Comparison of Storage Protocol Performance

ESX Server 3.5

This study provides performance comparisons of various storage connection options available to VMware® ESX Server. We used the widely used IOmeter benchmark for the comparison.

The results show that all four network storage options can reach wire-speed throughput when properly configured. The wire-speed throughput is also maintained with multiple virtual machines driving concurrent I/Os, indicating that the limiting factor in performance scalability is not in ESX Server.

The data also demonstrates that, although Fibre Channel has the best performance in throughput, latency, and CPU efficiency among the four options, iSCSI and NFS are also quite capable and may offer a better cost-to-performance ratio in certain deployment scenarios.

This study covers the following topics:

Hardware and Software Environment

Tables 1 and 2 provide details of the test environment.

Table 1. ESX Server host and storage components

Component	Details
Hypervisor	VMware ESX Server 3.5
Processors	Four 2.4GHz dual-core AMD Opteron processors
Memory	32GB
NICs for NFS and software iSCSI	1Gb on board NICs
Fibre Channel network	2Gb FC-AL
Fibre Channel HBA	QLogic QLA2340
IP network for NFS and iSCSI	1Gb Ethernet with a dedicated switch and VLAN
iSCSI HBA	QLogic QLA4050C
File system	Fibre Channel and iSCSI: None. RDM-physical was used. NFS: Native file system on the server.
Storage server/array	One server supporting FC, iSCSI, and NFS. (All protocols accessed the same LUNs.)

Table 2. Virtual machine configuration

Component	Details
Operating system	Windows
Version	2003 Enterprise
Number of virtual processor	1
Memory	256MB
Virtual disk size for data	100MB to achieve cached runs effect
File system	None. Physical drives are used.

To obtain optimal performance for large I/O sizes, we modified a registry setting in the guest operating system. For details, see http://www.lsilogic.com/files/support/ssp/fusionmpt/Win2003/symmpi2003_11002.txt.

NOTE Because we used no file system in the data path in the virtual machine, no caching is involved in the guest operating system. As a result, the amount of memory configured for the virtual machine has no impact on the performance in these tests if it satisfies the minimum requirement.

Workload

We used the IOmeter benchmarking tool (http://sourceforge.net/projects/iometer), originally developed at Intel, to generate I/O load for these tests. We held the IOmeter parameters shown in Table 3 constant for each individual test.

Table 3. IOmeter configuration

Parameter	Value
Number of outstanding I/Os	16
Run time	3 minutes
Ramp-up time	60 seconds
Number of workers to spawn automatically	1

To focus on the performance of storage protocols, we adopted the cached run approach often used in the industry when there is a need to minimize latency effects from the physical disk. In such an experimental setup, the entire I/O working set resides in the cache of the storage server or array. Because no mechanical movement is involved in serving a read request, maximum possible read performance can be achieved from the storage device. This also ensures that the degree of randomness in the workload has nearly no effect on throughput or response time and run-to-run variance becomes extremely low.

However, even in cached runs, write performance can still depend on the available disk bandwidth of the storage device. If the incoming rate of write requests outpaces the server's or array's ability to write the dirty blocks to disk, after the write cache is filled and steady state reached, a new write request can be completed only if certain blocks in the cache are physically written to disk and marked as free. For this reason, read performance in cached runs better represents the true performance capability of the host and the storage protocol regardless of the type of storage server or array used.

Performance Results

Figure 1 shows the sequential read throughput in MB/sec achieved by running a single virtual machine in the standard workload configuration through each storage connection option.

The Fibre Channel results indicate that for I/O sizes at or above 64KB, the limitation presented by a 2Gb link is reached. As for all IP-based storage connections, the 1Gb wire speed is reached with I/O sizes at or above 16KB.

Figure 1. Single virtual machine throughput comparison, 100 percent sequential read (higher is better)

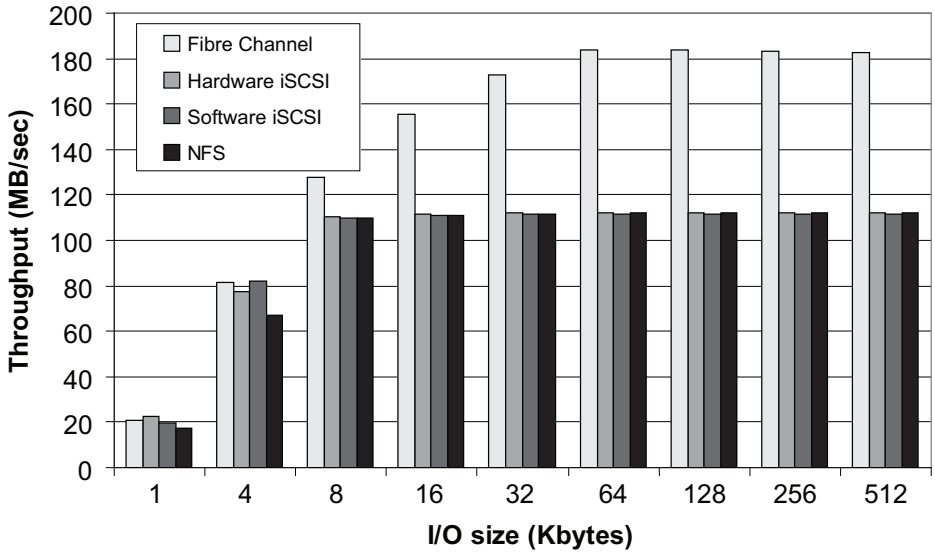


Figure 2 shows the sequential write throughput comparison. Because of the disk bandwidth limitation from the storage sever or array, the maximum write throughput is consistently lower than read throughput. Also, the slightly lower Fibre Channel performance with block sizes of 256KB and beyond is a result of the inability of the storage devices to handle large block sizes well. This is entirely an effect of the storage device, not the ESX Server host. If the intended workload constantly issues large writes, see VMware knowledge base article 1003469 "Tuning ESX Server 3.5 for Better Storage Performance by Modifying the Maximum I/O Block Size" (http://kb.vmware.com/kb/1003469) for a workaround for this issue.

Figure 2. Single virtual machine throughput comparison, 100 percent sequential write (higher is better)

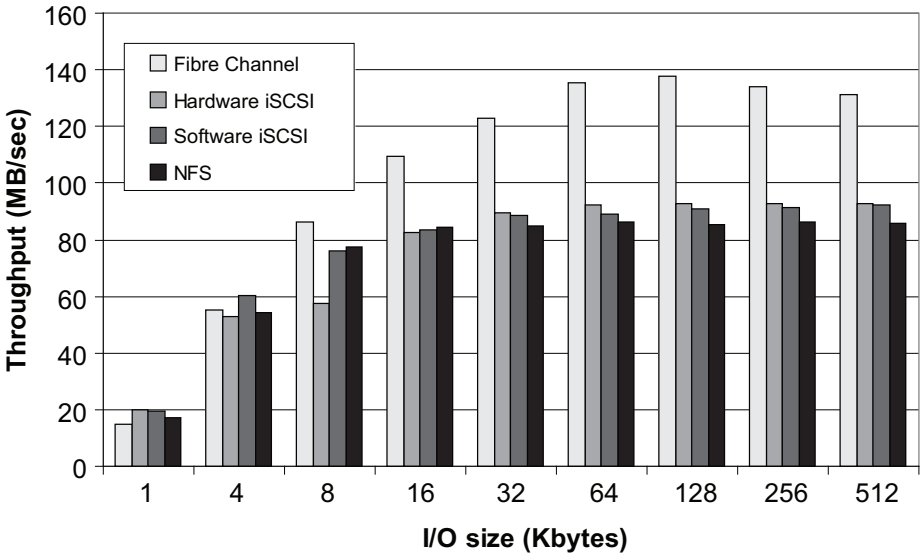


Figure 3 and Figure 4 show the response-time measurements from the same tests. All the IP-based options have very similar characteristics in processing read requests, though NFS shows slightly longer response times in processing large writes.

In general, Fibre Channel has the performance advantage in larger I/O sizes mainly because of its higher wire speed. The difference in throughput and response time among the three IP-based storage connections is negligible in nearly all cases.

Figure 3. Single virtual machine response time comparison, 100 percent sequential read (lower is better)

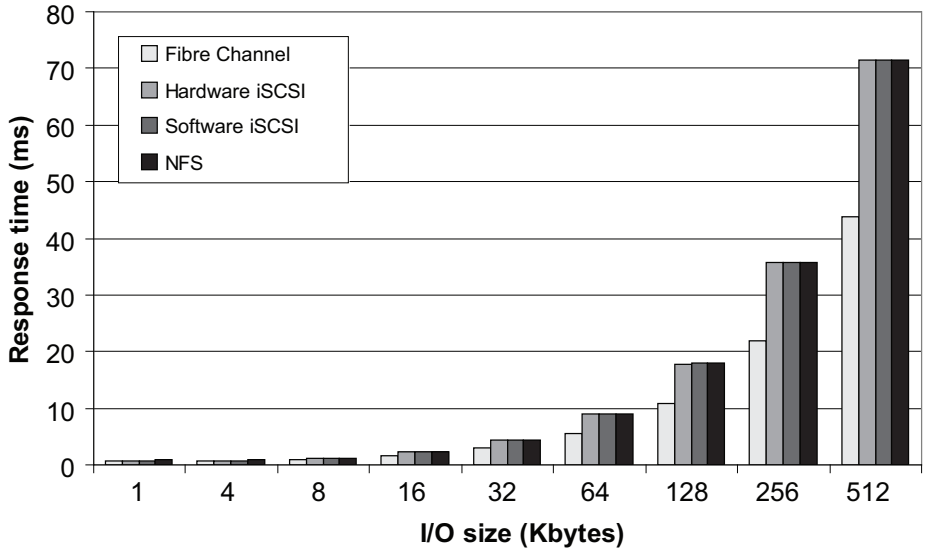
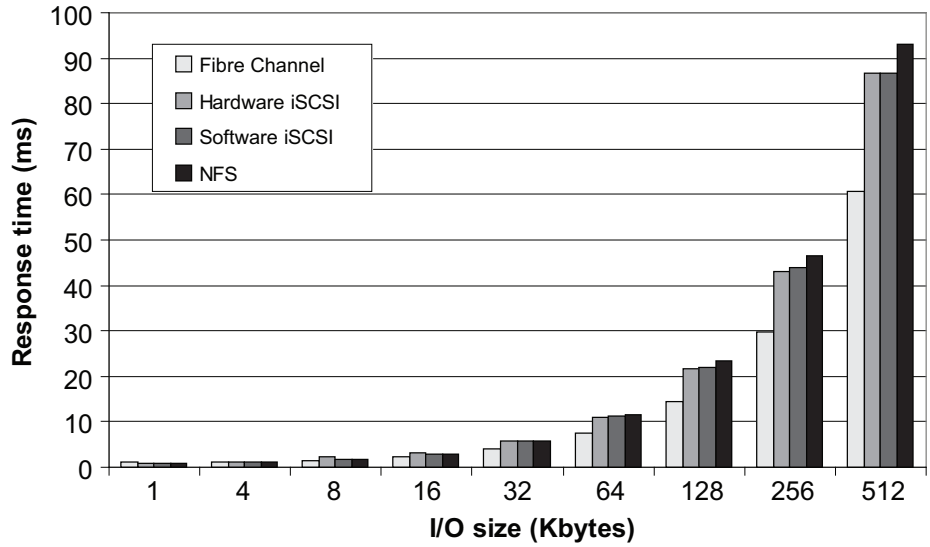
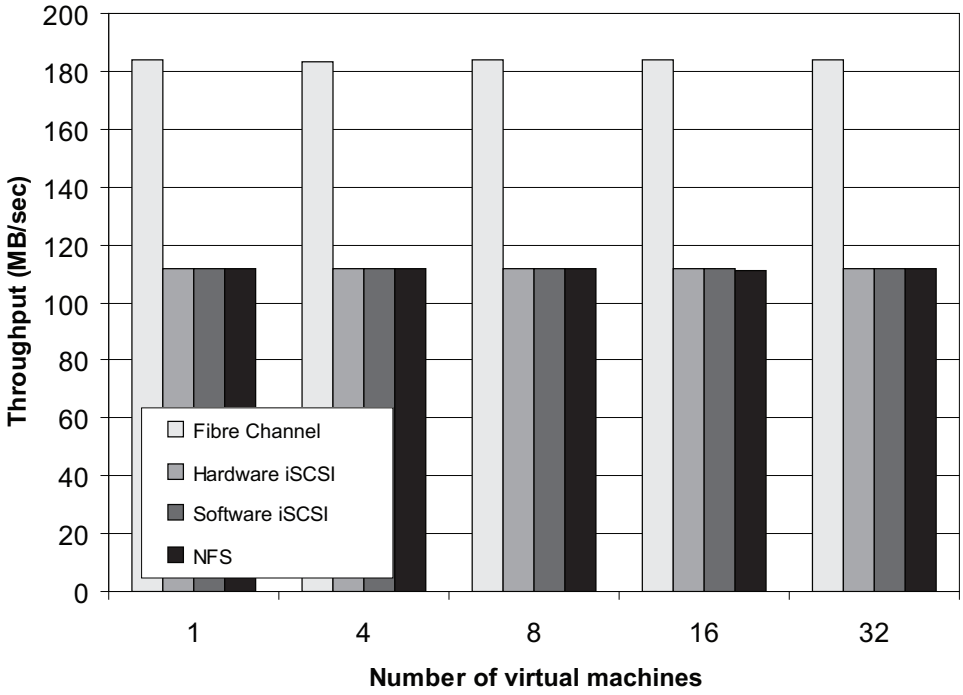


Figure 4. Single virtual machine response time comparison, 100 percent sequential write (lower is better)



When you deploy a hypervisor such as ESX Server to consolidate physical servers, you expect multiple virtual machines to perform I/O concurrently. It is therefore very important that the hypervisor itself does not become a performance bottleneck in such scenarios. Figure 5 shows the aggregate throughput of 1, 4, 8, 16, and 32 virtual machines driving 64KB sequential reads to each virtual machine's dedicated LUN. In all cases, the data is moving at wire speed. Clearly, in each storage connection option available to ESX Server, wire speed is maintained up to 32 virtual machines, showing that wire speed, rather than ESX Server, limits the scalability.

Figure 5. Multiple virtual machine scalability, 100 percent sequential read (higher is better)

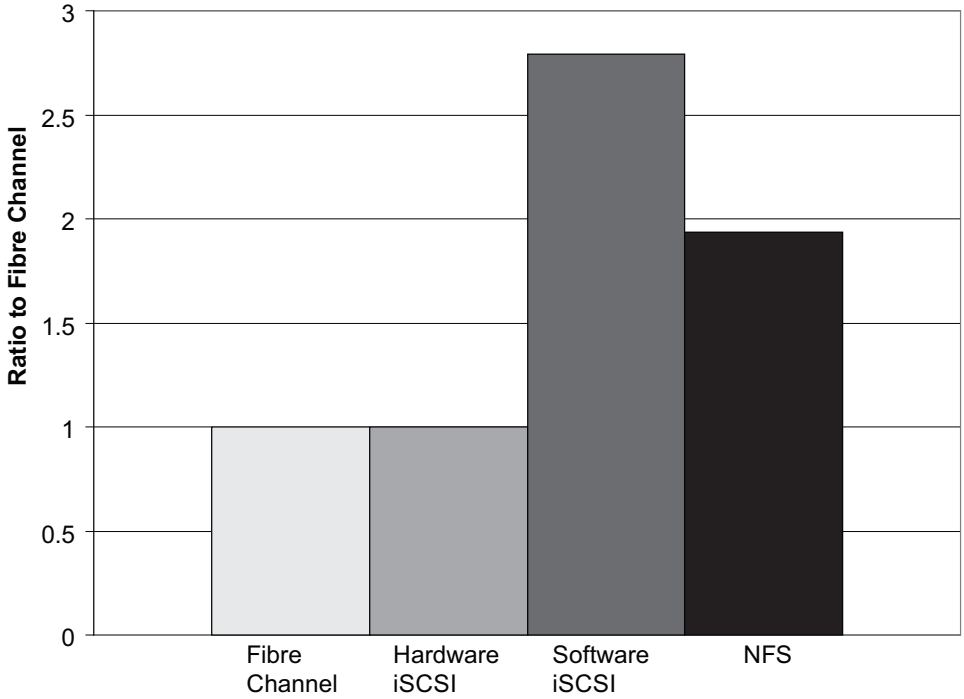


The data shown so far has demonstrated that all storage connection options available to ESX Server perform well, achieving and maintaining wire speed even in the cases with large numbers of virtual machines performing concurrent disk I/Os. In light of this, because of its intrinsic wire-speed advantage, especially if the 4Gb technology is deployed, Fibre Channel is likely the best choice if maximum possible throughput is the primary goal. Nevertheless, the data also demonstrates other options being quite capable and, depending on workloads and other factors, iSCSI and NFS may offer a better price-performance ratio.

In addition to aggregate throughput, another consideration is the CPU cost incurred for I/O operations. When you consolidate storage-I/O heavy applications, given a fixed amount of CPU power available to the ESX Server host, the CPU cost of performing I/O operations may play a major role in determining the consolidation ratio you can achieve.

Figure 6 shows the CPU cost compared to the Fibre Channel cost when a virtual machine is driving a mix of 50 percent reads and 50 percent writes at the 8KB block size, which is most common in enterprise-class applications such as database servers. Fibre Channel and hardware iSCSI HBAs both offload a major part of the protocol processing from the host CPUs and thus have very low CPU costs in most cases. For software iSCSI and NFS, host CPUs are involved in protocol processing, so higher CPU costs are expected. However, software iSCSI and NFS are fully capable of delivering the expected performance, as shown in earlier results, when CPU resource availability does not become a bottleneck.

Figure 6. CPU cost per I/O compared to Fibre Channel cost for 8KB 50 percent read and 50 percent write (lower is better)



Conclusion

This paper demonstrates that the four network storage connection options available to ESX Server are all capable of reaching a level of performance limited only by the media and storage devices. And even with multiple virtual machines running concurrently on the same ESX Server host, the high performance is maintained. The data on CPU costs indicates that Fibre Channel and hardware iSCSI are the most CPU efficient, but in cases in which CPU consumption is not a concern, software iSCSI and NFS can also be part of a high-performance solution.

VMware, Inc. 3401 Hillview Ave., Palo Alto, CA 94304 www.vmware.com

Revision 20080122 Item: PS-047-PRD-01-01