Performance Implications of Storage I/O Control–Enabled NFS Datastores in VMware vSphere® 5.0

Performance Study

TECHNICAL WHITE PAPER

Table of Contents

Introduction

Datacenters based on VMware's virtualization products often employ a shared storage infrastructure to service clusters of VMware vSphere hosts. External storage infrastructure such as storage area networks (SANs) and network attached storage (NAS) expose logical storage devices (LUNs) or volumes that can be shared across a cluster of vSphere hosts. VMFS datastores are created on the logical devices whereas volumes using the NFS access protocol can be directly exposed as NFS datastores to the vSphere hosts. Virtual disks are created in these datastores and are assigned to virtual machines (VMs). Consolidating VMs' virtual disks onto a single datastore backed by a higher number of disks has several advantages—ease of management, better resource utilization, and higher performance (when storage is not bottlenecked). Active VMs can take advantage of periods of inactivity in the shared environment to obtain a higher I/O performance compared to what they would have obtained had they used dedicated datastores backed by fewer disks.

However, there are instances when a higher than expected number of I/O intensive VMs sharing the same storage device become active at the same time. During this period of peak load, VMs contend with each other for storage resources. In such situations, the lack of a global control mechanism can lead to degraded performance of the VMs running high priority workloads as they compete for storage resources which have VMs running low priority workloads.

Storage I/O Control (SIOC), a feature that was introduced in vSphere 4.1, provides a fine-grained storage control mechanism which dynamically allocates portions of hosts' I/O queues to VMs whose data is located on the same datastore. This dynamic allocation is based on shares that an administrator assigns to the VMs. Using SIOC, vSphere administrators can allocate more shares to the VM that is running a high priority application, which ensures that the application can maintain a higher level of performance during peak load periods. SIOC is enhanced in vSphere 5.0 to include support for NFS datastores.

This paper shows the advantages of using SIOC when consolidating applications on a vSphere-based virtual infrastructure. For more information about using SIOC, see "Managing Storage I/O Resources" in the *vSphere 5.0 Resource Management Guide*[1].

Executive Summary

Application performance can be impacted when servers contend for I/O resources in a shared storage environment. There is a need for isolating the performance of high priority applications from other low priority applications by appropriately prioritizing the access to shared I/O resources. Storage I/O Control (SIOC) provides a dynamic control mechanism for managing I/O resources across virtual machines in a cluster. This feature was introduced in vSphere 4.1 with support for VMs that share a storage area network (SAN). In VMware vSphere 5.0, this feature has been extended to support network attached storage (NAS) datastores using the NFS application protocol (also known as NFS datastores).

Experiments conducted in the VMware performance labs show that:

- SIOC regulates VMs' access to shared I/O resources based on disk shares assigned to them. During the periods of I/O congestion, VMs are allowed to use only a fraction of the shared I/O resources in proportion to their relative priority as determined by the disk shares.
- SIOC helps in isolating the performance of latency-sensitive applications that issue small sized (≤8KB) I/O requests from the increase in I/O latency due to a larger sized (≥32KB) request issued to the same storage shared by other applications.
- If the VMs do not fully utilize their portion of the allocated I/O resources on a shared storage device, SIOC redistributes the unutilized resources to those VMs that need them in proportion to the VMs' disk shares. This results in a fair allocation of storage resources without any loss in their utilization.

- SIOC minimizes the fluctuations in the performance of a high priority workload during periods of I/O congestion. For the test cases executed at VMware labs, limiting the fluctuations to a small range resulted in as much as 11% performance benefit compared to that in an unmanaged scenario.

Terminology

The following terms are used in this paper:

- **Congestion threshold –** defines a threshold for I/O latency that signals when SIOC should begin to manage each VM's access to the I/O resources on a shared NFS datastore. Access is granted according to the disk shares that are assigned to each VM. During the congested period, SIOC strives to maintain the overall I/O latency on the datastore around the congestion threshold value.
- **Disk shares** – represent the relative importance of a virtual machine with regard to the distribution of storage I/O resources. Under resource contention, virtual machines with higher share values receive higher priority to access the shared storage device, which typically results in higher throughput and lower latency.
- **Host aggregate shares** – sum of disk shares of all VMs on a host sharing a datastore.
- **Datastore-wide normalized I/O latency** – weighted average of normalized I/O latency as seen by hosts sharing a datastore. The number of I/O requests completed per second by each host is used as the weight for averaging. Before computing the average, the device latencies are normalized based on the I/O request size. This is done because device latency depends on I/O size (for example, 256KB versus 8KB). This way, high device latency due to larger I/O size is distinguished from high device latency due to actual I/O congestion at the device.

Experimental Environment

The experiments were based on DVD Store workloads running in four VMs (two hosts, two VMs each). In all tests, the DVD Store workload in VM 4 was picked as the high priority workload. Note that VM 4 shared its host with VM 3.

Table 1 describes the vSphere host and system configuration for the experiments.

COMPONENT	DETAILS
Hypervisor	VMware vSphere 5.0
Processors	Four 2.27GHz Eight Core Intel X7560 Xeon (host 1) Two 2.93GHz Quad Core Intel Xeon X5570 (host 2)
Memory	256GB (host 1) 96GB (host 2)
Guest Operating System	Windows Server 2008 R2 Enterprise x64 edition
Database	SQL Server 2008 R2 Enterprise x64 edition
Virtual CPUs/Memory	VMs 1-3: 2 vCPUs, 4GB memory VM 4: 4 vCPUs, 8GB memory
Virtual Disk Size (OS)	50GB
Fibre Channel HBA	Dual port QLogic QLE2562

File System	NFS for data and index files of DVD Store; VMFS for operating system, SQL Server binaries, and log files of DVD Store
Storage Server/Array	NetApp FAS 6030
Test Application	DVD Store Version 2.1

Table 1. vSphere Host and Storage System Configuration Information

Test and Measurement Tools

DVD Store version 2.1 is an open source, online e-commerce test application, with a back-end database component, a Web application layer, and driver programs. For more information, see http://www.delltechcenter.com/page/dvd+store.

For these tests, a 50GB database was created.

Setup for the Tests

Four identical VMs running Windows Server 2008 R2 Enterprise x64 edition were used for the experiments. The VMs ran on two vSphere hosts (two VMs per host). A 50GB virtual disk created in a VMFS datastore was used to install the guest operating system and application binaries in each VM. All the VMs shared a separate NFS datastore created on a RAID 5 LUN (eight disks). Four identical thick virtual disks were created on the shared datastore and one was assigned to each VM. The virtual disks were used to store the data and index files of the DVD Store workload running in each VM. Four additional, identical virtual disks were created on a separate VMFS datastore and one virtual disk was assigned to each VM. These virtual disks were used to store the log files of the DVD Store workload running in each VM.

DVD Store in one of the VMs (VM 4) was identified as a high priority workload. Figure 1 shows the test bed layout.

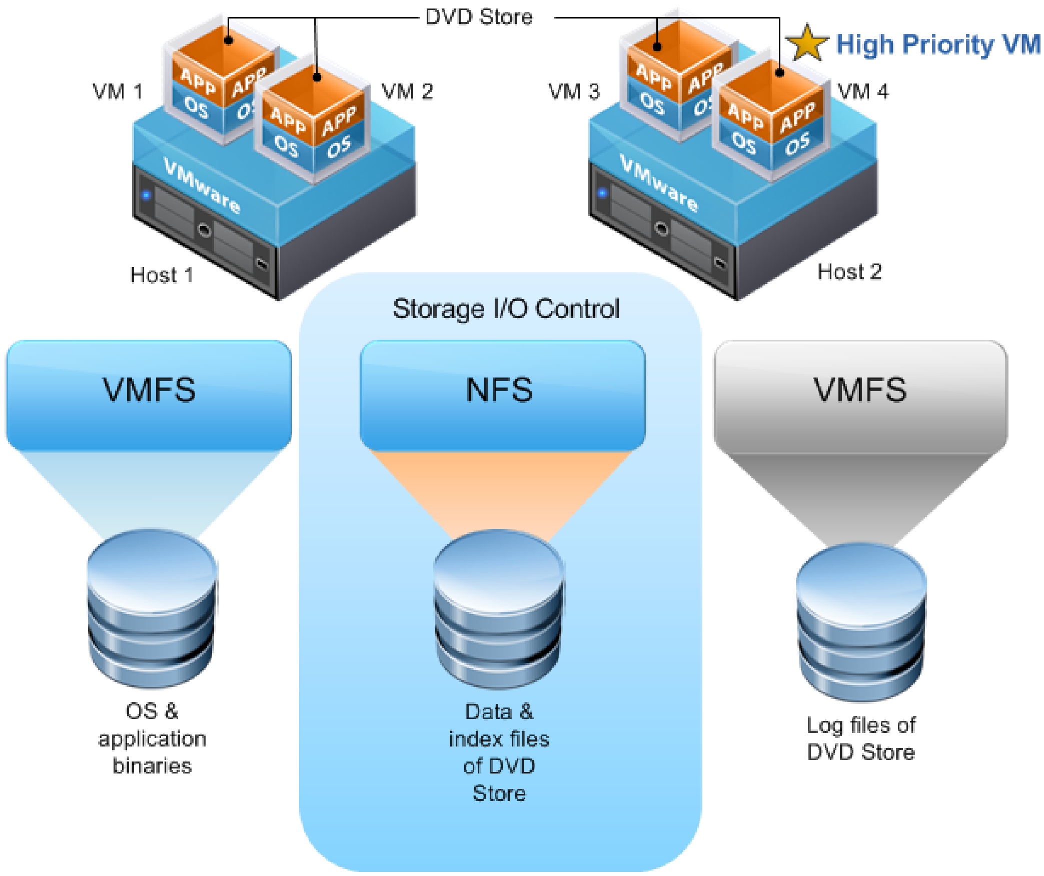


Figure 1. Test Bed Layout with 4-VM Configuration

Test 1: Performance Impact of Storage I/O Control

This test case shows the impact SIOC can have on the performance of an application. In this test, the DVD Store workloads, one in each VM, were run simultaneously for 10 minutes. Running DVD Store workloads in all VMs generated a high I/O load on the NFS datastore causing it to be nearly saturated. The performance metrics—total number of orders completed and average time to complete a single order—were recorded for the high priority DVD Store workload in each run. The DVD Store workload in VM 4 was configured to support twice the number of DVD Store users than in the other VMs.

Initially, the disk share value of each VM was left as "normal" (default value of 1000), which resulted in an equal I/O priority assignment to each VM. Next, the disk share value of VM 4 was increased to 2000. Then, SIOC was enabled on the shared NFS datastore and the congestion threshold value was set to 12 milliseconds. Table 2 shows the placement of VMs, their disk shares, and the DVD Store configuration in each VM.

VM ID	HOST ID	DISK SHARES		NUMBER OF DS2 USERS	THINK TIME BETWEEN TRANSACTIONS (MILLISECONDS)
		Default	**SIOC Enabled**		
1	1	1000	1000	36	500
2	1	1000	1000	36	500
3	2	1000	1000	36	500
4	**2**	**1000**	**2000**	**72**	**500**

Table 2. VM Configuration and Placement for Test 1

Figure 2 compares the performance of the high priority workload with and without SIOC enabled on the NFS datastore.

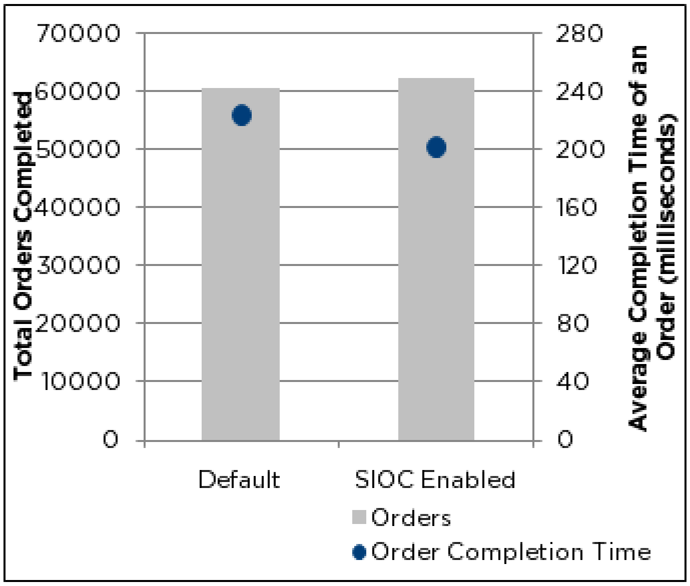


Figure 2. Impact of Storage I/O Control on Performance of DVD Store Workload in VM 4

In the default configuration, each VM was allowed to use the shared storage resources freely without any access control. The unmanaged access caused the I/O load on the datastore to increase. This resulted in the total completed orders from a 10-minute run of the high priority DVD Store workload (in VM 4) to be relatively lower. When disk shares were assigned to the VMs and SIOC was enabled on the shared NFS datastore, the total orders completed by the high priority DVD Store workload improved by approximately 3%. But a bigger improvement was observed in the completion time of an order. When SIOC was enabled, the order completion time of the high priority DVD Store workload dropped by 10%.

A closer look at the I/O performance of the VMs is useful in understanding the reason behind the improvement. Figures 3 and 4 show the I/O performance of the VMs in the default configuration without and with SIOC enabled on the shared datastore.

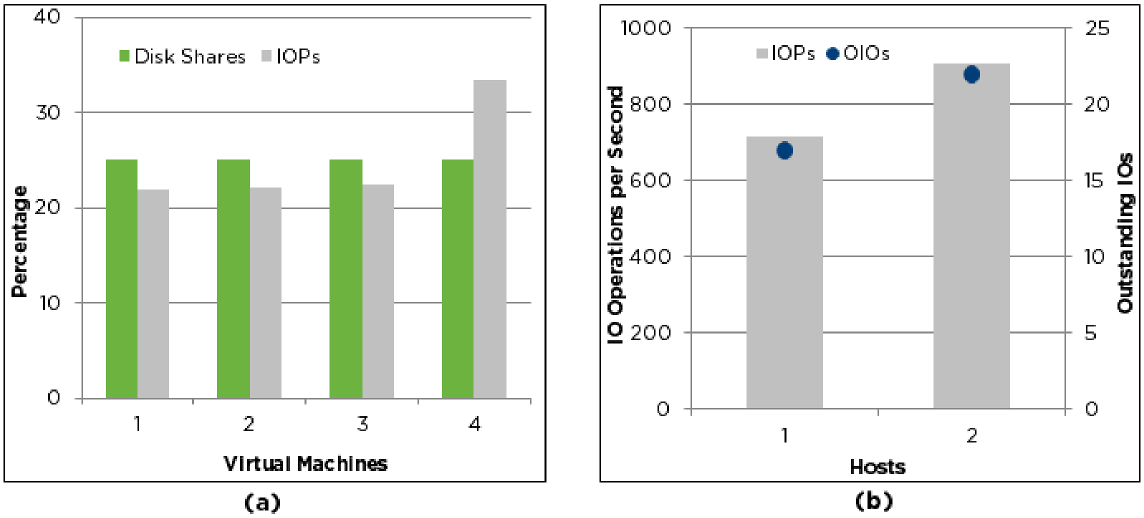


Figure 3. I/O Performance of the VMs in the Default Configuration (without SIOC)

By default, each VM sharing a datastore is allocated an equal share of access to the I/O resources of the datastore. In the default configuration, all the VMs were allowed equal access to the I/O queue of the shared NFS datastore due to their equal disk shares. Hence the VMs were allowed to issue as many concurrent I/Os to the shared NFS datastore as their DVD Store workload demanded without any restrictions. The I/O load from each of the low priority VMs constituted about 22% and the I/O load from VM 4 was about 33% of the total I/O load on the shared datastore (Figure 3). VM 4 issued more I/O requests than other VMs because it had twice the user load of the other VMs. As a result, there were approximately 40 outstanding I/O requests on the NFS datastore and the datastore delivered about 1624 IOPs at any given time throughout the duration of the test.

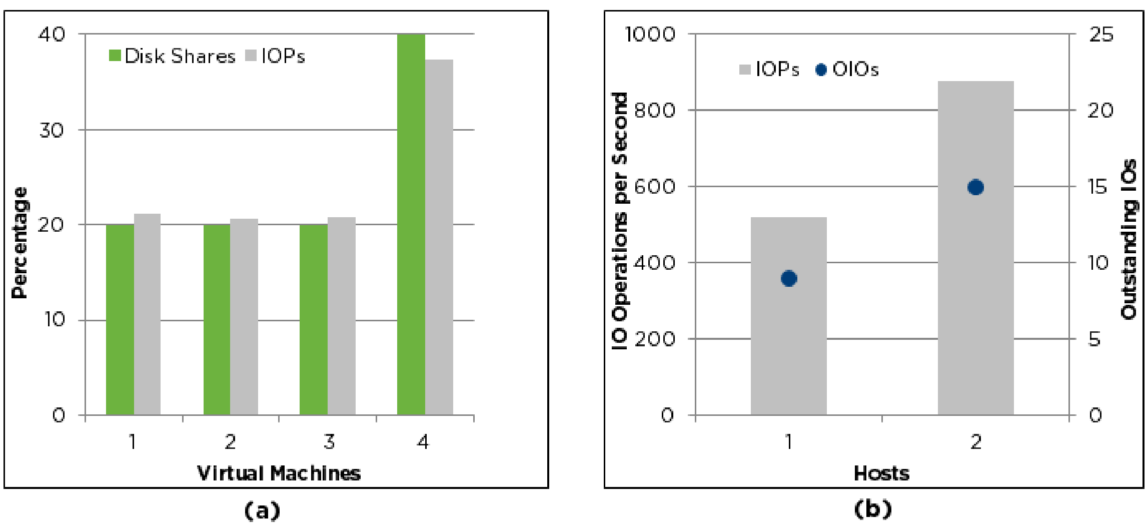


Figure 4. I/O Performance of the VMs when SIOC Was Enabled

When SIOC was enabled and different disk shares were assigned to each VM (as shown in Table 2), each VM's share of access to the I/O queue changed. SIOC detected I/O congestion on the datastore and started managing each VM's access to the I/O queue of the shared NFS datastore in proportion to its disk shares. This resulted in a reduction in the percentage of I/O load from each of the low priority VMs to about 20% and an increase in the percentage of I/O load from VM 4 to about 40% (Figure 4(a))The net result was that the overall I/O load on the datastore decreased (compare outstanding I/Os in Figure 4(b) to that in Figure 3(b) as a result of the managed access to the shared I/O resources, but the I/O load originating from each VM was maintained in proportion to each VM's disk shares. The total IOPs measured at host 1 reduced, but the total IOPs measured at host 2 increased (Figure 4(b)) compared to that in the default configuration in Figure 3(b).

The bulk of this increase in IOPs was seen by VM 4 as explained next.

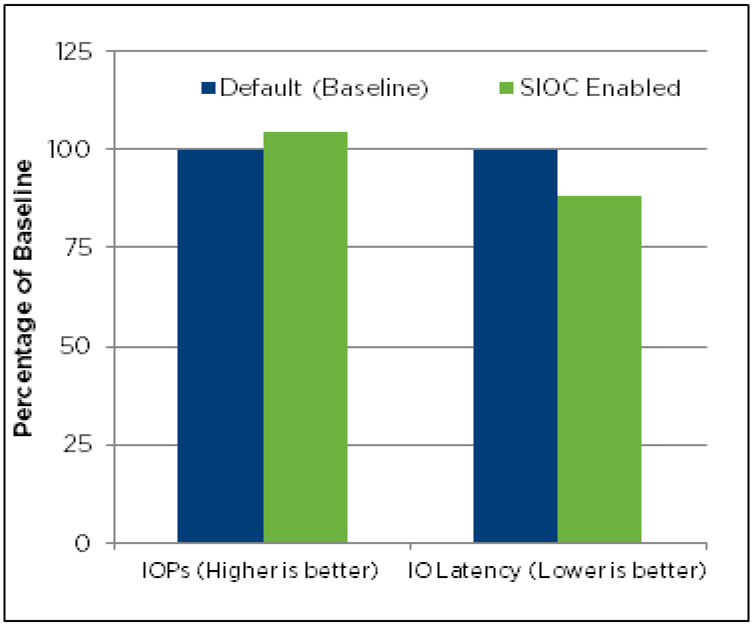


Figure 5. Comparison of I/O Performance of VM Running High Priority DVD Store Workload

Figure 5 compares the I/O performance of VM 4 in the default configuration with that when SIOC was enabled on the shared NFS datastore. Due to SIOC managing each VM's access to the I/O queue of the shared datastore in proportion to the VM's disk shares, VM 4's portion of the access increased. This resulted in a 5% increase in IOPs for VM 4. An even bigger improvement was seen in the latency of an I/O operation from VM 4. As the total load on the datastore became more manageable, the latency of I/O operations on the datastore reduced to 88% of the baseline value. This increase in throughput at a lower latency resulted in a performance improvement for the high priority DVD Store workload running in VM 4.

Test 2: Isolating Performance of Applications with Small Request Size

Due to a variety of applications being virtualized today, it is not uncommon to see storage resources shared across applications whose I/O characteristics vary vastly. Applications such as OLTP databases that employ small (8KB), random sized I/O requests can share storage resources with applications, such as Web and media servers, that issue large (32KB) sized I/O requests[2]. Sometimes the same application can consist of I/O requests that are completely opposite in nature. For example, in database applications, access to data and index files tend to be small random reads while access to log files tends to be large sequential writes.

Large I/O requests typically require a longer time to complete compared to small I/O requests. Hence the latency of an I/O operation on a storage device servicing large I/O requests typically tends to be higher than that on the same storage device with only small sized requests.

This test case shows the ability of SIOC to isolate the performance of applications that use small sized I/O requests in the presence of applications that issue large sized I/O requests. For this test case, two VMs were used to simulate the scenario described in the previous paragraph. Iometer was installed in one of the VMs and DVD Store in the other VM. The VMs shared an NFS datastore. The data and index files of the DVD Store workload were placed on a thick virtual disk which was created on the shared datastore. The Iometer workload issued requests against another thick virtual disk, which was also created on the shared datastore. The virtual disks for each VM were approximately the same size.

There were three test runs in this experiment: a baseline run, a run with the default SIOC settings, and a run with the SIOC settings changed to give the DVD Store workload higher priority.

Initially, only the DVD Store workload was made active for 10 minutes and its performance was recorded. This run served as a baseline with which to compare subsequent test runs.

For the second run, Iometer was configured to issue large I/O requests (12 concurrent random read requests of 64KB). The disk shares of both the VMs were left as "normal" (default value of 1000). Both Iometer and DVD Store workloads were run at the same time for 10 minutes.

For the third run, SIOC was enabled on the shared NFS datastore and the congestion threshold value was set to 10 milliseconds. Each VM was assigned a disk share with the highest share allocated to the DVD Store VM. Again, both of the Iometer and DVD Store workloads were run for 10 minutes as shown in Table 3. The performance metric "total number of orders completed" was recorded for the DVD Store workload in each case.

VM ID	HOST ID	WORKLOAD	NUMBER OF DS2 USERS	DISK SHARES	
				Default	**SIOC Enabled**
1	1	DVD Store	72	1000	High (2000)
4	2	Iometer	--	1000	Low (500)

Table 3. VM Configuration and Placement for Test 2

Figure 6 compares the performance of the DVD Store workload in all three of the situations previously described. The performance of the DVD Store workload running alone serves as the baseline for comparison. Total orders completed in the other cases are normalized to the baseline value.

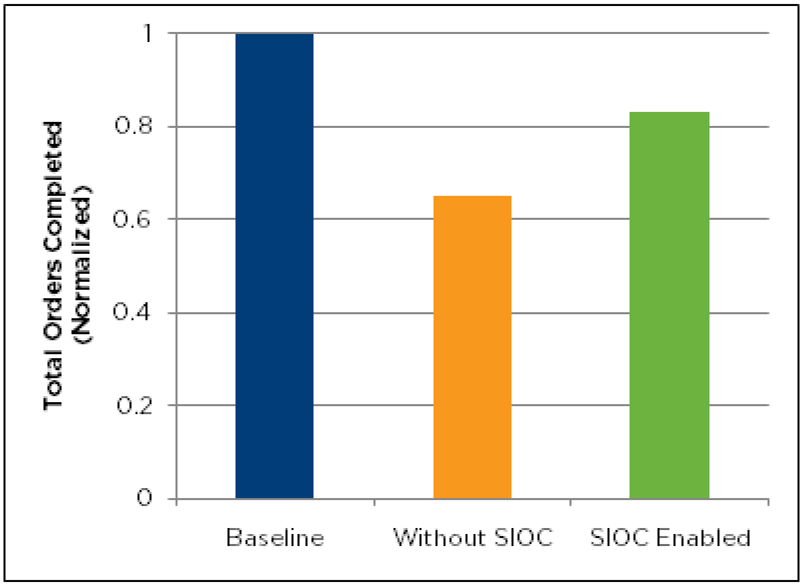


Figure 6. Performance of DVD Store Workload in Presence of an Iometer Workload Issuing 64KB Random Requests

As seen in the graph, in the default case (default disk share assignment without SIOC enabled on the datastore), the throughput of the DVD Store workload decreased to 65% of the baseline value. Because the disk shares assigned to both VMs were the same, each VM could access the I/O queue of the shared datastore equally. The large sized I/O requests from the Iometer workload caused the latency of all I/O requests to the datastore to increase, which resulted in an increase in the order completion time of the DVD Store workload. This led to a decrease in the throughput of the DVD Store workload.

When SIOC was enabled on the shared NFS datastore and the disk shares of the VMs were changed, the total orders completed by the DVD Store workload improved to approximately 83% of the baseline. By changing the disk shares, VM 2’s access to the I/O queue increased by four times. This reduced the number of large sized I/O requests coming from the Iometer VM and this reduction lowered the impact on the overall I/O latency. As a result, the throughput of the DVD Store workload improved.

Note that the throughput of the DVD Store workload could have been improved further by increasing the disk shares assigned to its VM.

Test 3: Intelligent Prioritization of I/O Resources

It is highly unlikely that the I/O load of the applications on a shared a storage device will remain constant. Any access control mechanism should be smart enough to detect the changing conditions quickly and re-provision the I/O resources to other VMs that need them. SIOC achieves this by monitoring the VMs' usage of the I/O queue at the host and dynamically redistributing any unutilized queue slots to those VMs that need them. The proportion of the VMs' disk shares determines the redistribution of unutilized queue slots. This results in a fair allocation of storage resources without any loss in their utilization.

This test case shows the dynamic nature of SIOC. A separate virtual disk created in a shared NFS datastore was assigned to each VM and was used to place the data and index files of the DVD Store workload running in that VM. The DVD Store workload in VM 4 was identified as the high priority workload for this test.

Running DVD Store workloads in all VMs generated a high I/O load on the NFS datastore, causing it to be nearly saturated. SIOC was enabled on the shared NFS datastore and the congestion threshold was set to 12 milliseconds. Table 4 shows the disk shares assigned to each VM. Because the DVD Store workload in VM 4 was identified as high priority, VM 4 was assigned the highest number of disk shares (2000) because it supported twice the users supported by other VMs.

VM ID	HOST ID	DISK SHARES	NUMBER OF DS2 USERS	THINK TIME BETWEEN TRANSACTIONS (MILLISECONDS)
1	1	1000	36	500
2	1	1000	36	500
3	2	1000	36	500
4	2	2000	72	500

Table 4. VM Configuration and Placement for Test 3

This test case had three phases:

- Phase 1: The DVD Store workloads in all four VMs were active for 300 seconds.
- Phase 2: The DVD Store workload in VM 4 went idle for 300 seconds. DVD Store in the remaining three VMs continued to be active.
- Phase 3: The DVD Store workload in VM 4 became active again for the next 300 seconds. DVD Store in the remaining three VMs continued to be active.

The performance metric "orders completed per second" was recorded for all the DVD Store workloads for the entire test duration at 10-second intervals.

Figure 7 shows the performance of the DVD Store workload in each VM during different phases of the test. When all VMs were active, VM 4 was allowed to use the highest number of queue slots from the host I/O queue because it was assigned the highest number of disk shares. As a result, the DVD Store workload in VM 4 delivered the highest throughput (nearly 100 orders per second) of all the VMs. When VM 4 went idle, the throughput of the DVD Store workload in VMs 1 through 3 improved by an equal amount, which indicated that SIOC quickly increased the active VMs' shares of access to the host's I/O queue. This helped to maintain a high utilization of the shared datastore.

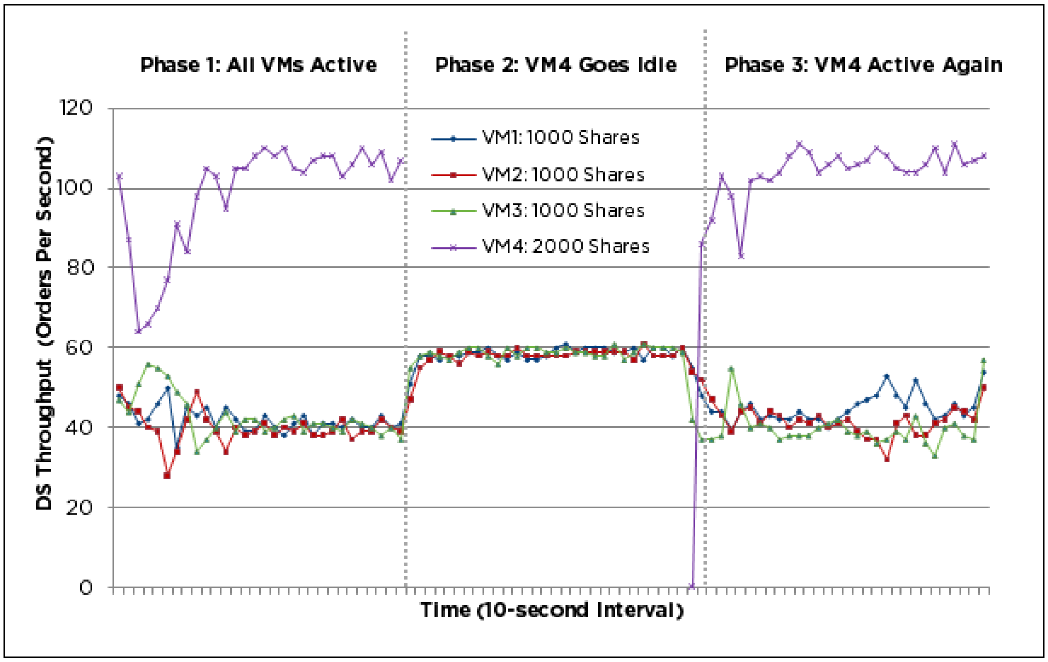
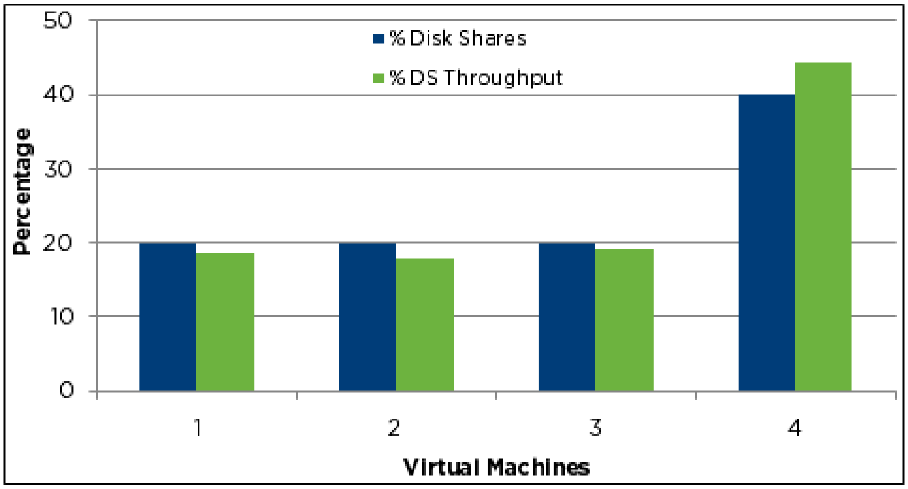


Figure 7. Application Throughput of all VMs in Different Phases

A more detailed look at what happened in each phase follows.

Phase 1: In this phase, the DVD Store workloads in all four VMs were active. This led to contention among all VMs for accessing the shared datastore, which resulted in an increase in the I/O latency. When the datastore-wide I/O latency increased beyond the congestion threshold limit, SIOC detected the I/O congestion. At this stage, SIOC started limiting each VM's access to its host's I/O queue based on the I/O priority of the VM (disk shares). As a result, each VM's I/O requests per second became proportional to its disk shares. This resulted in the DVD Store throughput of each VM being nearly proportional* to its disk shares as shown in Figure 8.



*SIOC's ability to accurately throttle each VM's access to the host device queue also depends on the read/write I/O access profile of the VMs sharing the datastore. This can cause the computation of each VM's allowed access to the shared datastore to be non-trivial. A detailed explanation of how they influence SIOC is beyond the scope of this paper.

Figure 8. Ratios of DVD Store Throughput and VM Disk Shares in Phase 1

Phase 2: At the end of phase 1, the high priority DVD Store workload in VM 4 went idle. However, the DVD Store workloads in the remaining three VMs continued to be active. VMs 1 through 3 still contended for access to the shared datastore. SIOC monitored the I/O contention, but this time detected that there were no active I/O requests from VM 4. Because SIOC is designed to be dynamic, it reconfigured the hosts' queue depth and increased each active VM's access to the queues almost instantaneously based on their disk shares as shown in Figure 9. This resulted in an increase of DVD Store throughput in all active VMs by as much as 45% compared to the VMs' phase 1 throughput.

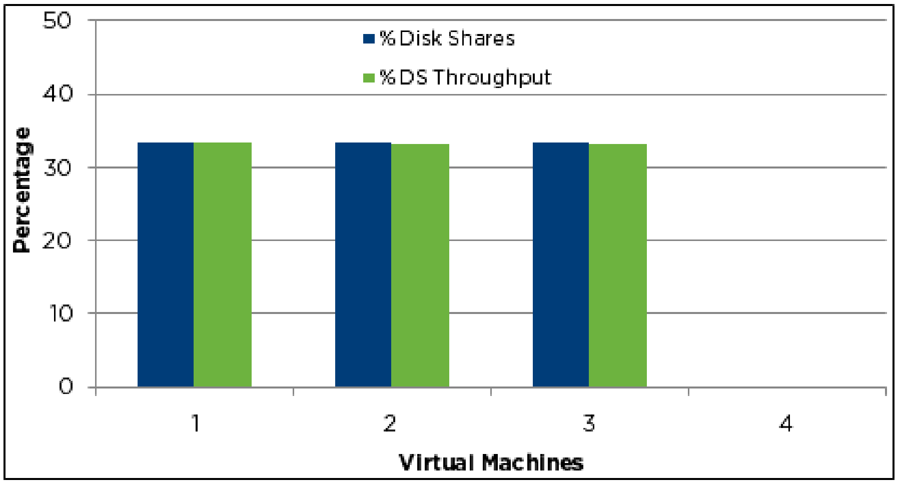


Figure 9. Ratios of DVD Store Throughput and VM Disk Shares in Phase 2

Phase 3: In phase 3, the high priority DVD Store workload in VM 4 became active again and started issuing I/O requests. SIOC detected that VM 4 was active once again and in less than a minute, SIOC modified each VM's access to the shared NFS datastore. Each VM's share of access returned to the phase 1 level. The DVD Store performance in VMs 1 through 3 dropped to their phase 1 values whereas the DVD Store performance of VM 4 increased to its phase 1 value as shown in Figure 10.

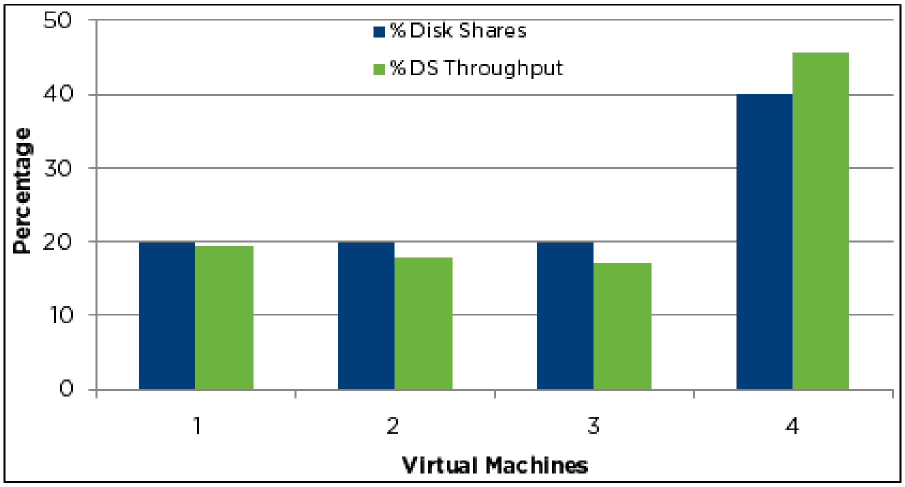


Figure 10. Ratios of DVD Store Throughput and VM Disk Shares in Phase 3

Device Queue Depth at the Host

When an NFS datastore is shared by more than one vSphere host, the array-level I/O queue of the logical device on which the datastore resides is shared among the hosts. SIOC, if enabled, manages each host's access to the array-level I/O queue by throttling the I/O queue of the NFS client in each host to keep the datastore latency close to the congestion threshold. The lower and upper bound on the current I/O queue in each host is as follows:

- Lower limit: 4
- Upper limit: will be set to 256 or the queue depth set by SIOC, whichever is lower

Throttling the host-level NFS I/O queue causes the local disk scheduler in ESX to further manage the I/O queue by allowing the individual VMs running on the host to use a specific number of queue slots. Access to queue slots is based on several factors—the relative priority (disk shares) of individual VMs, the congestion threshold, and the current usage of the slots by the individual VMs. By enabling the proportional allocation of queue slots to each VM, SIOC regulates the VMs' access to their host's device queue.

SIOC monitors the usage of the device queue in each host, aggregate I/O requests per second from each host, and datastore-wide I/O latency every four seconds and throttles the I/O queue in each host, if required. The I/O queue depth of both hosts in all three phases of this test is shown in Figure 11.

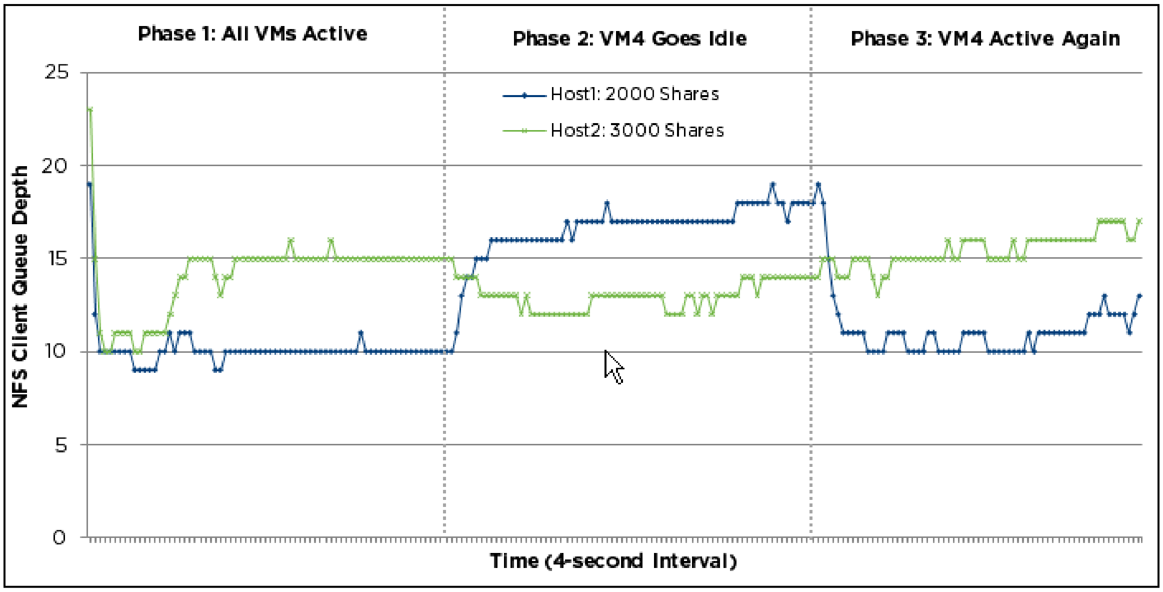


Figure 11. NFS Client Queue Depth in Each Host

As the DVD Store workloads started issuing I/O requests, the I/O latency increased beyond the congestion threshold value. SIOC detected the congestion immediately and started throttling the I/O queue depth of the NFS client in each host. Very soon the I/O queue depths in hosts 1 and 2 settled to 10 and 15 respectively. At this stage, the queue depths in hosts 1 and 2 were in the ratio 2:3, which was in proportion to the aggregate disk shares of VMs running on them (2000:3000).

In the second phase, the DVD Store workload in VM 4 (on host 2) became inactive and stopped issuing I/O requests to the datastore. SIOC detected the change in the I/O profile of VM 4 and immediately increased the NFS client queue depth of host 1 and reduced the queue depth of host 2 as the aggregate disk shares of active VMs on host 1 was now higher than that on host 2.

In the third phase, the DVD Store workload in VM 4 became active, causing SIOC to immediately throttle back the queue depths of host 1 and 2 to their phase 1 values[†].

Test 4: SIOC in a Datacenter

The final test case shows SIOC's effectiveness in preventing performance fluctuations of a high priority workload, caused by contention for storage resources. The setup used for this test case represents a centralized datacenter servicing various offices on different coasts. Imagine VMs 1 through 3 to be database servers supporting order processing, and VM 4 to be a database server supporting product catalogs, with each VM operating in different time zones. In this test case, the order entry servers were active at different time periods based on the time zone in which they operated. The product catalog server serviced regions in all time zones; therefore, it remained active throughout the test duration.

- VM 1 - East coast (order processing server)
- VM 2 - Central (order processing server)
- VM 3 - West coast (order processing server)
- VM 4 - Servicing all the offices (product catalog server)

The order processing servers (VMs 1 through 3) were identified as low priority VMs and were assigned lower priority (500 disk shares). The product catalog server (VM 4) was identified as a high priority VM and assigned the highest priority (4000 disk shares). All VMs ran the DVD Store workload. The VM configuration and placement, and the DVD Store configuration are described in Table 5.

VIRTUAL MACHINES	HOST	VCPU	MEMORY (IN GB)	DISK SHARES	NUMBER OF DVD STORE USERS	THINK TIME BETWEEN TRANSACTIONS (MILLISECONDS)
1	1	2	8	500	36	500
2	1	2	4	500	36	500
3	2	2	4	500	36	500
4	2	4	8	4000	72	300

Table 5. VM Configuration and Placement for Test 4

[†] The minor difference observed in the NFS client queue depths of host 1 and 2 during phase 1 and 3 can be attributed to the changing workload conditions of DVD Store. As time progressed, more data was cached in the SQL Server buffer, which caused the overall I/O latency to drop a bit. This resulted in increased queue depths for both the hosts. A detailed explanation of this behavior is beyond the scope of this paper.

Load Period: The workloads in all VMs were started and stopped according to the schedule shown in Table 6.

TIME IN SECONDS	ACTIVITY
0	VM 4 starts
300	VM 3 starts
600	VM 2 starts
900	VM 1 starts
1200	VM 3 stops
1500	VM 2 stops
1800	VM 1 stops
2100	VM 4 stops

Table 6. Load Period Schedule

Figure 12 shows performance of the DS2 workload in all four VMs with default settings.

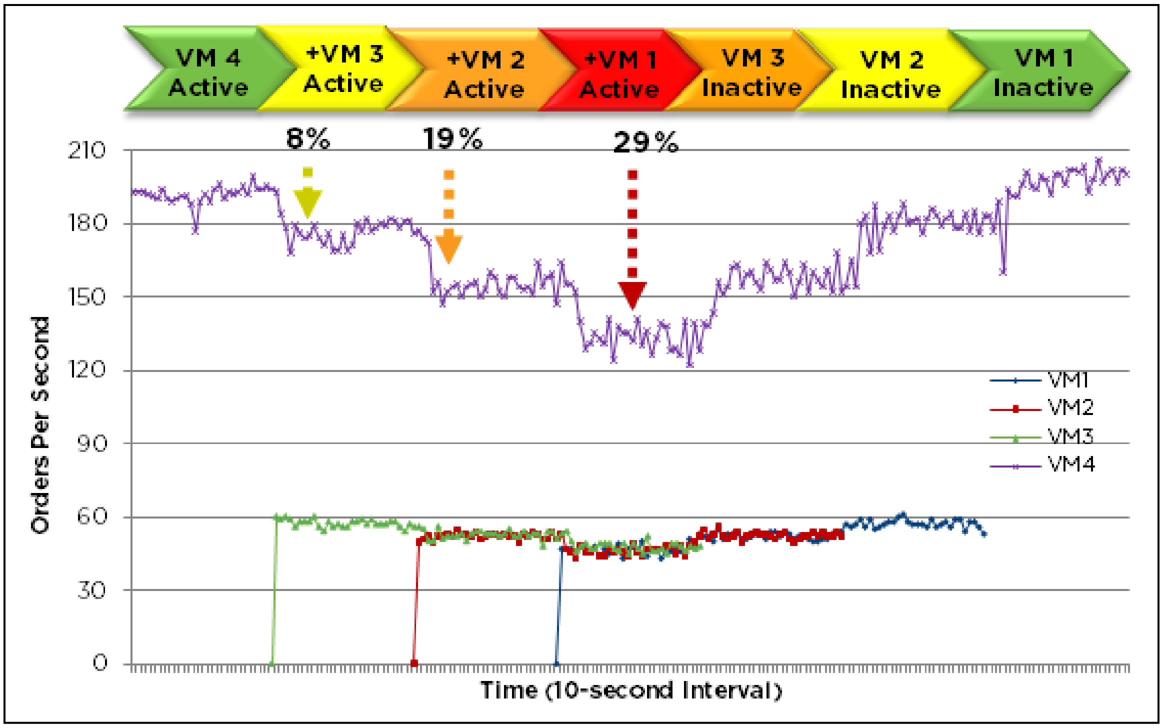


Figure 12. Performance of the DS2 Workload in All 4 VMs with Default Settings

As seen in Figure 12, every time one of the low priority order processing server VMs was active, the performance of the catalog server VM was affected. When order processing server 1 (VM 1) was active, the performance of the catalog server VM dropped by 8%. When order processing server 2 became active, the performance dropped by 19%. Ultimately, when order processing server 3 also became active, the performance of the catalog server VM degraded by 29%. The bottom line is that the performance of the catalog server VM suffered from 7% to 11% every time an order processing server became active, and eventually degraded by as much as 29% from its peak performance because there was no mechanism to manage each VM's access to the shared NFS datastore.

The same experiment was repeated after enabling SIOC for the NFS datastore that was shared between the VMs. This time the congestion threshold was set to 15 milliseconds. Figure 13 shows the results.

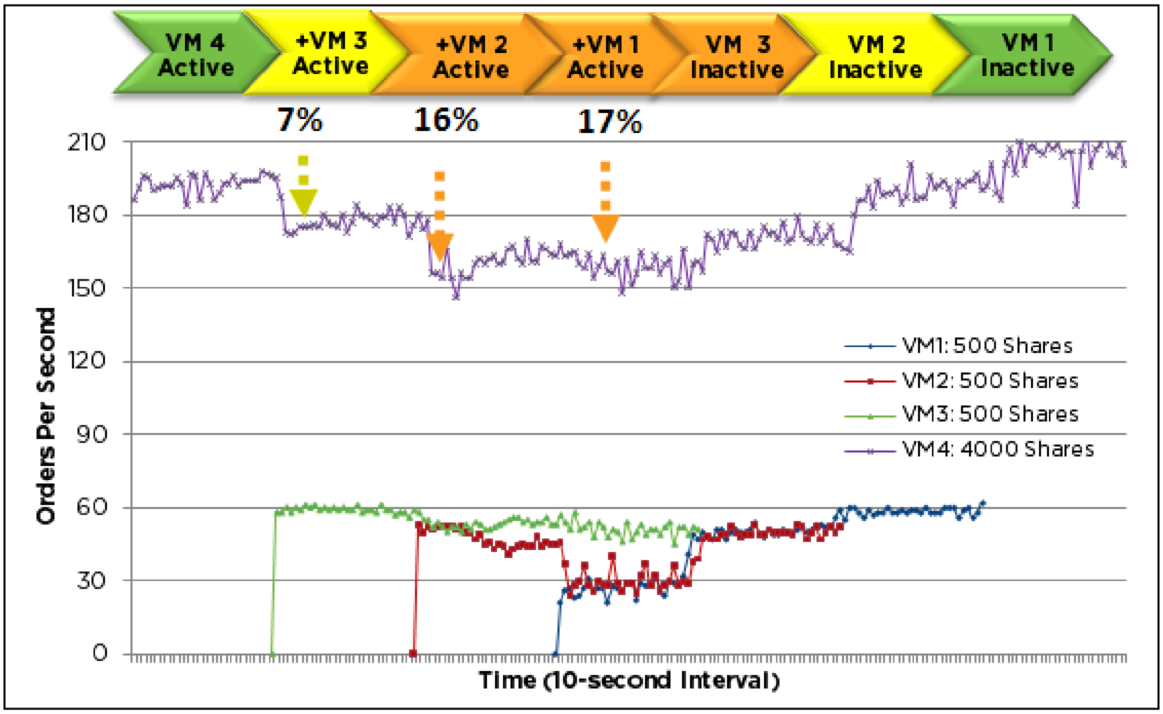


Figure 13. Test Results when SIOC Was Turned On

As one of the order processing server VMs became active and started issuing I/O requests to the shared datastore, the latency of the I/O accesses increased noticeably, but the datastore-wide I/O latency was still under the congestion threshold limit. Therefore, vSphere did not assign any priority to I/O requests from either of the VMs; the VMs were allowed to access the shared datastore equally. The increase in I/O latency resulted in an approximately 8% decrease in performance of the catalog server VM.

When more than one of the order processing server VMs became active, the I/O latency increased above the congestion threshold limit. At this stage, SIOC started managing each VM's access to the shared datastore in proportion to the VM's disk shares. The catalog server VM received the highest share of access because it had the highest number of disk shares. Hence its performance remained mostly unaffected irrespective of the number of active order processing server VMs. The performance of the catalog server VM was 83% of its peak value throughout the congestion period.

Benefits of Storage I/O Control

Figure 14 provides a closer look at the application throughput of the catalog server VM (VM 4) with the default setting and with SIOC enabled on the shared NFS datastore when order processing server VMs 2 and 3 were active in the previous test case.

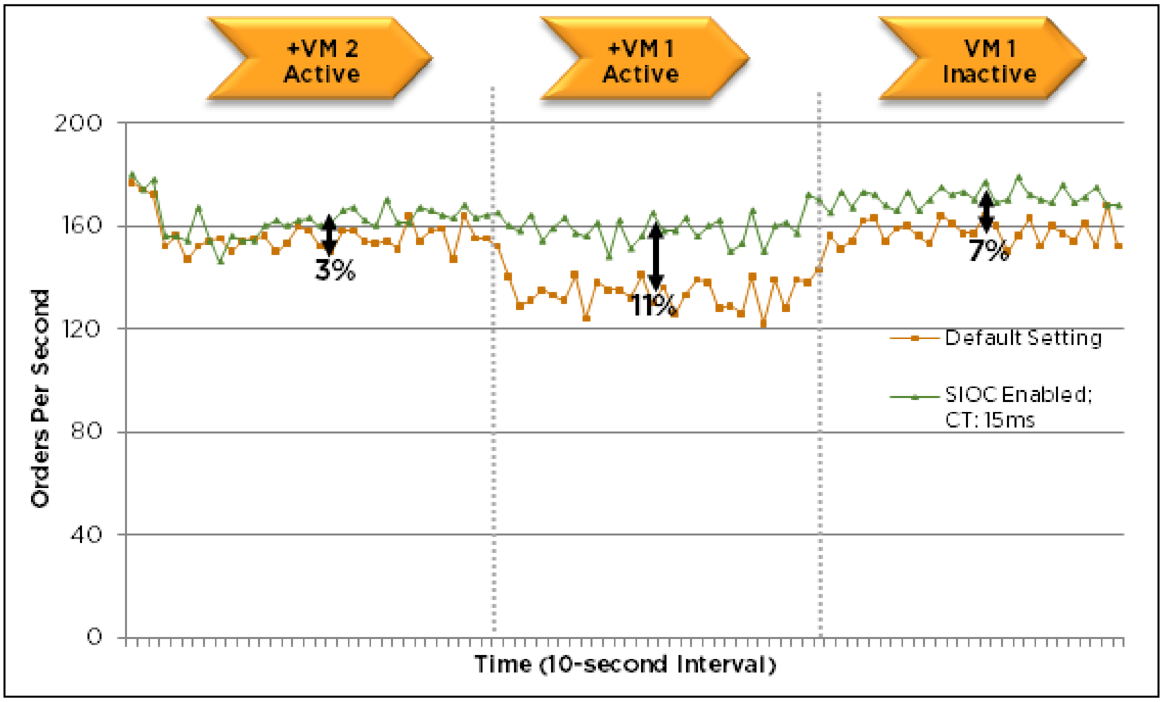


Figure 14. DS2 Throughput of the High Priority Workload with and without SIOC Enabled

From the graph, we can infer that:

- When SIOC was enabled, the loss of application performance in the catalog server VM was predictable and manageable. Without SIOC, the loss in performance was unmanageable. Performance deteriorated every time an order processing server VM became active.
- When SIOC was enabled, the performance of the catalog server VM recovered by 3% to 11% during the congestion period compared to that in the default configuration.

Conclusion

Sharing storage infrastructure to service I/O demands of applications in a virtual infrastructure offers many benefits—better utilization of resources, reduced cost of ownership, higher performance, and ease of management. However, peak demands for I/O resources can lead to contention for I/O resources among the virtual machines. This results in a performance reduction for applications running in those VMs. Storage I/O Control provides a dynamic control mechanism for managing VMs' access to I/O resources in a cluster. The experiments conducted in VMware performance labs showed that:

- SIOC managed each VM's access to shared I/O resources based on the VM's disk shares. This allowed VMs with higher disk shares to maintain a much higher level of performance during peak loads when many VMs were competing for disk I/Os.
- SIOC prevented a significant increase in the I/O latency on a shared NFS datastore when an application issued large sized I/O requests to the datastore. This helped mitigate the impact of the increase in latency on the performance of a latency-sensitive application that issued small sized I/O requests to the same datastore.
- SIOC monitored the utilization of I/O resources and dynamically re-distributed the unused I/O resources to those VMs that actively needed them. This resulted in higher utilization of the storage resource at all times.
- For the test cases executed at VMware labs, SIOC limited the performance fluctuations of a high priority workload to a small range during periods of I/O congestion, resulting in as much as 11% performance benefit compared to that in an unmanaged scenario.

References

1. *vSphere Resource Management Guide*. VMware, Inc., 2011. http://pubs.vmware.com/vsphere-50/topic/com.vmware.vsphere.resmgmt.doc_50/GUID-98BD5A8A-260A-494F-BAAE-74781F5C4B87.html.

2. Voellm, Tony. *Useful I/O Profiles for Simulating Various Workloads*. MSDN, 2009. http://blogs.msdn.com/b/tvoellm/archive/2009/05/07/useful-io-profiles-for-simulating-various-workloads.aspx.

3. Kumar, Chethan. *Managing Performance Variance of Applications Using Storage I/O Control*. VMware, Inc., 2010. http://www.vmware.com/files/pdf/techpaper/vsp_41_perf_SIOC.pdf.

4. Kumar, Chethan, and Ajay Gulati. *TA8233 - Prioritizing Storage Resource Allocation in ESX Based Virtual Environments Using Storage I/O Control*. VMworld, 2010. http://www.vmworld.com/docs/DOC-5117.

5. Gulati, Ajay, Irfan Ahmad, and Carl A. Waldspurger. *PARDA: Proportional Allocation of Resources for Distributed Storage Access*. USENIX Association, 2009. http://www.usenix.org/events/fast09/tech/full_papers/gulati/gulati.pdf.

About the Author

Chethan Kumar is a senior member of Performance Engineering at VMware, where his work focuses on performance-related topics concerning database and storage. He has presented his findings in white papers and blog articles, and technical papers in academic conferences and at VMworld.

Acknowledgements

The author would like to thank Jinpyo Kim for allowing the author to use his hardware resources for the experiments. He would also like to thank members of the Outbound and Storage Performance Engineering, Resource Management, Technical Marketing, and Sales Engineering teams for reviewing this document and providing valuable feedback.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com