WHITE PAPER

Comparison of Storage Protocol Performance in VMware vSphere™ 4

vmware®

Table of Contents

Introduction

This paper compares the performance of various storage protocols available on VMware vSphere™ 4. The protocols Fibre Channel, Hardware iSCSI, Software iSCSI, and NFS are tested using virtual machines on an ESX 4.0 host. Iometer is used to generate the I/O workload.

The Fibre Channel experiments were conducted over a 4Gb Fibre Channel network. The Hardware iSCSI, Software iSCSI, and NFS experiments were conducted over a Gigabit Ethernet connection. Experiments over 10Gb Ethernet and 8Gb Fibre Channel will be included in a future update to this paper.

This paper will start by describing key aspects of the test environment: ESX host, storage array, virtual machines, and Iometer workload. Next, performance results for throughput and CPU cost are presented from experiments involving one or more virtual machines. Finally, the key findings of the experiments are summarized.

The terms "storage server" and "storage array" will be used interchangeably in this paper.

Executive Summary

The experiments in this paper show that each of the four storage protocols (Fibre Channel, Hardware iSCSI, Software iSCSI, and NFS) can achieve line-rate throughput for both single virtual machine and multiple virtual machines on an ESX host. These experiments also show that Fibre Channel and Hardware iSCSI have substantially lower CPU cost than Software iSCSI and NFS.

Experimental Setup

The table below shows key aspects of the test environment for the ESX host, the storage array, and the virtual machine.

ESX Host

Component	Details
Hypervisor	VMware ESX 4.0
Processors	Four Intel Xeon E7340 Quad-Core 2.4GHz processors
Memory	32GB
Fibre Channel HBA	QLogic QLA2432 4Gb
Fibre Channel network	4Gb FC switch
NIC for NFS and SW iSCSI	1Gb (Intel 82571EB)
MTU for NFS, SW iSCSI, HW iSCSI	1500 bytes
iSCSI HBA	QLogic QL4062c 1Gb (Firmware: 3.0.1.49)
IP network for NFS and SW/HW iSCSI	1Gb Ethernet with dedicated switch and VLAN (Extreme Summit 400-48t)
File system for NFS	Native file system on NFS server
File system for FC and SW/HW iSCSI	None (RDM-physical was used)

Storage Array

Component	Details
Storage server	One server supporting FC, iSCSI, and NFS
Disk Drives: Number per data LUN	9
Disk Drives: Size	300Gb
Disk Drives: Speed	15K RPM
Disk Drives: Type	Fibre Channel

Virtual Machine

Component	Details
Guest OS	Windows Server 2008 Enterprise SP1
Virtual processors	1
Memory	512MB
Virtual disk for data	100MB Mapped Raw LUN (RDM-physical)
File system	None (Physical drives were used)
SCSI controller	LSI Logic Parallel

VMware's VMFS file system is recommended for production deployments of virtual machines on iSCSI and Fibre Channel arrays. Because NFS storage presents files and not blocks, VMFS is not needed or possible. VMFS was therefore not used in the Fibre Channel and iSCSI experiments to attempt to produce results that could be compared across all protocols.

I/O Workload

Iometer (http://sourceforge.net/projects/iometer) was used to generate the I/O workload for these experiments. Iometer is a free storage performance testing tool that can be configured to measure throughput and latency under a wide variety of access profiles.

Iometer Workload

Component	Details
Number of outstanding I/Os	16
Run time	2 min
Ramp-up time	2 min
Number of workers	1 (per VM)

Each virtual (data) disk of the virtual machines used in these experiments is 100MB in size. The small size of these virtual disks ensures that the I/O working set will fit into the cache of the storage array. An experiment with a working set size that fits into the cache of the storage array is commonly referred to as a *cached run*.

For read operations in a cached run experiment, the data is served from the storage array's cache, and read performance is independent of disk latencies.

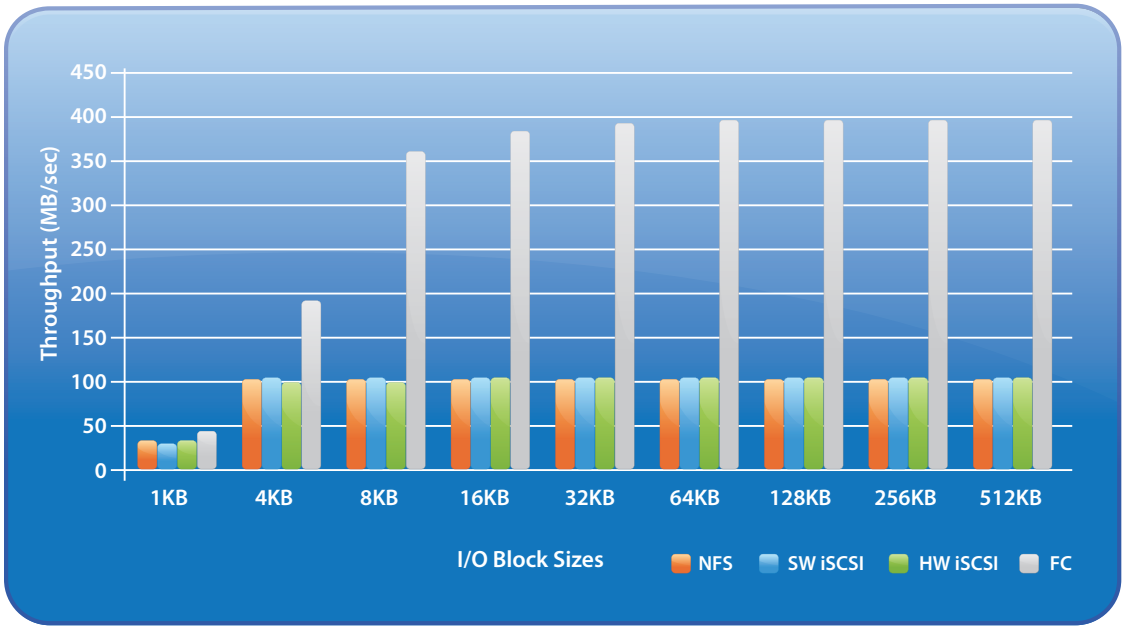
For write operations in a cached run experiment, the rate of write requests at the storage array may exceed the storage array's rate of writing the dirty blocks from the write cache to disk. If this happens, the write cache will eventually fill up. Once the write cache is full, write performance is limited by the rate at which dirty blocks in the write cache are written to disk. This rate is limited by the latency of the disks in the storage array, the RAID configuration, and the number of disk spindles used for the LUN.

For these reasons, read performance for cached runs is a better indication of the true performance of a storage protocol on the ESX host, irrespective of the storage array used.

Experiment: Single VM: Throughput (Read and Write)

Figure 1 shows the sequential read throughput (in MB/sec) of running a single virtual machine in the standard workload configuration for different I/O block sizes, for each of the storage protocols.

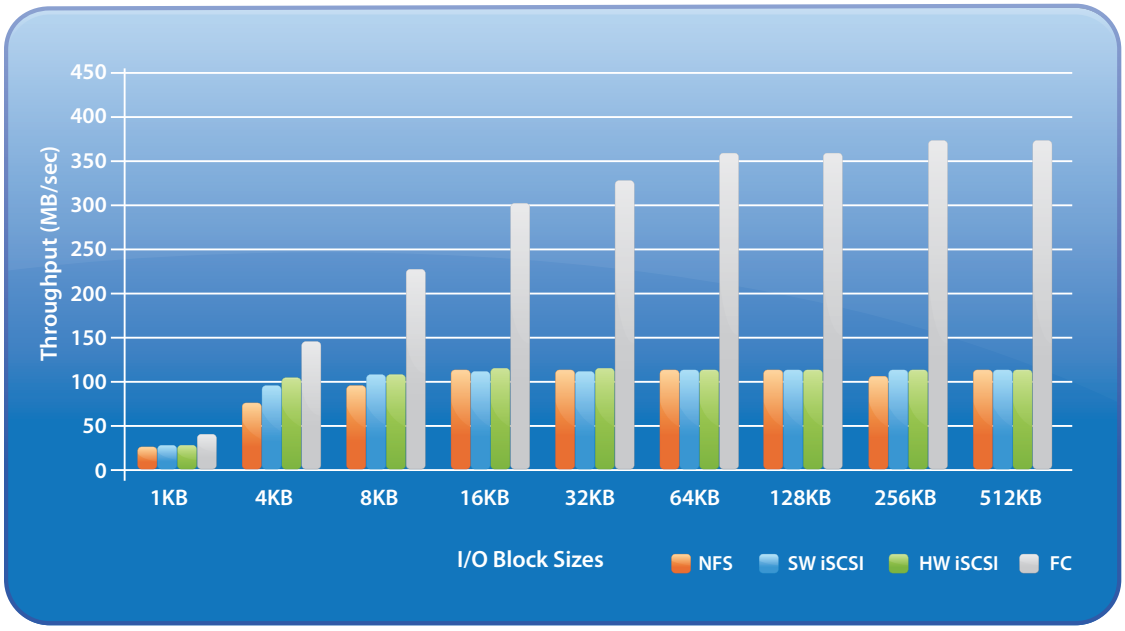
Figure 1: Read throughput for different I/O block sizes



For Fibre Channel, read throughput is limited by the bandwidth of the 4Gb Fibre Channel link for I/O sizes at or above 64KB. For IP-based protocols, read throughput is limited by the bandwidth of the 1Gb Ethernet link for I/O sizes at or above 32KB.

Figure 2 shows the sequential write throughput (in MB/sec) of running a single virtual machine in the standard workload configuration for different I/O block sizes, for each of the storage protocols.

Figure 2: Write throughput for different I/O block sizes



For Fibre Channel, the maximum write throughput for any I/O block size is consistently lower than read throughput of the same I/O block size. This is the result of disk write bandwidth limitations on the storage array. For the IP-based protocols, write throughput for block sizes at or above 16KB is limited by the bandwidth of the 1Gb Ethernet link.

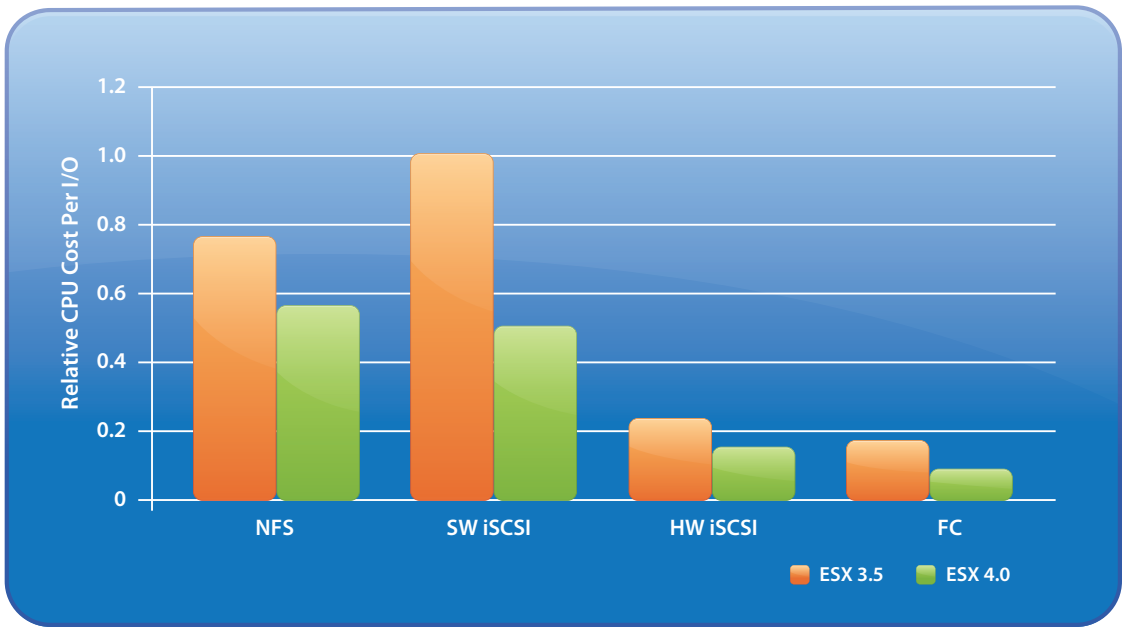
To summarize, a single Iometer thread running in a virtual machine can saturate the bandwidth of the respective networks for all four storage protocols, for both read and write. Fibre Channel throughput performance is higher because of the higher bandwidth of the Fibre Channel link. For the IP-based protocols, there is no significant throughput difference for most block sizes.

Experiment: Single VM: CPU Cost Per I/O (Read)

CPU cost is a measure of the amount of CPU resources used by ESX to perform a given amount of I/O. In this paper, the CPU cost of each storage protocol is measured in units of CPU cycles per I/O operation. The cost for different storage protocols is normalized with respect to the cost of software iSCSI on ESX 3.5.

Figure 3 shows the relative CPU cost of sequential reads in a single virtual machine in the standard workload configuration for a block size of 64 KB for each of the storage protocols. Results on ESX 4.0 are shown next to ESX 3.5 to highlight efficiency improvements on all protocols. The CPU cost of write operations for different storage protocols was not compared as write performance is strongly dependent on the choice of the storage array.

Figure 3: Relative CPU cost of 64 KB sequential reads in a single virtual machine



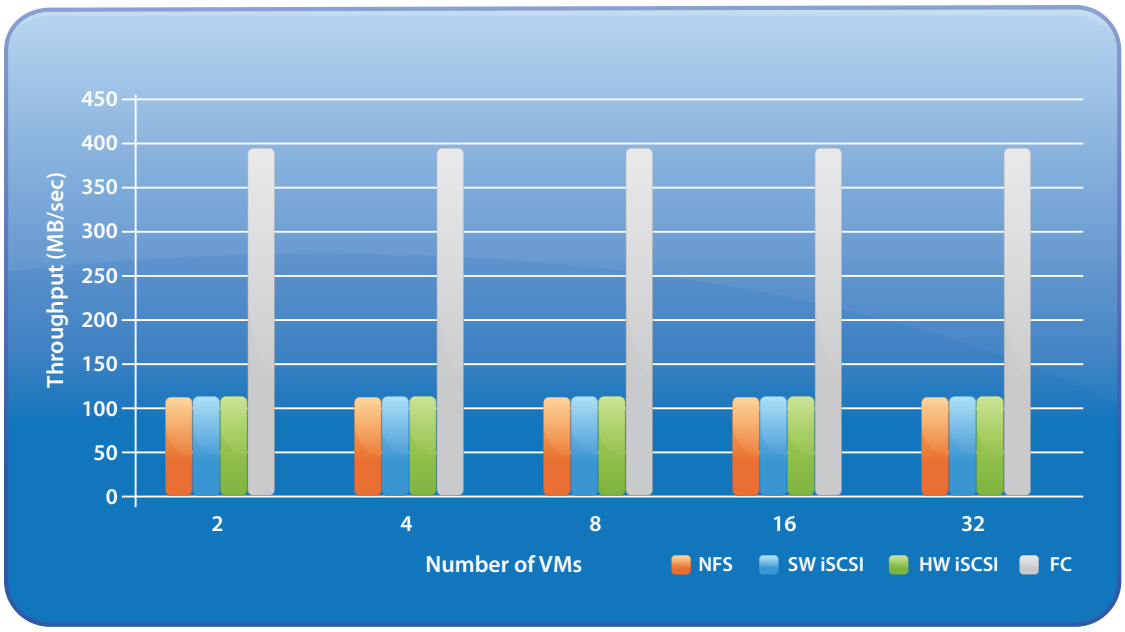
For Fibre Channel and Hardware iSCSI, a major part of the protocol processing is offloaded to the HBA, and consequently the cost of each I/O is very low. For Software iSCSI and NFS, host CPUs are used for protocol processing which increases cost. Furthermore, the cost of NFS and Software iSCSI is higher with larger block sizes, such as 64 KB. This is due to the additional CPU cycles needed for each block for check summing, blocking, etc. Software iSCSI and NFS are more efficient at smaller blocks and are both capable of delivering high throughput performance when CPU resource is not a bottleneck, as will be shown in the next section.

The cost per I/O is dependent on a variety of test parameters, such as platform architecture, block size, and other factors. However, these tests demonstrate improved efficiency in vSphere 4's storage stack over the previous version.

Experiment: Multiple VMs: Aggregate Throughput (Read)

Figure 4 shows the aggregate sequential read throughput (in MB/sec) of running 2, 4, 8, 16, and 32 virtual machines in the standard workload configuration for a block size of 64KB. Each virtual machine performs I/O to its dedicated 100MB LUN.

Figure 4: Throughput of running multiple VMs in the standard workload configuration



For each of the storage protocols, the maximum aggregate throughput is limited by the network bandwidth. There is no degradation in aggregate throughput for a large number of virtual machines.

Conclusion

In this paper, the performance of four storage protocols was compared for accessing shared storage available on VMware ESX 4.0: Fibre Channel, Hardware iSCSI, Software iSCSI, and NFS.

All four storage protocols for shared storage on ESX are shown to be capable of achieving throughput levels that are only limited by the capabilities of the storage array and the connection between it and the ESX server. ESX shows excellent scalability by maintaining these performance levels in cases of heavy consolidation. For CPU cost, Fibre Channel and Hardware iSCSI are more efficient than Software iSCSI and NFS. However, when CPU resources are not a bottleneck, Software iSCSI and NFS can also be part of a high-performance solution.

Earlier versions of ESX were also able to achieve throughput levels that are only limited by the array and bandwidth to it. VMware vSphere 4 continues to support this maximized throughput but can do so with greater efficiency. Improved efficiency on vSphere 4 means the same high levels of performance with more virtual machines.

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

VMW_09Q2_WP_VSPHERE_StorageProtocols_P8_R1

vmware®