Performance of VSA in VMware vSphere® 5

Performance Study

TECHNICAL WHITE PAPER

vmware®

Table of Contents

Introduction

vSphere 5 combined with vSphere Storage Appliance (VSA) allows for the creation of shared storage from local storage. This enables the use of vSphere features like vMotion, distributed resource scheduling (DRS), and high availability (HA) without a SAN.

VSA uses either two or three vSphere 5 hosts and the local storage that resides on them to create a VSA storage cluster. This cluster provides shared storage in the form of NFS shares that can be used to host virtual machines (VMs). Additionally, VSA provides redundancy with hardware RAID and data replication across the cluster so that even if one of the server nodes becomes unavailable or a hard disk fails, the storage will continue to be available and useable via the remaining hosts. This results in a highly available, shared storage array created out of local storage with no requirement for additional dedicated storage hardware.

The performance of VSA is directly related to the hardware configuration of the systems used for its cluster. Big differences in capacity, performance, and price exist depending on the exact configuration used. The data replication performed by VSA across its cluster nodes provides high availability, but also has an effect on performance. Testing was done with a mixed workload to examine application performance and infrastructure operations. A set of tests with an I/O generation tool were also run to examine what happens across the hosts in a VSA cluster and to illustrate how to monitor and manage performance.

Executive Summary

VSA provides the basic capabilities of shared storage to environments where it was not possible before. This enables advanced features of a virtual infrastructure for environments that are as simple as just two hosts and an Ethernet switch. VSA is able to run a mix of application workloads while also supporting a range of dynamic virtual infrastructure operations. As with any storage system, there are tradeoffs of capacity, price, and performance that can be made to achieve an optimal solution based on requirements.

Test Environment

A VSA cluster can be built on two or three vSphere 5 hosts. The local storage of the hosts are configured as hardware RAID 1/0 LUNs and used to create shared storage that is presented as an NFS share on each host. The NFS shares reside on each vSphere 5 host and can be used to host VMs with vSphere 5 hosts using NFS to access VMs that are stored on the NFS datastores. VSA installation and management was designed to be very simple and easy to use. The VSA installer does a thorough check of requirements and will prompt for anything that is either missing or incorrect to be fixed before proceeding. Once installation is complete, there is nothing further that needs to be configured or adjusted. All of the local storage on the hosts used in the VSA cluster is taken and used for the VSA cluster. The data replication between nodes will be enabled and working at all times after setup and cannot be disabled. In the event of a node failure, the NFS share that was hosted on the failed node will be automatically brought up on the host with the replica copy of the data in the VSA cluster.

The server resources used by the VSA cluster are small in terms of CPU and memory, leaving the majority of the server available for running other VMs. Each VSA appliance VM uses one vCPU with a 2GHz reservation and 1GB of RAM.

A test environment for VSA was configured with three vSphere 5 hosts. Each host was a two-socket Intel Xeon x5680 3.33GHz-based server with 96GB of RAM and eight 300GB 10,000 RPM SAS disks attached to an internal RAID controller with 512MB of cache and set up in a RAID 1/0 LUN (the RAID level VSA supports).

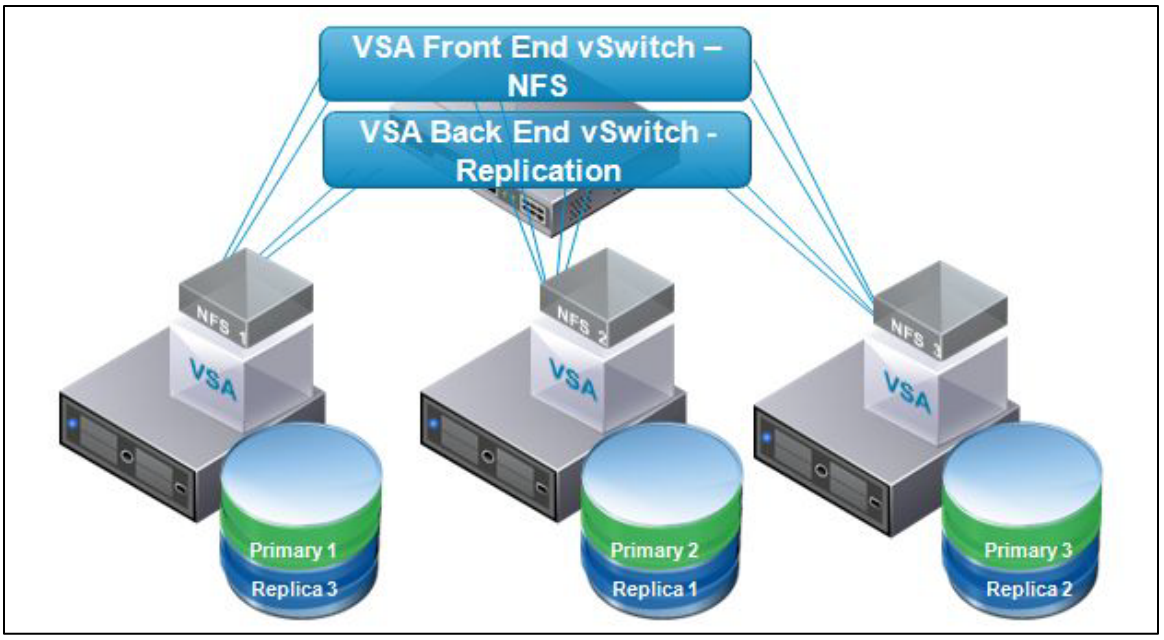

Figure 1. VSA Test Configuration with Three Hosts

The VSA cluster is layered on top of this simple hardware. A 1-vCPU VSA appliance VM on each host used four 1Gb Ethernet network connections and the local storage to create an NFS share on each host. Each NFS share was backed by a local copy and a replicated copy on another host in the cluster. Two of the network connections per host were dedicated for a back-end cluster communication and replication network where the VSA VMs replicated the data between the hosts. The other two ports were used to provide access to the NFS shares. Figure 1 provides a diagram showing the test configuration and key aspects of the VSA cluster.

Key Factors of VSA Performance

In most aspects, the performance of VSA is determined by the same things that determine the performance of any storage array or LUN. There are, however, some aspects that are different or have a bigger impact for VSA than in other environments. Understanding these aspects of VSA performance will aid in planning a deployment, and monitoring or managing the performance of a VSA cluster.

Key factors in the performance of VSA:

- Common Storage Performance Factors
  - Number of disks
  - Rotational speeds of disks
  - Size of disks
  - RAID type
  - I/O request size
  - Randomness of I/O workloads
- Local RAID Controller
- Replication of Data Between VSA hosts
- Mix of Reads and Writes

Common Storage Performance Considerations

Each disk is capable of a certain performance given a certain type of workload. Depending on the type of disk, this performance varies. For many I/O-intensive enterprise applications, the key factor is how many read and write operations the disk can complete quickly. This is usually called Input / Output Operations Per Second or IOPS. The speed of the disk has a direct effect on the number of IOPS. Estimates vary quite a bit on the IOPS capabilities of different speeds of disks. A 10,000 RPM disk is often estimated to get from 120 to 150 IOPS, and 15,000 RPM disks from 170 to 200 IOPS.

In a storage array, these disks are combined together into a SCSI logical unit (sometimes referred to as a LUN or virtual disk) where the data is spread across all of the disks. This means that the number of IOPS that the SCSI logical unit is capable of is now the sum of all the individual disks, with consideration for the RAID type used to create the logical unit. With VSA clusters, RAID 1/0 is the only supported RAID type.

Local RAID Adapter

The performance of the local SCSI RAID adapter is a big factor for VSA. It is important to get a good RAID adapter with ample onboard non-volatile write cache and ensure that the cache settings are enabled. If non-volatile write cache is disabled on the RAID adapter, it will lead to lower-than-expected performance.

VSA Data Replication

The replication of data in a VSA cluster happens automatically and there is nothing that needs to be done to manage it. When planning or examining the performance of a VSA cluster, it is important to understand the impact that the replication has on available storage and the number of disk write operations.

VSA presents an NFS datastore to each host in its cluster, which is how ESXi servers connect to and use the shared storage that VSA creates. Additionally, all data is synchronously replicated from its primary node to a RAID 1 mirror, located on a second host in the VSA cluster. This allows for the data in a VSA cluster to remain available in the event that one node goes offline or is lost for any reason. In order to do this replication, VSA divides the available local storage on each host and uses half for a primary and half for a RAID 1 replica.

Additionally, in order to keep the replica in sync, all writes that occur on the exported primary datastore must also occur on the replica. Each write operation will result in a write to the primary and a write to the replica. Workloads using exported datastore will use resources on two hosts—the host for the primary and the host for the replica. Replication is synchronous in the VSA cluster, meaning that a write won't be acknowledged and completed until it is committed to both the primary datastore and its replica.

Mix of Reads and Writes

The mix of reads and writes is always an important aspect of storage performance. Depending on the RAID type used, the number of physical reads or writes that occur for each logical read or write is different. RAID 1/0 is used for the physical storage configuration of the hosts participating in a VSA cluster. In a RAID 1/0, each logical read results in one physical read, but each logical write results in two physical writes so that the data can be kept in sync on both sides of the RAID 1/0 mirror. In addition to this, VSA replicates all data to a second host via a software-based network RAID 1 mirror. This means that each logical write from a VM will result in four physical writes. Two for the replication and then two at the SCSI logical unit (LUN) level due to RAID 1/0. When planning a VSA deployment, expect each write by a VM to result in four writes to disk. Two physical writes will occur on the primary host and two will occur on the host with the replica copy.

Performance Metrics

Storage performance for VMs using a VSA cluster can be measured with a variety of metrics. To obtain a complete analysis of VSA storage performance, evaluate three key aspects: application / VM, NFS datastore, and physical SCSI adapter.

Application / VM

The performance of the actual application running inside the VM is measured in a variety of ways. An application that reads and writes its data to VSA storage will report its performance in some type of response time or throughput metric. Inside the VM there are also OS-level tools such as perfmon (Windows) or iostat (Linux). These OS-level performance monitoring tools can report storage-specific performance in terms of IOPS and latency. At the VM level, there are also counters in vCenter and esxtop that can provide the same type of information as perfmon and iostat provide for storage. In vCenter, information from these counters is found by going to the performance tab of a VM and looking at Storage. In esxtop interactive mode, press "v" to get to the Virtual Disks output screen.

NFS Datastore

VSA provides an NFS datastore on each host in its cluster which can be used to host many VMs. These NFS datastores are visible to the vSphere hosts and are used to host the VMs virtual disk files. Performance metrics at this level report a summary for all the VMs on the datastore. In vCenter, look at the datastores labeled VSADs to access the performance information for the VSA NFS datastores. In esxtop interactive mode, press "u" to get to the disk device screen.

Physical SCSI Adapter

Viewing performance for the physical SCSI adapter for the local storage used by VSA will show all of the I/O data for VSA on that host. This includes both the primary and replica. The physical SCSI adapter is below the VSA appliance VMs in the stack, which means that all I/O will be measured here regardless of whether it is for the VSA primary or replica. This is different from the NFS datastore view, where only the primary operations are measured. This additional data that includes the replica operations is needed to create the full picture of VSA performance. In vCenter, this data is accessed under the Storage Adapters and in esxtop it is reached by pressing "d" to look at the disk screen. It will most likely be the adapter labeled vmhba1, but it could be different if there are multiple local SCSI adapters present.

Mixed Workload Test

In order to test the performance of VSA in such a way as to simulate a real-world mix of applications, VMware VMmark® 2.0 was selected as a test workload. VMmark 2.0 includes a Microsoft Exchange mailserver, DVD Store database server, three DVD Store application servers, an OLIO database server, an OLIO application server, and a standby VM. Additionally, it includes infrastructure operations vMotion, Storage vMotion, and VM deployment.

The VMs were spread across the three hosts in the VSA cluster with the mail server and standby on host 1, the DVD Store VMs on host 2, and the Olio VMs on host 3. vMotion and VM deploy infrastructure operations were run in addition to the standard VMmark2 application workloads. Storage vMotion was disabled for these tests because there was not another VSA cluster available to be used as a target.

VSA was able to handle this mixed workload simulating thousands of users with the 10,000 RPM SAS disk configuration. The number of sustained IOPS for the test was approximately 2,500 across all three hosts. This IOPS number was measured by adding the IOPS of the physical adapters for all three hosts. It includes the primary, replica, and infrastructure operations.

A key measure of performance is how the application performs and a key factor in VSA performance is the local RAID controller. In order to show the effect of the local RAID controller on application performance, VMmark 2.0 tests were run with the cache enabled and disabled. In order for the test to run successfully with the cache disabled, it was necessary to disable the infrastructure operations. This meant that it was necessary to also disable them for the cache-enabled test to get comparable test results. Figure 2 shows the results from these tests.

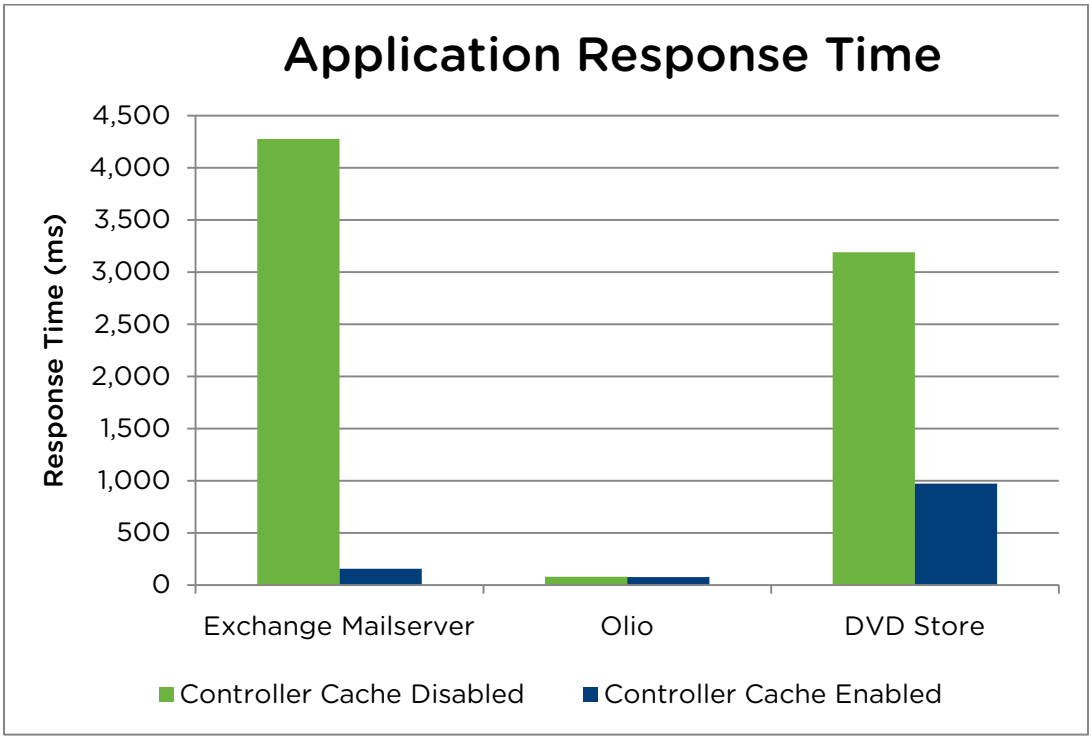

Figure 2. VMmark 2.0 Application Response Time with RAID Controller Cache Disabled and Enabled

The results show that the cache of the RAID controller has a huge effect on performance. The two I/O-intensive workloads showed dramatic changes in response time. By default, the write cache is usually enabled on RAID controllers and this test was done only to show its importance. The most likely scenario for the write controller

cache becoming disabled is when its onboard battery loses its charge and the controller disables the cache to protect data integrity. It is also interesting that OLIO, which is not very I/O-intensive, had almost no change in response, showing the storage performance does not affect all workloads.

IOBlazer Test

IOBlazer was used to illustrate how workload affects the VSA cluster. IOBlazer produces as many disk I/Os as it can, based on the outstanding I/Os parameter, and reports the number of IOPS and average response time or latency for those operations. Because each VSA node is a host for a primary datastore and a replica datastore, a workload on one node will cause IOPS to occur on two nodes.

A simple test scenario with the VSA-based cluster was run with three phases to show what happens. Three Windows Server 2008 64-bit VMs were set up, one on each of the three VSA NFS datastores, with IOBlazer installed to generate an I/O workload. The I/O profile used was 8k block size, random, 50% reads, 50% writes, and 32 outstanding I/Os.

Five IOBlazer tests were run in succession with one immediately following the next, in three phases. In the first phase, each of the VMs ran the IOBlazer workload one at a time so that only one VM was active at a time. In the second phase, two VMs ran IOBlazer at the same time. In the third phase, all three VMs ran IOBlazer at the same time. Performance was recorded using esxtop on all three hosts. The key storage performance metrics from these tests are shown in the following graphs.

The placement of the three test VMs is important to understand the results of the tests. Each host in the VSA cluster hosted one of the NFS datastores. One of the VMs was placed on each datastore, . For the purposes of measuring storage performance, this designeffectively simulates the placement of one VM on each of the hosts in the VSA cluster..

From IOBlazer's perspective, the IOPS and latency numbers for the tests showed approximately a 2x increase in IOPS from one VM to all three VMs, while latency increased by approximately 3ms. Figure 3 shows the IOBlazer reported results across all three test phases.

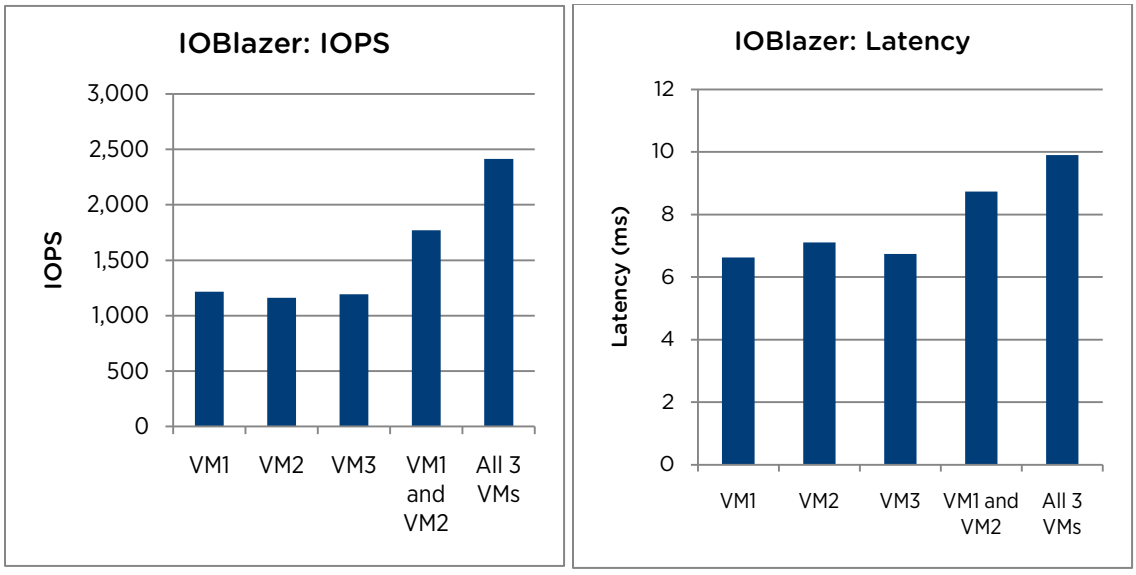

Figure 3. IOBlazer IOPS and Latency Results as Reported by IOBlazer During Testing with one VM per VSA Datastore

Figure 4 shows the IOPS from the same IOBlazer test as measured at the NFS datastore. The esxtop IOPS data for the NFS volumes shows the same thing as the IOBlazer data that was measured from within the VM. This is because only one VM was running on each datastore.

The graph also shows the total IOPS for the VSA cluster across all three NFS datastores. The total line is the same as the individual host lines in phase one of the test when only one VM is under load. In phases two and three, as multiple datastores become busy, the total for the cluster rises, reflecting the increased load. The amount of IOPS across the cluster does not triple, but doubles in the final phase when all three datastores are active.

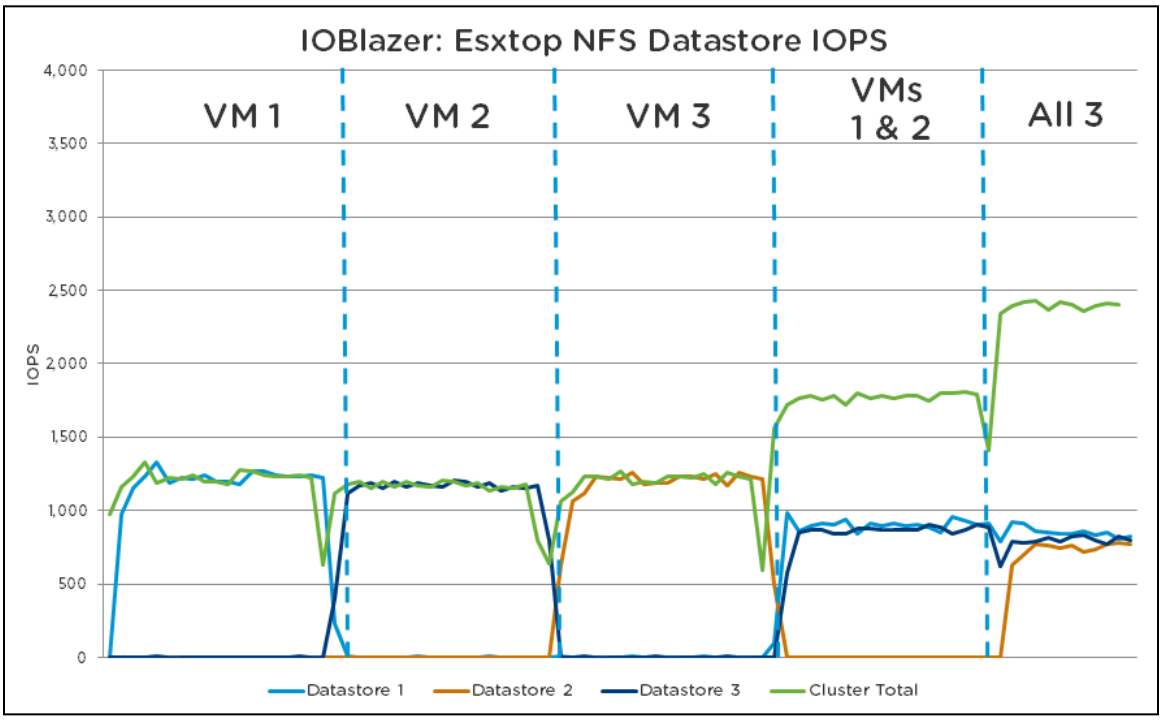

Figure 4. Performance of VSA NFS Datastore Level During IOBlazer Testing

The esxtop data for the storage adapter allows us to see what the IOPS were across each host, including the replication activity as shown in Figure 5. This explains why effective IOPS at the VM and datastore levels do not triple as the test progresses from one to three hosts. Even though only one VM is active on one NFS datastore, the I/Os are occurring on two hosts because data is being written to both the primary and replica. Additionally, reads can be done on either the primary or replica depending on load across the VSA nodes.

Because one active workload actually affects two hosts, once two hosts are active, all of the hosts are now active. In Figure 5, the total number of IOPS does not increase linearly between phases two and three of the test because all of the disks were already busy once two hosts were active. Using the storage adapter performance metrics shows the impact of the replicas and how many IOPS are actually occurring at the physical adapter level.

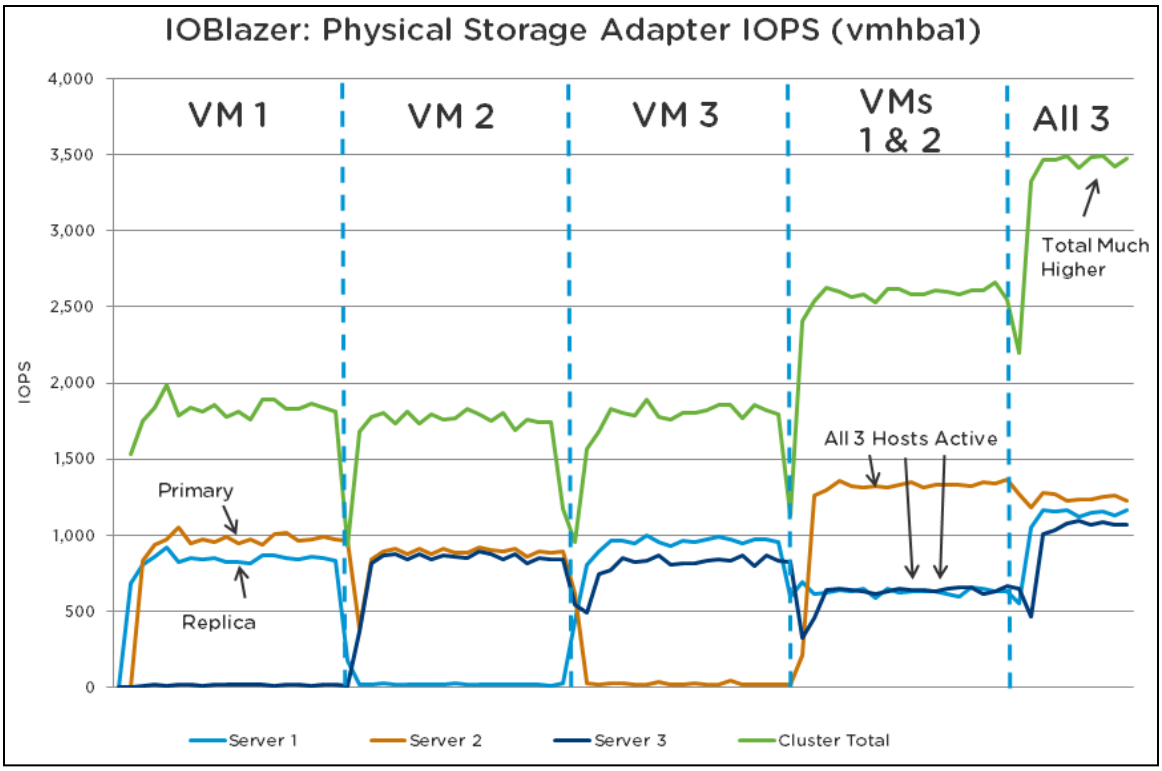

Figure 5. Performance Measured for the Physical Adapter During IOBlazer Tests

In Figure 3 and Figure 4, the number of IOPS is approximately 1,200 when a single VM is running the IOBlazer workload. In Figure 3, this is reported from within the VM from the test application's perspective. In Figure 4, this is reported from the NFS datastore level perspective. Because there is only one VM per datastore, the number is the same. The mix of read and writes used in this test was 50/50, meaning that there are 600 reads and 600 writes occurring.

In Figure 5, the number of IOPS across the cluster is approximately 1800 in each of the three cases where only one VM is actively running IOBlazer. This is as expected because the 600 reads are only occurring once, while the 600 writes are occurring twice. Because the reads can occur on either the primary or replica, the number of IOPS on either VSA node is similar.

Best Practices

There is a tradeoff between capacity, performance, and price when deciding on the disk configuration for the VSA hosts. SATA disks provide higher capacities at lower prices and lower reliability levels, but do not perform as well as SAS disks. SAS disks are more reliable and have higher performance, but cost more than SATA disks. Increasing the number of disks will increase performance and so will increasing the speed of disks, but doing either of these will result in higher costs.

The hardware used for VSA is a big factor in performance. Ensure that the RAID adapter in the VSA hosts has sizeable cache and the cache is enabled.

VSA requires a RAID 1/0 configuration for the local storage on the hosts and also replicates data across the cluster to provide highly available shared storage. When planning a deployment of VSA, the combination of RAID 1/0 and data replication mean that useable capacity will be ¼ of raw capacity and each write by a VM will result in two physical writes on the primary host and two physical writes on the replica host.

VSA performance can be monitored at three levels to get a complete picture of the environment. The application- or VM-level performance provides the view of the "end user." The NFS datastore performance shows how much each of the VSA datastores is being loaded by all of the VMs that they serve. The physical SCSI adapter view shows the total impact of both VSA primary and replica data copies on the host.

Conclusion

VSA allows features like vMotion, DRS, Storage vMotion, and HA to be possible using only local storage. This enables advanced vSphere capabilities now to be possible for environments as small as just two servers. vSphere provides the VSA environment with tools to monitor and manage performance and understanding the key factors of VSA performance helps drive a successful deployment..

About the Author

Todd Muirhead is a performance engineer at VMware focusing on database, mail server, and storage performance with vSphere.

Acknowledgements

Thanks to the VSA, VMmark, and Performance Engineering teams for their assistance and feedback throughout.