Implementing VMware's Virtual SAN™ with Micron SSDs and the Mellanox Interconnect Solution

This document discusses an implementation of VMware Virtual SAN™ (VSAN) that supports the storage requirements of a VMware Horizon™ View™ Virtual Desktop (VDI) environment. Although VDI was used to benchmark the performance of this Virtual SAN implementation, any application supported by ESXi 5.5 can be used. VSAN is VMware's hypervisor-converged storage software that creates a shared datastore across SSDs and HDDs using multiple x86 server hosts.

To measure VDI performance, the Login VSI workload generator software tool was used to test the performance when using Horizon View. VDI performance is measured as the number of virtual desktops that can be hosted while delivering a user experience equal to or better than a physical desktop including consistent, fast response times and a short boot time. Supporting more desktops per server reduces CAPEX and OPEX requirements.

Benefits of Virtual SAN

Data storage in VMware ESX environments has historically been supported using NAS- or SAN-connected shared storage from vendors such as EMC, Netapp, and HDS. These products often have considerable CAPEX requirements and because they need specially-trained personnel to support them, OPEX increases as well. VSAN eliminates the need for NAS- or SAN-connected shared storage by using SSDs and HDDs attached locally to the servers in the cluster. A minimum of three servers are required in order to survive a server failure. Information is protected from storage device failure by replicating data on multiple servers. A dedicated network connection between the servers provides low latency storage transactions.

SSDs Boost Virtual SAN Performance

Application performance is often constrained by storage. Flash-based SSDs reduce delays (latency) when reading or writing data to hard drives, thereby boosting performance.

READ Caching: By caching commonly accessed data in SSDs READ latency is significantly reduced because it is faster to retrieve data directly from the cache than from slow, spinning HDDs. Because DRS[1] may cause VMs to move occasionally from one server to another, VSAN does not attempt to store a VM's data on a SSD connected to the server that hosts the VM. This means many READ transactions may need to traverse the network, so high bandwidth and low latency is critical.

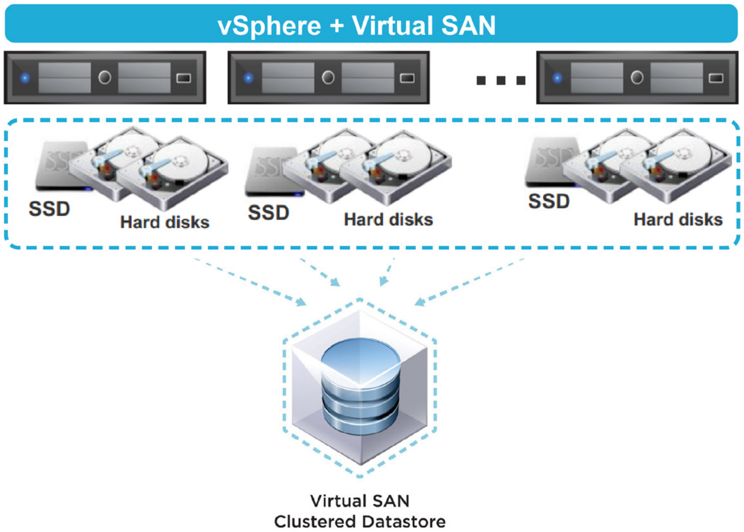

***Figure 1.** Virtual SAN*

[1] VMware's Dynamic Resource Scheduling performs application load balancing every 5 minutes.

WRITE Buffering: VSAN temporarily buffers all WRITEs in SSDs to significantly reduce latency. To protect against SSD or server failure, this data is also stored on a SSD located on different server. At regular intervals, the WRITE data in the SSDs are de-staged to HDDs. Because Flash is non-volatile, data that has not been de-staged is retained during a power loss. In the event of a server failure, the copy of the buffered or de-staged data on the other server ensures that no data loss will occur.

Dedicated Network Enables Low Latency for VSAN

Most READ and all WRITE transactions must traverse over a network. VSAN does not try to cache data that is local to the application because it results in poor balancing of SSD utilization across the cluster. Because caching is distributed across multiple servers, a dedicated network is required to lower contention for LAN resources. For data redundancy and to enable high availability, data is written to HDDs located on separate servers. Since two traverses across the network are typically required for a READ and one for a WRITE, the latency of the LAN must be sub-millisecond. VMware recommends at least a 10GbE connection.

VSAN-Approved SSD Products

VMware has a compatibility guide specifically listing I/O controllers, SSDs, and HDDs approved for implementing VSAN. Micron's P320h and P420m PCIe HHHL SSD cards are listed[2] in the compatibility list.

Tested Configuration

Three servers, each with dual Intel Xeon E5-2680 v2 processors and 384GB of memory, were used for this test. Each server included one disk group consisting of one SSD and six HDDs. Western Digital 1.2TB 10K rpm SAS hard drives were connected using an LSI 9207-8i host bus adapter set to a queue depth of 600. A 1.4TB Micron P420m PCIe card was used for the SSD. A dedicated storage network supporting VSAN used Mellanox's end-to-end 10GbE interconnect solution, including their SX1012 twelve-port 10GbE switch, ConnectX®-3 10GbE NICs and copper interconnect cables.

On the software side, ESXi 5.5.0 Build 1623387 and Horizon View 5.3.2 Build 1887719 were used. Within the desktop sessions, Windows 7 64-bit was used. Each persistent desktop used 2GB of memory and one vCPU. VDI performance was measured as the number of virtual desktops that could be hosted while delivering a user experience equal to or better than a physical desktop.

Results

Version 4.1.0.757 of the Login VSI load simulator was used for testing. This benchmark creates the workload representative of an office worker using Microsoft Office applications. The number of desktop sessions is steadily increased until a maximum is reached, in this case 450 sessions. Increasing the number of sessions raises the load on the servers and the VSAN-connected storage, which causes response times to lengthen. Based on minimum, average, and maximum response times, the benchmark software will calculate VSImax, which is their recommendation for the maximum number of desktops that can be supported. The following figure shows that using the three-server configuration, up to 356 desktops can be supported.

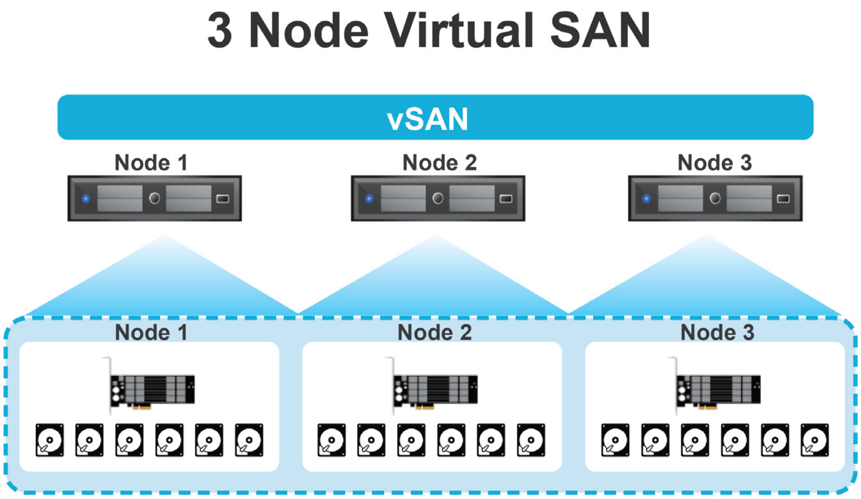

Figure 2. VSAN Test Configuration

[2] The Virtual SAN compatibility guide is located at https://www.vmware.com/resources/compatibility/search.php?deviceCategory=vsan

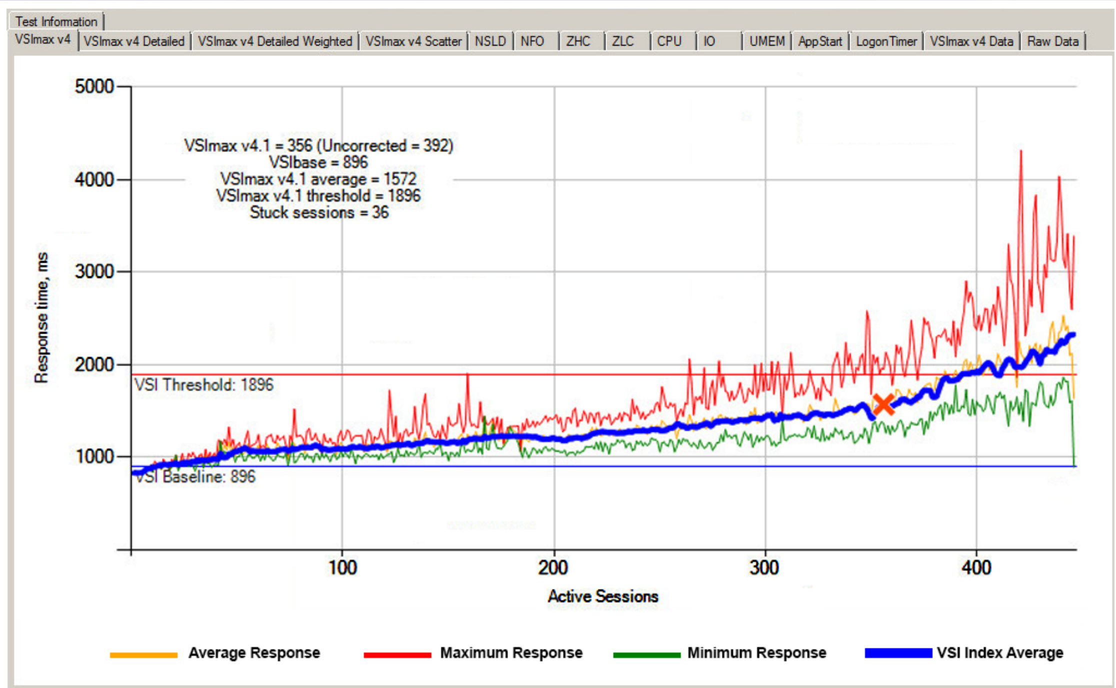

Figure 3. VSAN Results

Other critical factors in VDI environments are the times required to boot, deploy, and recompose desktops. Boot is when an office worker arrives at work and wants to access their desktop. Deploy is the creation of a desktop session, and recompose is the update of an existing session. An update may be required after a patch release has been distributed. For this test, 450 desktops were simultaneously booted, deployed, and recomposed.

- Boot: 0.7 seconds/desktop
- Deployment: 7.5 seconds/desktop
- Recompose: 9.2 seconds/desktop

Conclusion

Software-defined storage appears to be a viable alternative to SAN or NAS storage from our experience using VSAN. Using directly-attached SSDs and HDDs can provide superior performance by bypassing the need for shared storage. The VSAN implementation provides the fault tolerance and high availability necessary for enterprise environments that has historically has been the limitation of DAS.

Read caching and write buffering using the Micron P420m PCIe SSD sufficiently masks the latency limitations of HDDs, allowing VMs to run at high performances. Since VMs frequently move between servers for load balancing, there is no guarantee that SSDs that are local to the VM will have cached data. The Mellanox interconnect provides low latencies whenever accessing data between servers is necessary.

To evaluate the Micron and Mellanox hardware supporting VSAN, VMware's Horizon View virtual desktop application was implemented. Using the Login VSI workload simulator, 356 desktops were hosted across three servers. This number is comparable to what a SAN- or NAS-connected shared storage implementation can support, but at a fraction of the cost.

About Login VSI

Login Virtual Session Indexer (Login VSI) is a software tool that simulates realistic user workloads for Horizon View and other major desktop implementations. It is an industry standard for measuring the VDI performance that a software and hardware implementation can support.

Mellanox TECHNOLOGIES

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

15-4175SB
Rev1.0