelerating iSCSI SAN Storage with Mellanox End-to-End Networking Solution



StarWind iSCSI SAN software coupled with Mellanox Ethernet-based high speed end-to-end infrastructure accelerates storage solution networking connectivity and lowers overall power consumption

Summary

Current data centers require frequent high speed access between server and storage infrastructures for timely response to customers. Web-based service providers, such as infrastructure as a service (IaaS) or platform as a service (PaaS), provide high speed access to their customers while keeping infrastructure costs low. For a cost-efficient and higher speed requirement, 40GigE infrastructure combined with software-based iSCSI storage provides the best solution.

StarWind SAN iSCSI software running over a Mellanox ConnectX[®]-2 40GigE networking solution provides better performance, high availability (HA) and redundant iSCSI storage solutions at 40Gb/s bandwidth and higher IOPs. The installation time for StarWind iSCSI SAN is very short, and takes only a few minutes. It requires no reboot and is entirely plug-and-play with no downtime.

Solution Emerges

Under Mellanox Enterprise Datacenter's initiative, StarWind used Mellanox Zorro cluster (See: www.mellanox.com/content/pages.php?pg=edc_cluster), to run StarWind iSCSI Initiator on HP DL380 systems with Mellanox ConnectX-2 40GigE NICs. The servers were connected, as shown in the connectivity diagram below (Figure 1). Full

performance benefits were realized over the maximum bandwidth allowed by the PCIe Gen2. A record level of 27Gb/s throughput and 350K IOPs were achieved. The higher speed solution provides faster access to the storage data in an iSCSI fabric.

Solution components:

- 3 servers of Zorro (HP DL380 G6 with 2*167GB disks, 24GB RAM, 8 cores).
- 6 Mellanox HCAs with 40Gb/s per PCIe Gen2 slot single port, 2 HCAs in each server. FW: 2.7.9470. Ordering Part: MNQH19-XTR
- Connected with 3 subnets (not 40GigE switch), copper QSFP cables.
- OS: Windows 2008 Server R2.
- OFED: driver version 2.1.3.7064.

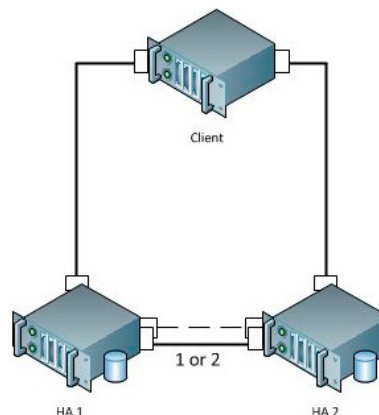


Figure 1: Connectivity Diagram



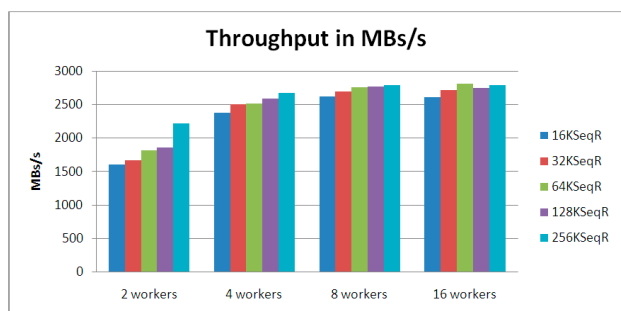
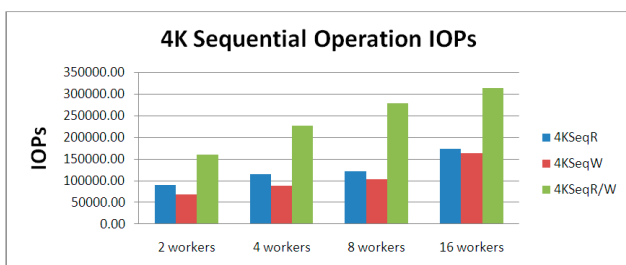
OVERVIEW

Storage networking software provider StarWind used Mellanox Zorro cluster to run StarWind iSCSI Initiator on HP DL380 systems with Mellanox ConnectX-2 40GigE NICs. Full performance benefits were realized over the maximum bandwidth allowed by the PCIe Gen2. A record level of 27Gbps throughput and 350K IOPs were achieved. The solution provides faster access to the storage data in an iSCSI fabric.

StarWind achieves record storage connectivity performance

With non-HA configuration (the only node called HA 1 and the client were performing I/O through the single 40Gb/s connection in both directions utilizing full-duplex Ethernet) cluster got 25Gb/s of an iSCSI traffic (due to PCIe Gen 2 system limitations full 40Gb/s wire speed was not reached). Tests with updated PCIe Gen3 hardware wire speed of 40Gb/s over iSCSI can be achieved, without any change in networking hardware, software and StarWind iSCSI SAN software. 300K IOPs at 25Gb/s were achieved with 16 clients using StarWind iSCSI SAN software. This represents a superior performance gain over current iSCSI HBA solutions.

The full HA version (both with HA 1 and HA 2 nodes are processing requests served under Round-Robin policy) achieves the same results as the non-HA configuration but with more workers and deeper I/O queue, which is a shortcoming for Microsoft SW initiator. Every single write went through the wire twice: first from Client to HA 1 (or 2) and then from HA 1 (or 2) to HA 2 (or 1), partner HA node before it got acknowledged as "OK" to Client. Full HA versions can reach non-HA performance with heavy and non-pulsating I/O traffic. A full set of graphs for both IOPs and Gb/s are published below. System RAM was used as a destination I/O target, which mirrors the rising trend of SSD usage in storage servers. A traditional storage stack (another PCIe controller, set of hard disks, etc.) would only add latency to the whole system and limit performance.



Accelerated cost-effective storage solutions

Consolidating over 40Gb/s networking solution for iSCSI in a virtual environment is recommended for blade server setup, as it will replace four 10GigE individual links. Trunking over 4X 10Gb/s links to achieve 40Gb/s does not work well for iSCSI, as 4KB iSCSI PDU will crawl through only two of four cards put into the trunk, leaving half of the theoretical bandwidth under-loaded. Increased latency with trunking and combined four 10Gb/s cards at a higher cost compared to single 40Gb/s makes consolidated 40Gb/s a more future proof option.

Summary - Higher ROI has been demonstrated

- 1) Fewer PCIe slots have been used to achieve the same bandwidth (one instead of four).
- 2) Less cables between cluster nodes.
- 3) Lower power solution (single silicon powered and one copper wire instead of four).
- 4) Easier to install and manage.
- 5) Superior scalability solution - higher bandwidth can be achieved using the same number of slots.

Using StarWind iSCSI SAN software version 5.6 with Mellanox ConnectX-2 EN Adapters plus 10Gb/s or higher speeds (preferably 40Gb/s), for current and future x86 servers along with PCIe Gen2 and PCIe Gen 3-enabled systems, is the most cost and power-efficient software-based iSCSI end-to-end solution for virtualized and non-virtualized environments.

Companies' Information:

StarWind Software is a global leader in storage management and SAN provider software for small to mid-sized businesses and entry-level enterprise companies. StarWind's flagship product iSCSI SAN target turns any industry-standard Windows or Linux server into a fault-tolerant, high performance and fail-safe iSCSI SAN appliance. It has been qualified to support virtualized datacenter hypervisor providers such as VMware ESX & ESXi, Microsoft Hyper-V and Citrix Xen hypervisors, as well as standard operating systems like Microsoft Windows, Linux & various UNIX environments. StarWind Software provides affordable and flexible high availability SAN solutions at a lower cost for small businesses and cost-conscious IT departments. Advanced enterprise-class features include Automated Failover and Failback, De-Duplication with variable block size, Distributed Write-Back Caching, Remote Replication across a WAN, Continuous Data Protection (CDP), Snapshots, Thin Provisioning and Virtual Tape Library management.

Mellanox Technologies (NASDAQ: MLNX, TASE: MLNX) is a leading supplier of end-to-end InfiniBand and Ethernet connectivity solutions and services for servers and storage. Mellanox products optimize data center performance and deliver industry-leading bandwidth, scalability, power conservation and cost-effectiveness while converging multiple legacy network technologies into one future-proof architecture. The company offers innovative solutions that address a wide range of markets including HPC, enterprise, mega warehouse data centers, cloud computing, Internet and Web 2.0.



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