Solution Brief

Open vStorage
The GeoScale Storage Platform

Building a Vastly Geographically Distributed Block Storage Cluster

An in-depth look at the networking of a multi-region storage cluster

Open vStorage provides a reliable, high performance, software-based scale-out storage platform, offering a block storage interface with the benefits of object storage. For one of its customers Open vStorage designed and now manages a multi-petabyte distributed block storage cluster across the U.S.A. The storage cluster offers hosting services (VMware ESXi, OpenStack, cloud Infrastructure, secondary storage for backups) with extreme resilience even in the event of a metro area failure.

THE OPEN vSTORAGE BLOCK STORAGE SOLUTION

The total storage solution actually consists of 2 interlinked Open vStorage clusters running within the same set of data centers.

- The first cluster consists of two data centers in New York and one data center in Santa Clara, California.
- The second cluster consists of two data centers in California, and one data center in New York.
- Both clusters are identically set up with two data centers in the same metro area (NY and Silicon Valley); one data center is located in another metro area.

Open vStorage uses the NC-ECC (Network Connected-Error Correction Code) algorithm that was developed in-house to store the data safely across these data centers. This way an outage, or even the complete destruction of a data center or metro area, does not result in data loss. In the event of a major disaster, one would simply boot the VMs in the metro area's second data center.

This data resilience doesn't detract from performance as the SSDs in each data center are combined into a giant distributed cache, which can contain the whole active data set. To ensure consistent write performance, the applications write to a Write Buffer on NVMe.

The Write Buffer accumulates incoming (random) IO and acknowledges the writes to the application. Only when enough data is in the Write Buffer does the buffer gets flushed to the backend, which is spread across the 3 data centers. The data in the Write Buffer itself is protected in the second data center of the same metro area.

HIGHLIGHTS

- High bandwidth storage cluster offers low latency block storage across the East and West Coasts of the US.
- Simple and linear network scalability achieved by connecting additional servers to the Mellanox Ethernet switches as ports (configurable from 10GbE to 40GbE).
- Combines cost-effective 40GbE switches and a low datacenter footprint , helping to reduce TCO.

"The network technology of Mellanox simply makes it possible to deliver a fast and scalable storage cluster which meets the requirements of our customer."

Tony Bogaert

VP Corporate Development, Open vStorage

NETWORK TOPOLOGY

From the beginning it was clear that this storage solution would require best-in-class network equipment and specialized knowledge to design the right underlying network topology. Together with the Mellanox team, the Open vStorage team designed the core network infrastructure. For ultra-low latency and easy-to-use networking, Open vStorage selected Mellanox SwitchX® Ethernet switches as spine and leaf switches. The Ethernet core switching builds on the Mellanox Virtual Modular Switch (VMS) network architecture. The Spine and Leaf implementation offers 40Gbps high-speed aggregation communication. Each data center (two data centers in each metro area) contains two Mellanox SN2100/SN2700 spine switches.

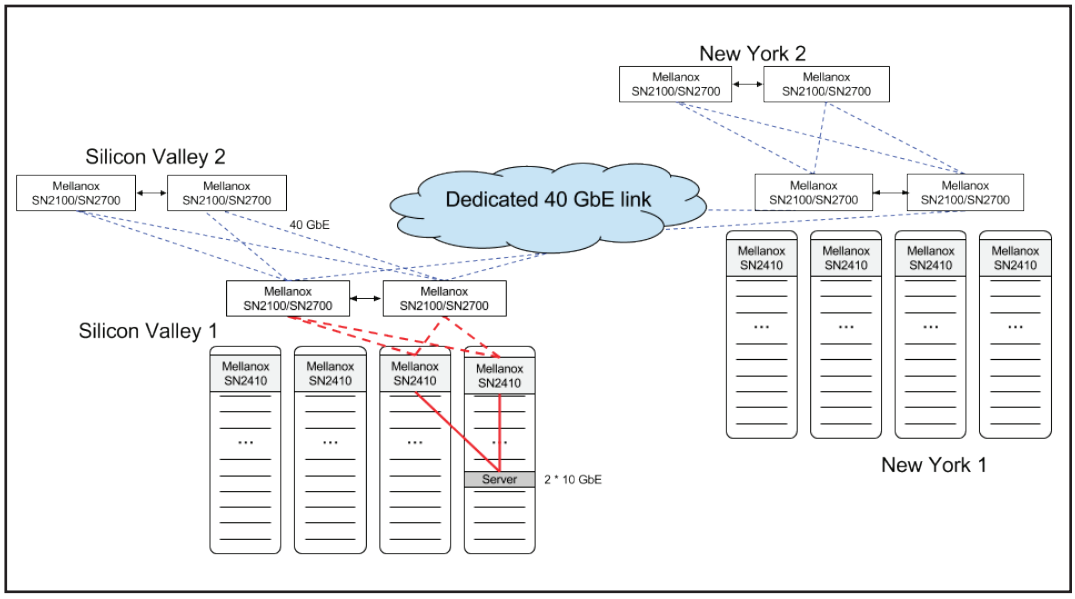

Figure 1: Co-Designed Core Network Infrastructure: Open vStorage and Mellanox

- Each rack in the data center contains a Mellanox SN2410 leaf switch.
- Each server of the storage cluster has a Mellanox ConnectX®-4 Dual-Port Adapter.
- One port of the adapter is linked to the Top Of Rack (TOR) leaf switch within the same rack while another port is linked to the TOR leaf switch in the adjacent rack.
- The ports on each server are bonded into a logical interface by means of MLAG on the switches.
- The OSPF+ECMP load balancing scheme is employed so that routing can be recovered automatically when any equipment or link fails, ensuring network resilience and reliability.

About Open vStorage

Open vStorage posesses a proven track record in data center evolution and cloud storage technologies, offering expert knowledge in cloud computing using open source technologies. Open vStorage strives to revolutionize software-defined storage, to enable large solution and service providers to continuously grow and globalize their business and compete with legacy public cloud offerings.
For more information, visit: www.openvstorage.com.

About Mellanox

Mellanox Technologies (NASDAQ: MLNX) is a leading supplier of end-to-end Ethernet and InfiniBand intelligent interconnect solutions and services for servers, storage, and hyper-converged infrastructure. Mellanox intelligent interconnect solutions increase data center efficiency by providing the highest throughput and lowest latency, delivering data faster to applications and unlocking system performance. Mellanox offers a choice of high performance solutions: network and multi-core processors, network adapters, switches, cables, software and silicon, that accelerate application runtime and maximize business results for a wide range of markets including high performance computing, enterprise data centers, Web 2.0, cloud, storage, network security, AI, telecom and financial services. More information is available at www.mellanox.com.

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

53851SB
Rev 1.0