



blade.org™

BLADE
NETWORK
TECHNOLOGIES

Chelsio
Communications
Accelerate

NetApp



EXECUTIVE BRIEF

10GbE iSCSI PERFORMANCE WITH BLADE SERVERS

Blade Network Technologies | Chelsio Communications | Network Appliance, Inc.
October 2007

PERFORMANCE TESTING 10GbE WITH iSCSI ON IBM BLADECENTER

Our test of network throughput capabilities compared performance of 10GbE iSCSI with 4Gb/s Fibre Channel using an actual database workload running on a reference architecture consisting of an IBM BladeCenter H with embedded 10GbE blade switches from Blade Network Technologies (BLADE) and mezzanine network cards for the blade servers and storage controllers from Chelsio Communications (Chelsio), and a NetApp FAS6030C clustered storage system. Results demonstrate that Oracle 10g RAC runs at near wire speeds—8.9 Gb/s using NFS and 7.2 Gb/s using iSCSI—in consolidated blade server architectures using networked storage. When compared with Fibre Channel results of 3.4Gb/s, these results prove that 10GbE environments are viable and can support large-scale business applications.

TABLE OF CONTENTS

Table of Contents	2
1 BACKGROUND	3
2 ACKNOWLEDGEMENTS	4
3 ORACLE TEST	4
3.1 ENVIRONMENT	4
3.2 ORACLE TEST PLAN	6
3.3 ORACLE RESULTS	7
3.4 ORACLE KEY FINDINGS	7
4 PRELIMINARY TESTS USING MICROSOFT® EXCHANGE.....	7
4.1 PRELIMINARY OBSERVATIONS.....	8
5 IMPLICATIONS.....	9
6 TRENDS AND EVOLUTION.....	10
7 CONCLUSION	10
8 RESOURCES.....	11

1 BACKGROUND

As data continues to grow at an alarming rate, data centers are hitting the limits of their growth. They are often constrained by both power and floor space. Blade servers were developed in response to a wide-spread need in the data center for increased server performance and availability without a corresponding increase in footprint, cost, and management complexity. Blade server technology greatly increases server density, lowers power and cooling costs, eases server expansion and simplifies datacenter management. It is no wonder that blade servers have become an important component in server consolidation and virtualization efforts.

To gain the full benefits of blade server technologies and to provide headroom for future growth require the latest in networking technology. Today that means 10 Gigabit Ethernet (10GbE)—the most recent and fastest Ethernet standard. The combination of the iSCSI protocol and 10GbE offers key advantages over other networking technologies used in blade environments. The first advantage is lower cost. The cost per gigabit of throughput of iSCSI with 10GbE is one third the cost of 4Gb/s Fibre Channel. Because IP networks are everywhere, iSCSI is easier to implement than Fibre Channel (FC) and requires no specialized knowledge. Equipment costs for iSCSI are also lower than for FC.

Together, Blade Network Technologies (BLADE), Chelsio Communications (Chelsio), and NetApp have created a reference architecture to demonstrate that consolidated environments comprised of blade servers with state-of-the-art networks and flexible and scalable storage can help control costs, infrastructure sprawl, and data growth. Multi-protocol support (NFS, iSCSI, CIFS) across a 10GbE network combined with integrated data management practices and techniques ideally suited for blade servers, such as thin provisioning and network boot, combine to provide the ultimate in flexibility and investment protection.

We realize that until a new combination of technologies is characterized, tested, and proven, the perceived risks of deploying the new technologies can hinder their adoption. To demonstrate that you do not have to sacrifice performance for a lower-cost network alternative, we created a test of network throughput capabilities to compare the performance of 10GbE NAS and iSCSI with 4Gb/s Fibre Channel using an Oracle® database workload. Our reference architecture consisted of an IBM BladeCenter H with embedded 10GbE blade switches from BLADE, mezzanine network cards from Chelsio for each blade server, and a NetApp FAS6030C clustered storage system with network interface cards from Chelsio.

We hope that our work will contribute to the adoption of blade technologies with 10GbE networks.

2 ACKNOWLEDGEMENTS

This project originated within blade.org, an industry organization dedicated to accelerating the development and adoption of open blade server platforms and fostering the blade community. Blade Network Technologies, Chelsio, IBM, and Network Appliance collaborated in the design and execution of the performance testing described in this brief.

We would like to acknowledge the team members who contributed to this project.

BLADE NETWORK TECHNOLOGIES	CHELSIO COMMUNICATIONS	IBM	NETAPP
Charles Ferland	Bruck Girmay	Silvio Erdenberger	Bharat Badrinath
Tracy Hickox	Troy Leedberg	Olaf Menke	Lee Dorrier
David Iles	Kianoosh Naghshineh	Ishan Seghal	John Elliott
Scott Lorditch		Marek Uroda	Gregg Ferguson
Shailesh Naik			Ilka Fock
Fred Rabert			Vinod Herur
William Scull			Chris Reno
			Eric Saillard
			Stefan Schiechl
			Juergen Seipel

3 ORACLE TEST

3.1 ENVIRONMENT

The primary goal of our testing was to compare the throughput potential of 10GbE to 4Gb/s Fibre Channel in a blade-server environment with a database workload to validate the viability of using 10GbE and iSCSI with blade technologies. To test performance, we simulated a production-ready data center environment using components that are readily available to most customers. The test environment consisted of an IBM BladeCenter H, embedded 10GbE blade switches from BLADE, mezzanine network cards for each blade server provided by Chelsio, and a NetApp FAS6030C clustered storage system which also used network interface cards from Chelsio. To simulate typical I/O work loads, we used a Decision Support System (DSS) workload generator for Oracle10g™.

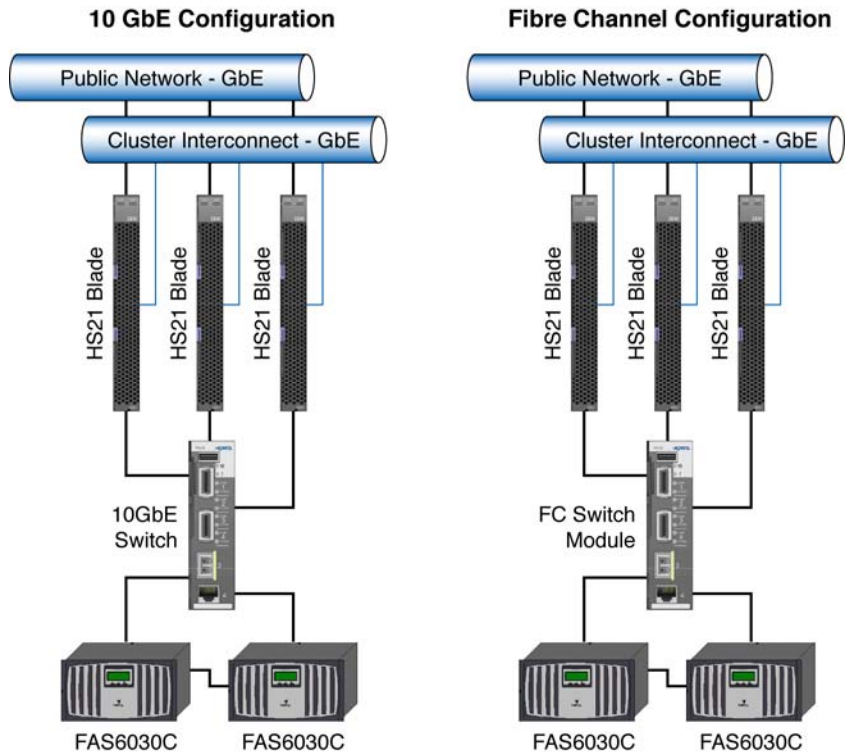


Figure 1) 10GbE and 4Gb/s Fibre Channel Test Bed Topologies

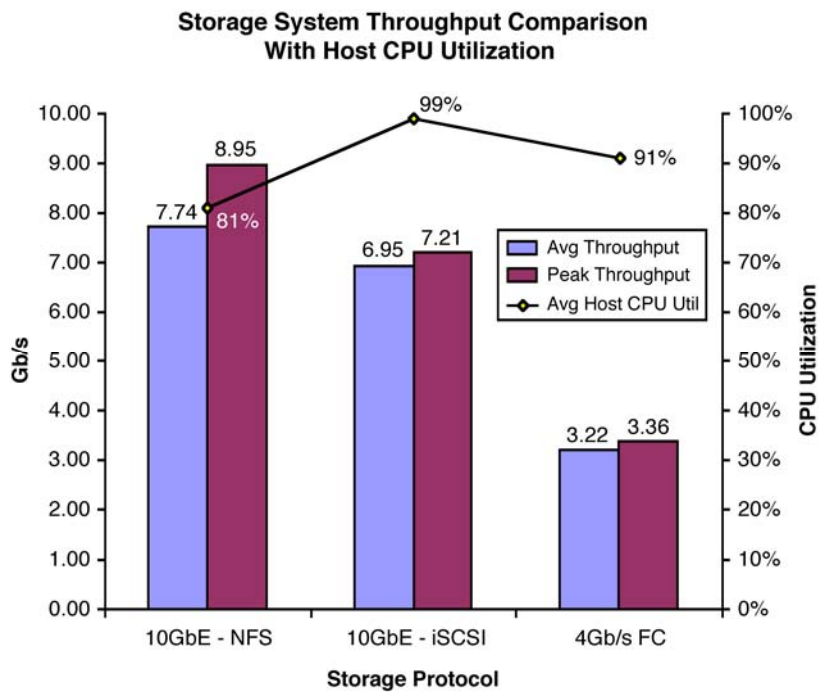


Figure 2) I/O Throughput Comparison with Host CPU Utilization

The test bed consisted of a three-node Oracle RAC database configured on three IBM HS-21 blades. Each node had a single 10GbE or 4Gb/s Fibre Channel path to the NetApp storage system depending on the test. Each node executed database queries designed to maximize the throughput generated across respective 10GbE and Fibre Channel interconnects. The generated I/O was sequential in nature and generally typical of a read-intensive Decision Support System workload. Refer to Figure 1 for a picture of the test bed topology and Table 1 for a description of the tested configurations.

Table 1) Oracle Test Environment Components

	10GbE with NFS	10GbE with iSCSI	FC
Operating System	Red Hat Enterprise Linux® Release 4, Update 4		
Storage Controller Software	NetApp Data ONTAP® 7.2.3		
RDBMS	Oracle10g R2 RAC		
Blade Chassis	IBM BladeCenter H		
Blade Servers	3 IBM HS-21		
Processors per blade	2 Dual Core 2.66 GHz Xeon		
PCI Express Mezzanine Cards	Chelsio S320em-BC cards		
Fibre Channel HBAs per blade			4Gb/s FC HBAs
Switches	Blade Network Technologies 39Y9265: 10GbE Layer 2/3 switch module blades		Embedded FC switch
Networked Storage	NetApp FAS6030C Active-Active, highly scalable, enterprise-class storage		
Storage Controller Network Interface Cards	Chelsio T210-SR cards		4Gb/s FC HBAs
Test Data	3-node Oracle database		
Load-generation tool	DSS Workload Generator		

3.2 ORACLE TEST PLAN

Our plan was to test the same Oracle workload running on the same servers and storage with three different protocols: 10GbE with NFS, 10GbE with iSCSI, and 4Gb/s Fibre Channel (FC). We used a workload generator tool to create an Oracle query that simulates a typical DSS workload. The function of the query is to analyze the parts inventories of suppliers across several different nations and determine their suitability for promotional offers. Query execution involved sequential scanning of several tables along with joins and sort operations, which in turn generated a high volume of direct sequential read I/O from storage. Due to the high data throughput requirements of the queries, we used the Oracle parallel query feature. In order to isolate disk performance from network throughput, we optimized the database to minimize disk I/O. As a result, the storage system was able to cache in memory most of the data accessed by the test queries, thereby avoiding latencies associated with normal disk I/O. By controlling disk I/O in this way, we have ensured that the data presented in the paper very closely reflects actual network throughput capabilities. By design, the configurations were tuned so that the DSS load generation tool running on each host node generated as much I/O as possible across each interconnect. This accounts for the high host CPU utilization measured during the tests.

3.3 ORACLE RESULTS

The throughput of each test configuration is reported in Gigabits per second (Gb/s). Average Throughput is defined as the average I/O throughput measured on the interconnects (GbE or FC) from both controllers of the NetApp storage system for the duration of a single test run. Peak Throughput is defined as the actual peak I/O throughput during a single test run.

3.4 ORACLE KEY FINDINGS

10GbE out-performs FC. Throughput rates of the two 10GbE configurations were clearly higher than the 4Gb/s FC (the average 10GbE throughput was more than double that of FC).

Peak throughput approaches theoretical limits. In each case, the peak results were close to the theoretical throughput maximum for the given interconnect.

NFS currently offers best throughput. Throughput rates using NFS averaged 11 percent greater than rates achieved using iSCSI. The difference can best be explained by the increased CPU requirements of the iSCSI software initiator. Note: This gap between the two protocols is expected to disappear as the implementation of TCP offload (TOE) capabilities for iSCSI becomes available in future driver releases.

Table 2) Throughput and Host CPU Utilization Comparisons

Metric	10GbE NFS	10GbE iSCSI	4Gb/s/s FC
Avg Throughput Across Both Storage Controllers (Gb/s)	7.74	6.95	3.22
Peak Storage Throughput (Gb/s)	8.95	7.21	3.36
Host CPU Utilization (Avg Across 3 Nodes)	81%	99%	91%
Peak Host CPU Utilization (Across 3 Nodes)	85%	100%	98%

4 PRELIMINARY TESTS USING MICROSOFT® EXCHANGE

We were interested to know if our Oracle results would hold up for other applications. To check it out, we used our 10GbE network reference architecture and substituted Windows® Server 2003 for RedHat Enterprise Linux and substituted Microsoft Exchange for an Oracle database. Testing at IBM's testing lab center in Frankfurt, Germany, focused on throughput and IOPS comparisons between 10GbE with iSCSI and 4Gb/s Fibre Channel. We used the iSCSI Software Initiator tool from Microsoft. This tool can be downloaded from the Microsoft website. Network components were implemented with default values, and there was no special tuning of the environment for performance to simulate a typical Exchange environment.

Table 3) Microsoft Exchange Test Environment Components

	10GbE with iSCSI	FC
Operating System	Windows Server 2003	
Storage Controller Software	NetApp Data ONTAP 7.2.3	
Blade Chassis	IBM BladeCenter H	
Blade Servers	2 IBM HS-21	
Processors per blade	2 Dual Core 2.66 GHz Xeon	
PCI Express Mezzanine Cards	Chelsio S320em-BC cards	
Switches	Blade Network Technologies 39Y9265: 10GbE Layer 2/3 switch module blades	Embedded 4G FC switch
Networked Storage	NetApp FAS6030C Active-Active, highly scalable, enterprise-class storage	
Network Interface Cards	Chelsio T210-SR cards	
Host Bus Adapters		4G Fibre Channel Target cards per controller
Load-generation tool	Exchange Jetstress Exchange Workload Generator	
Scalable I/O (SIO) tool	Basic I/O load generator	

4.1 PRELIMINARY OBSERVATIONS

10GbE out-performs FC. iSCSI with 10GbE using the Microsoft Software iSCSI Initiator enables superior storage performance using 4Gb/s Fibre Channel SANs, with a cost that can be 1/3 as much per gigabit of throughput.

More users per blade. Improved latency under Microsoft Exchange enables more users per blade server. The network is no longer the bottleneck.

Easy implementation. iSCSI was easy to set up using NetApp tools, which are fully integrated into the Windows OS and Microsoft Exchange.

The SIO tool enabled us to simulate traffic between the blade servers and the storage and allowed us to measure end-to-end throughput. Notice on the following graphic that block size has a profound impact on throughput performance for both 10GbE and 4Gb/s Fibre Channel. It is expected that as applications are further aggregated in consolidated and virtualized environments, that optimal block sizes and the number of simultaneous threads will require careful tuning to prevent bottlenecks throughout the infrastructure. The lessons learned from this study will assist customers as they deploy 10GbE environments.

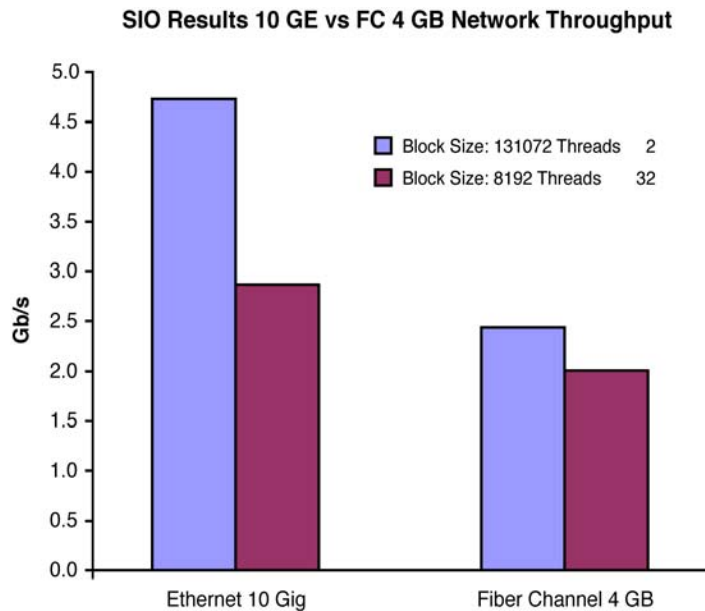


Figure 3) Microsoft Server 2003 Throughput Comparisons

More work in this area is required in order to truly characterize the performance of an Exchange environment supporting thousands of mailboxes and users.

Based upon the results of our preliminary tests, 10GbE environments have dramatically better throughput than 4 Gb/s Fibre Channel. It can also be stated that 10GbE is better positioned to support heterogeneous applications, operating systems, and protocols (NFS, iSCSI). *Unified Wire* environments are not a thing of the future—they exist today with 10GbE.

5 IMPLICATIONS

Deploying our reference architecture for 10GbE server-to-switch-to-storage infrastructure in blade environments offers several potential advantages:

Investment protection. Blade server architectures not only enable consolidated processing environments, they also have the interconnect technologies to provide the type of investment longevity that customers expect. Because you can use existing network expertise, management tools, and processes, you can avoid costs associated with implementing a new network.

Cost savings. When combined with Storage Systems that can support all aspects of data management services, from boot from SAN to non-disruptive data protection to secure data archive, customers can realize the flexibility and bandwidth they require from this “unified wire” technology for their dynamic data center environments without adding parallel management networks.

Headroom for growth. We expect that typical business workloads will not reach the same levels of CPU utilization and throughput that we simulated in our tests. As a result, it is reasonable to assume 10GbE networks will provide the necessary headroom for deploying highly consolidated workloads such as virtualized application and operating system instances, with technology that is available today.

More improvements. TCP Offload Engine (TOE) was not used for these performance tests, and it is expected that enabling this capability will only result in improved throughput and better CPU utilization.

6 TRENDS AND EVOLUTION

Several near-term industry trends in storage and network technology will affect the decision process for considering 10GbE and other high-speed options in data center environments. By far the most significant variable will be operating system virtualization technologies from VMware, Microsoft, Oracle and other Xen-based vendors (Citrix/XenSource and Virtual Iron).

CPU Utilization. A major concern with deploying 10GbE has been the high levels of CPU utilization that it causes. Heavy CPU utilization occurs because of the increased number of packets that must be received, processed, managed, and replied to at significantly higher rates. Previous throughput of 1GbE is now multiplied by 10, regardless of latency considerations. A number of emerging technologies will relieve this processing burden.

Quad Core Processors. Both Intel (Core 2, Q6600, Jan. 2007) and AMD ("Barcelona", Sept. 10, 2007) have announced and made available quad-core processors that are designed specifically for consolidated workloads, video streaming, and multimedia environments. According to each vendor's claims, it is expected that this type of processing, along with advance features in terms of cooling and raw performance, will continue to allow large scale and multi-threaded workloads to be easily supported over the next two years or so. We expect that the projected price points for these processors will make their adoption and use in specialized network interface cards a near-term, cost-effective reality.

TCP/IP Off-Load Engines. TCP/IP off-load engines relieve high CPU utilization by incorporating a processor that performs data packet handling in the network interface card. The idea is to make certain to utilize the full bandwidth of the network connection without overwhelming the CPU responsible for business application processing. However, an even more compelling reason arose that is driving the adoption of technologies that off-load processing—consolidation. As more applications, and therefore more users, are consolidated, processing that was focused on network traffic would have a noticeable impact on application performance and therefore user satisfaction. Combine consolidated day-to-day processing with other network-intensive activities, such as data backup and restore, and it becomes clear that TOE implementations can help keep operations running at optimal levels.

pNFS and File System Advances. In addition to advances in connectivity technology, file system updates from storage and operating system vendors will also be affecting the performance and characterization of application and data management. One advance in particular, Parallel NFS (pNFS), will improve performance of clustered environments by allowing file requests to multiple servers or storage devices at the same time (gigabytes/sec) as opposed to the serial request process prevalent in NFS implementations today (megabytes/sec). The IETF has a working group called NFSv4 that is expected to ratify the standard by the end of calendar year 2007.

7 CONCLUSION

Based upon the performance results summarized above, 10GbE server-to-switch-to-storage infrastructures are a viable alternative to other interconnect technologies used in blade environments. 10GbE provided better than twice the throughput of 4Gb/s Fibre Channel. Given our results, it is reasonable to expect that 10GbE throughput performance will compare favorably with the next generation of 8Gb/s Fibre Channel.

8 RESOURCES

About BLADE Network Technologies

BLADE Network Technologies is the market-leading supplier of Gigabit and 10G Ethernet, IP, and Application Switches for blade server systems globally. BLADE is the first vendor to focus exclusively on serving the network infrastructure needs of the rapidly growing blade server market. The company's end-users include Fortune 500 companies across 26 different industry segments. As the number-one supplier of network switches for blade servers, BLADE boasts an installed base of more than 120,000 blade switches (representing approximately 2.7 million switch ports) around the world. In 2006, BLADE was established as a fully independent company when it purchased certain assets of Nortel's Blade Server Switch Business Unit. For more information, visit BLADE at www.bladenetwork.net.

Tim Shaughnessy
(408) 850-8963
press@bladenetwork.net

About Chelsio Communications

Chelsio Communications is leading the convergence of networking, storage and clustering interconnects with its robust, high-performance and proven unified wire technology. Featuring a highly scalable and programmable architecture, Chelsio is shipping 10-Gigabit Ethernet and multi-port Gigabit Ethernet adapter cards, delivering the low latency and superior throughput required for high-performance computing applications. For more information, visit the company online at www.chelsio.com.

Tim Helms
(925) 606-6936
press@chelsio.com

About Network Appliance, Inc.

Network Appliance is a leading provider of innovative data management solutions that simplify the complexity of storing, managing, protecting, and retaining enterprise data. Market leaders around the world choose NetApp to help them reduce cost, minimize risk, and adapt to change. For solutions that deliver unmatched simplicity and value, visit us on the Web at www.netapp.com.

Adam Mendoza
(210) 240-1738
Adam.Mendoza@netapp.com

[c] 2007 Blade Network Technologies Inc., Chelsio Communications Inc., and Network Appliance, Inc. All rights reserved. Specifications subject to change without notice. NetApp, the Network Appliance logo, and Data ONTAP are registered trademarks and Network Appliance is a trademark of Network Appliance, Inc. in the U.S. and other countries. Microsoft and Windows are registered trademarks of Microsoft Corporation. Linux is a registered trademark of Linus Torvalds. Oracle is a registered trademark and Oracle10g is a trademark of Oracle Corporation. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such.