Choosing the optimum fabric for your DSS workloads
April 2010

Fibre Channel, FCoE and iSCSI Performance Comparison for DSS Workloads with Microsoft SQL Server 2008

By Vinay Kulkarni
IBM System x

Kent Swalin
IBM Center for Microsoft Technologies

Wei Liu
NetApp

Executive Overview

This document details a series of tests conducted in the Microsoft® SQL Performance Lab in Redmond by IBM® and NetApp® to compare the performance of 10Gb Fibre Channel over Ethernet (FCoE) and iSCSI protocols with 4Gb Fibre Channel (FC), using a Microsoft SQL Server® 2008 Decision Support System (DSS) workload on an IBM System x3850 X5 server, NetApp FAS3070 storage systems and Emulex® OneConnect OCe10102-F 10Gbps FCoE Adapters.

Introduction

Two common storage protocols, Fibre Channel (FC) and iSCSI, have been used and deployed extensively in the data center. Fibre Channel is the prevalent technology standard in the Storage Area Network (SAN) data center environment. iSCSI solutions have been primarily been used for smaller and mid-sized storage using 1Gb Ethernet. Until recently, the iSCSI bandwidth was limited to 1Gbps on a single connection, due to the use of Gigabit Ethernet (GbE) in existing networks. At least four 1GbE connections were needed to equal the bandwidth of a 4Gbps FC connection. This bandwidth limitation made it complex to implement 1GbE iSCSI solutions in large data centers where high data throughput is a requirement.

Another new protocol, Fibre Channel over Ethernet (FCoE), has emerged. FCoE is not meant to displace or replace FC. Rather, FCoE unifies the two leading-edge technologies, Fibre Channel protocol and enhanced 10Gb Ethernet physical transport, to provide customers with more options for SAN connectivity and networking.

With today's 10GbE networks becoming more affordable, 10GbE iSCSI and FCoE are options for data centers looking for high performance and cost-effective solutions.

The systems tested in this paper ran identical SQL Server DSS workloads across the three different protocols: FC, FCoE and iSCSI. In these tests the server and storage remained the same, only the storage protocols were changed. All three protocol tests were configured similarly using a switch (or two switches for FC) between the storage and the server. It is important to note that we also kept the total maximum bandwidth as close as possible for these three protocol tests. We used eight 4Gb FC ports from the server for a total bandwidth of 32Gbps. For iSCSI and FCoE we used three 10GbE ports from the server for a total bandwidth of 30Gbps. The total bandwidth difference in these protocols was no more than 6.3%.

The IBM server in these tests was running Microsoft Windows® Server 2008 R2 and SQL server 2008. The Decision support workload was chosen for the comparison test because this workload is frequently very I/O intensive and bandwidth-intensive. In these tests IBM, NetApp, and Emulex wanted to push the interconnects between the storage and the server to see how they compare under an intensive load.

Test Topology

Figure 1 illustrates the topology of the 10GbE and 4Gb FC testing. The server and storage remained the same for all three tests, while the storage protocol and network were changed for each different test.

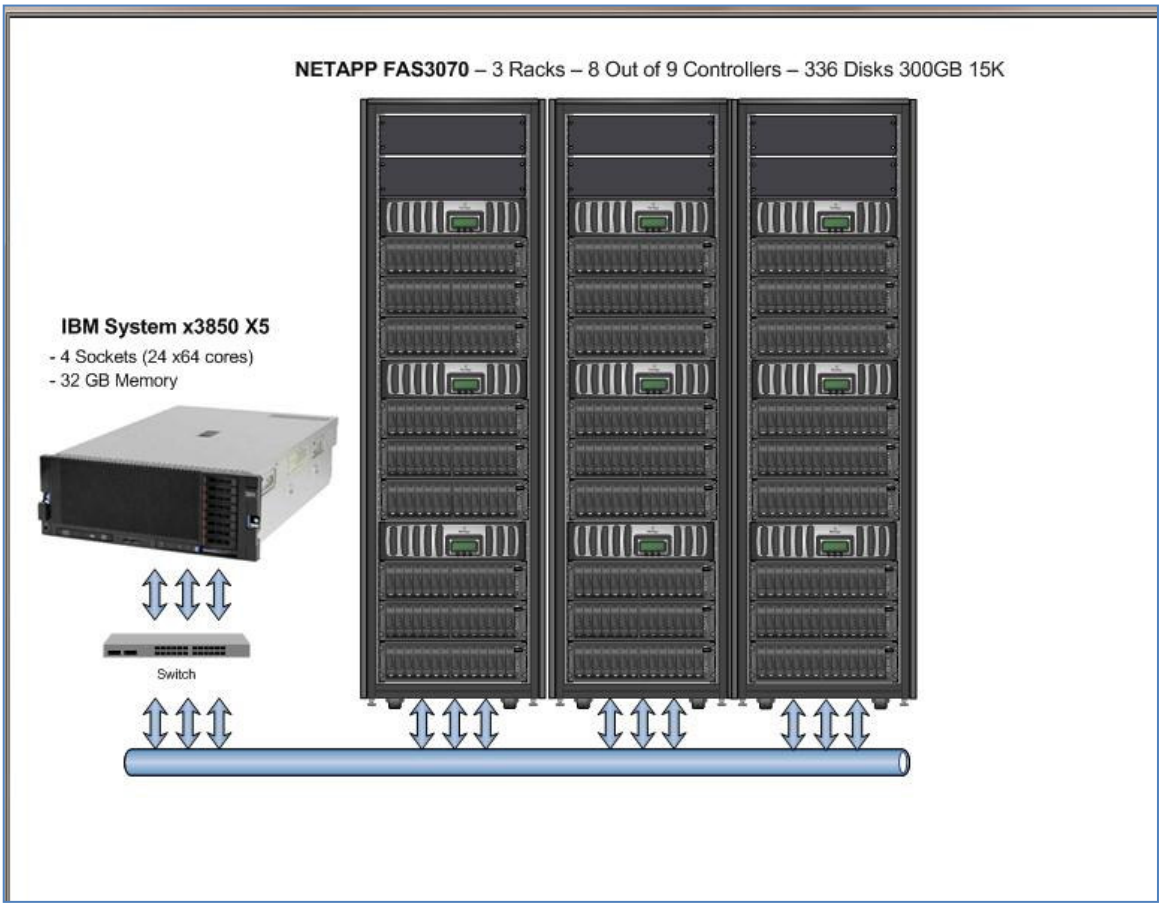

Figure 1. *Topology for the FC, FCoE and iSCSI tests*

FC Test

The high I/O requirement of this test was met by using NetApp FAS3070 storage systems connected to the IBM x3850 X5 server using four dual-port Emulex LPE12002-M8 adapters running at 4Gbps via a pair of Brocade 200E 4Gbps FC switches. The server was connected using eight 4Gbps FC interconnects for a maximum bandwidth of **32Gbps**.

FCoE and iSCSI Tests

The high I/O requirements of these tests were met by using NetApp FAS3070 storage systems connected to the IBM System x3850 X5 server using two dual-port Emulex OneConnect OCe10102 10GbE adapters. The server was connected using three 10GbE interconnects via 50/125 µm multi-mode fibre (OM3) for a maximum bandwidth of **30Gbps**. A Cisco Nexus 5010 switch was used for the iSCSI and FCoE tests.

Components Used to Build the Test Environment

The following sections describe the major components used in the test configurations.

IBM System x3850 X5 Server

The release of the IBM System x3850 X5 server marks another significant advance forward in server technology. These new systems incorporate the latest in hardware acceleration and

scalability providing resource configurations that push the boundaries of x86/x64-based systems while providing industry-leading flexibility.

The System x3850 X5 enterprise servers offer up to 64 logical processors, up to 1TB of memory with 64 16GB DIMMs, 7 PCIe Gen 2 slots and up to 8 internal drives per chassis. (Two chassis can be interconnected to create an **8-processor/64-core/128-thread/14 I/O-slot** system.) Powered by Intel® Xeon® 7500 Series processors with up to eight cores, the System x3850 X5 is designed to provide superior processing, memory, and I/O for high-performance environments. In fact, if 2TB of memory isn't enough, you can use the exclusive IBM MAX5 memory expansion unit to add another 512GB of memory per chassis, for up to 1.5TB for a single chassis server, or 3TB for a dual-chassis system. Industry-leading performance per watt delivers significantly reduced energy consumption compared to previous generation servers. An easy upgrade path provides the necessary flexibility to continue delivering an optimized solution for scale-up database and enterprise applications as your requirements continue to grow.

The System x3850 X5 was set up running Windows Server 2008 R2 and SQL Server 2008. The server had two dual-port Emulex OneConnect OCe10102-F 10Gbps FCoE adapters running the standard Emulex driver for Windows Server 2008 R2 x64 (driver version 2.30.016) . The Emulex OneCommand™ Manager application was used to monitor FCoE connectivity on the server.

The server was configured with the following hardware:

- 4 Intel Xeon 6-core X7542 2.66GHz processors
- 16 2GB DIMMs, for a total of 32GB memory
- 2 146.8GB HDDs with RAID-1
- 1 IBM ServeRAID®-M5015 SAS/SATA controller

Intel Xeon Processor 7500 Series

Get a dramatic increase in performance, efficiency, and reliability with servers powered by the highly intelligent and scalable Intel Xeon 7500 Series processor. Built to handle your most demanding applications, the 7500 Series delivers a huge leap in enterprise computing performance over previous generations of processors and servers.

It combines up to 8 cores and 16 processing threads in a single device and offers four advanced, high-bandwidth Quick Path Interconnect (QPI) links that allow multiple processors to be directly connected to one another. The result is eight-socket systems capable of processing 128 threads simultaneously.

You also get the reliability you need to run the most demanding applications with complete confidence while maintaining data integrity and minimizing downtime. The 7500 Series features Intel Advanced Reliability Technology that provides automatic detection and correction of errors, dynamic reassignment of workloads across processors, interconnect error detection/recovery, and individual virtual machine recovery in virtualized environments. In essence, it delivers everything you need to maintain data integrity, minimize downtime, and maximize productivity.

NetApp FAS3070 Storage

NetApp fabric-attached storage (FAS) systems simplify data management, enabling enterprise customers to reduce costs and complexities, minimize risks, and control change. NetApp FAS systems are the most versatile storage systems in the industry for storage consolidation. The FAS3070 addresses the core requirements of the midrange enterprise storage market, delivering a superb blend of price, performance, and scalability for SQL Server databases and business applications. The compact, modular design provides native support for FCoE, FC, iSCSI and NAS storage with scalability to over 500 disk drives. The FAS3070 storage controller supports FC,

SAS and SATA disk drives for tiered storage. FAS3070 systems support as many as 32 FC ports or 32 Ethernet ports, including support for 2Gb, 4Gb, and 8Gb FC, and 10Gb Ethernet.

The FAS3070 runs the NetApp Data ONTAP® operating system, which is optimized for fast, efficient, and reliable data access and retention. Data ONTAP 7G dramatically simplifies common storage provisioning and management operations. LUNs and volumes created and configured using FlexVol® technology can be dynamically expanded or contracted with a single command. FlexVol volumes also enable thin provisioning, which avoids the cost of overprovisioning and the time-consuming reconfiguration typical with other storage solutions. Host-based NetApp SnapDrive® extends this flexible storage provisioning capability to databases and applications. Another Data ONTAP 7G feature, FlexClone®, instantaneously creates cloned LUNs or volumes without requiring additional storage. FlexClone technology can dramatically improve the effectiveness and productivity of application and database development and predeployment testing.

FAS hardware design and the Data ONTAP operating system are tightly integrated to provide resilient system operation and high data availability. FAS systems incorporate redundant and hot-swappable components and patented double-parity RAID-DP®. NetApp RAID-DP, a high-performance implementation of RAID 6, provides superior data protection with negligible impact on performance. NetApp Snapshot™ technology provides up to 255 data-in-place, point-in-time images per LUN or file system, available for near instantaneous file-level or full data set recovery. The minimal performance overhead of NetApp Snapshot technology makes it well suited for protecting production data. Host-based SnapManager® software integrates Snapshot management with applications, providing consistent backup images and application-level recovery in minutes. SnapMirror® uses Snapshot copies to provide incremental block-level synchronous and asynchronous replication; SnapVault® uses it for block-level incremental backups to another system. Together, these SnapSuite™ products help deliver the high application-level availability that enterprises require for 24×7 operation.

For the series of tests described in this paper, NetApp Snapshot technology was used to streamline and speed up the testing process. *Figure 2* shows the test procedures. After the initial database creation, NetApp Snapshots were taken on all eight controllers. After the FC and FCoE tests, the Snapshots were restored and the databases were returned to their initial state, saving several hours needed for recreating databases. The SnapRestore operation took less than a minute.

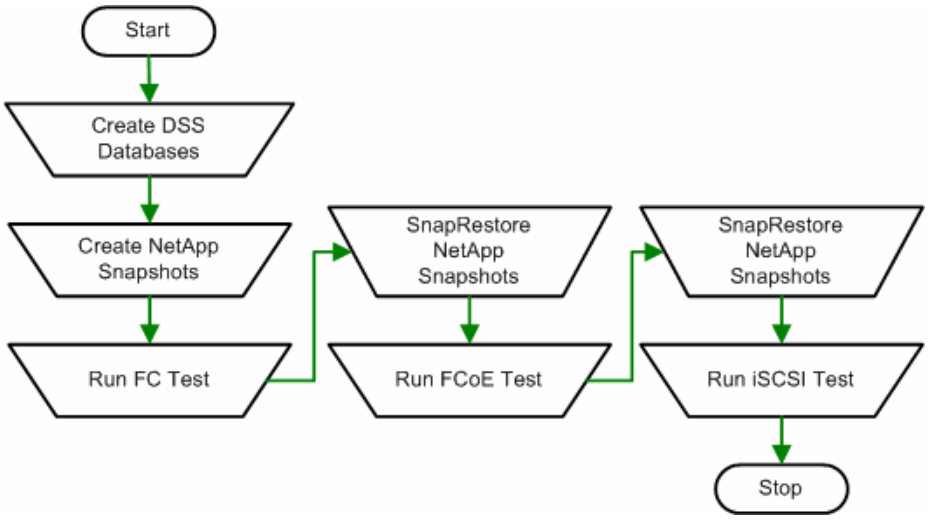

Figure 2. *Test procedures and the use of NetApp Snapshot technology*

Emulex LPE12002 Adapter

Emulex creates enterprise-class products that intelligently connect storage, servers and networks. Every Emulex LightPulse® Host Bus Adapter (HBA) leverages eight generations of advanced, field-proven technologies to deliver a distinctive set of benefits that are relied upon by the world's largest enterprises.

Key Features

- PCIe 2.0 bus-compliant, with PCIe 1.0 backward compatibility supported
- Support for Message Signaled Interrupts eXtended (MSI-X), improves host utilization and enhances application performance
- Support for 8Gbps, 4Gbps, and 2Gbps Fibre Channel devices

Emulex OCe10102 Adapter

Emulex OneConnect Universal Converged Network Adapters (CNAs) are a family of single-chip, high-performance 10Gbps Ethernet (10GbE) multi-function adapters that provide server connectivity for network and storage traffic. Unlike standard 10GbE NICs or first generation CNAs, OneConnect provides optimized network and storage performance with protocol offloads for TCP/IP, TCP Offload Engine (TOE), Fibre Channel over Ethernet (FCoE) and iSCSI. OneConnect adapters deliver maximum performance, regardless of the mix of network traffic.

Key Features

- Enhanced Ethernet convergence with high-performance FCoE protocol offload
- Superior network support with TCIP/IP and TCP Chimney offloads
- Easy to deploy and manage with the OneCommand™ Manager application
- Enterprise-ready with hardware parity, CRC, ECC and other advanced error checking

Results and Analysis

The DSS workload consists of a series of SQL Queries that stress the system from both a CPU performance and I/O performance perspective. We have chosen to highlight one of the queries that requires the highest I/O bandwidth to show the difference between the protocols in the charts below.

Peak and Average Bandwidth

Figure 3 shows the peak and average throughput for this I/O-intensive query. The *peak* throughput for FC and FCoE was identical at **3.17GBps**. The FCoE test during peak loads was running at nearly the wire speed of the three interconnects from the server to the storage systems. The peak throughput for iSCSI was slightly lower at **2.93GBps**, due to the protocol overhead. The difference is within 7.6%. For these tests the Emulex iSCSI initiator was used rather than the Microsoft software iSCSI initiator, because Emulex CNAs offered both TCP and iSCSI offloading.

The *average* throughput for the FC, FCoE and iSCSI tests was virtually identical. The FCoE test showed a slightly higher average throughput than either FC or iSCSI. The average bandwidth used for FCoE was **2.82GBps;** FC was 2% slower at **2.76GBps**, and iSCSI was 4% slower at **2.66GBps**.

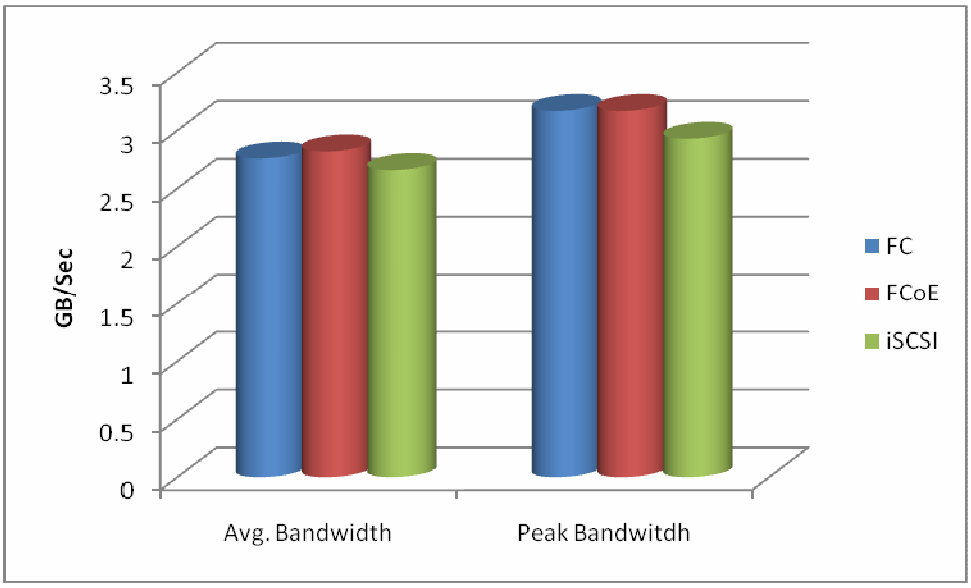

Figure 3. *Average and peak bandwidths used by the DSS query*

Processor Utilization

Figure 4 shows the processor utilization of the 4-socket 24-core server running the protocols during the I/O-intensive query that drove the bandwidth above. The average CPU utilization for iSCSI at **35%** is, as expected, slightly higher than that of FCoE at **30%** or FC at **29%**.

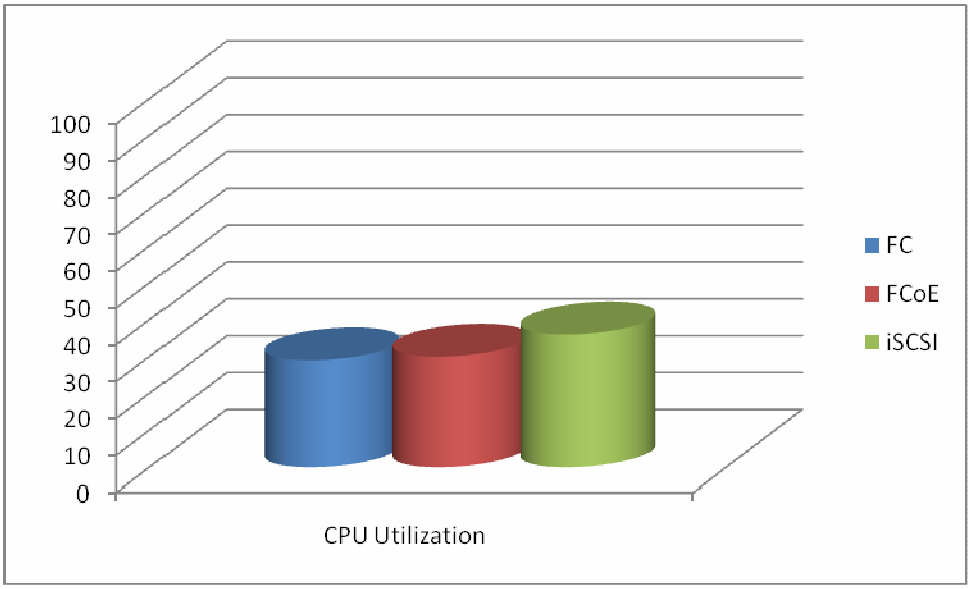

Figure 4. *Server average CPU utilization during the DSS query*

Total Execution Time

Figure 5 shows the total execution time for all queries in this workload. Using FC as the baseline at 100%, FCoE is faster by **5.73%** and iSCSI trails by **0.8%**.

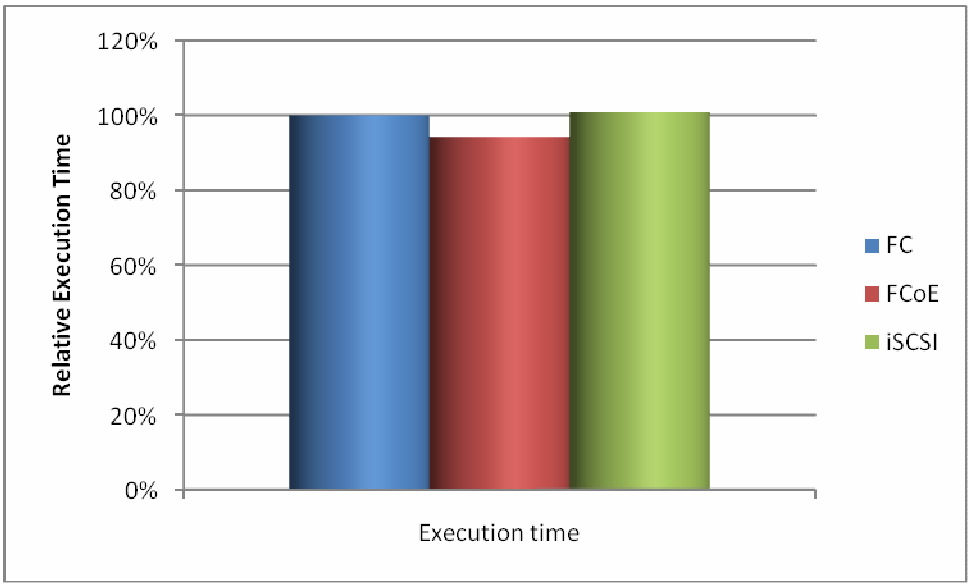

Figure 5. *Comparison of the total DSS workload execution time relative to that of FC*

Conclusion

Our results demonstrated that for an SQL Server 2008 DSS workload, all three protocols performed equally well on average (within 4% of the FC baseline), while 10GbE FCoE offered slightly better performance than 4Gb FC or 10GbE iSCSI in terms of peak throughput and query execution time, given a comparable maximum bandwidth.

The IBM System x3850 X5 server demonstrated that it possesses the processing power, memory, and I/O bandwidth performance needed to support this DSS workload.

The NetApp FAS3070 storage systems selected for this SQL Server 2008 DSS workload provided the required performance for this workload type, with no noticeable bottleneck.

With the lower infrastructural and administrative costs of Ethernet, and more knowledge workers in the network industry, it is evident that 10GbE FCoE and iSCSI can offer performance comparable to FC at a much lower TCO. With FCoE or iSCSI, network and storage data traffic can be consolidated using a single network switch. This consolidation has the following advantages:

- Reduces the number of network interface cards required to connect to storage and IP networks
- Reduces the number of cables and switches required to connect to storage and network devices
- Lowers energy and cooling costs by eliminating switches and multiple adapters

For more information

- http://www.redbooks.**ibm.com**/abstracts/redp4493.html?Open
- *http://netapp.com/us/communities/tech-ontap/tot-fcoe.html*
- *http://netapp.com/us/products/protocols/*
- *http://media.netapp.com/documents/tr-3719.pdf*
- *http://ietf.org/proceedings/49/I-D/draft-ietf-ips-iscsi-boot-00.txt*
- *http://msdn.microsoft.com/en-us/library/ee410012.aspx*

About the Authors

Vinay Kulkarni is an IBM System x performance engineer working onsite at the Microsoft Redmond campus. He has been working with Microsoft over the past eight years to optimize performance of System x servers running Microsoft Windows and SQL Server software. He also works with IBM clients to tune the performance of System x and storage environments and supports publication of industry-leading benchmarks alongside System x benchmark teams.

Kent Swalin is an IBM System x systems engineer working at the IBM Center for Microsoft Technologies and on the Microsoft Redmond campus. He has been working with Microsoft over the past 13 years to optimize the performance of IBM's x86-based servers running Microsoft Windows Server operating systems and Microsoft application software. He has extensive experience working with IBM customers to understand the issues they are facing and develop solutions that address them.

Wei Liu is a Microsoft Alliance engineer working at the NetApp office in Bellevue, Washington, and on the Microsoft Redmond campus. Prior to NetApp, he worked with Microsoft to optimize the performance of Exchange Server and Exchange Server backup and restore process. Over the past 4 years, he has been working with Microsoft to showcase the benefits of using NetApp solutions for Microsoft SQL Server, Exchange Server, and other popular business applications.

Acknowledgements

The authors wish to thank Scott Hinckley and Richio Aikawa from Emulex, Richard Preston and Shawn Dutton of NetApp, and the Microsoft SQL Server Performance team for assistance in setting up and getting these configurations tested.

For More Information

IBM System x Servers	http://**ibm.com**/systems/x
IBM BladeCenter Server and options	http://**ibm.com**/systems/bladecenter
IBM Systems Director Service and Support Manager	http://**ibm.com**/support/electronic
IBM System x and BladeCenter Power Configurator	http://**ibm.com**/systems/bladecenter/powerconfig
IBM Standalone Solutions Configuration Tool	http://**ibm.com**/servers/eserver/xseries/library/configtools.html
IBM Configuration and Options Guide	http://**ibm.com**/servers/eserver/xseries/cog
IBM ServerProven Program	http://**ibm.com**/servers/eserver/serverproven/compat/us
IBM Technical Support	http://**ibm.com**/server/support
Other IBM Technical Support Resources	http://**ibm.com**/servers/eserver/techsupport.html

Legal Information

IBM Systems and Technology Group
Dept. U2SA
3039 Cornwallis Road
Research Triangle Park, NC 27709

Produced in the USA
April 2010

For a copy of applicable product warranties, write to: Warranty Information, P.O. Box 12195, RTP, NC 27709, Attn: Dept. JDJA/B203. IBM makes no representation or warranty regarding third-party products or services including those designated as ServerProven or ClusterProven. Telephone support may be subject to additional charges. For onsite labor, IBM will attempt to diagnose and resolve the problem remotely before sending a technician.

IBM, the IBM logo, ibm.com, ServeRAID, System Storage, System x, and X-Architecture are trademarks of IBM Corporation in the United States and/or other countries. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. For a list of additional IBM trademarks, please see http://**ibm.com**/legal/copytrade.shtml.

Intel, the Intel logo, and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds.

Microsoft, Windows, Windows Server, the Windows logo, Hyper-V, and SQL Server are trademarks or registered trademarks of Microsoft Corporation.

Other company, product and service names may be trademarks or service marks of others.

IBM reserves the right to change specifications or other product information without notice. References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. IBM PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS

OR IMPLIED, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This publication may contain links to third party sites that are not under the control of or maintained by IBM. Access to any such third party site is at the user's own risk and IBM is not responsible for the accuracy or reliability of any information, data, opinions, advice or statements made on these sites. IBM provides these links merely as a convenience and the inclusion of such links does not imply an endorsement.

Information in this presentation concerning non-IBM products was obtained from the suppliers of these products, published announcement material or other publicly available sources. IBM has not tested these products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

MB, GB and TB = 1,000,000, 1,000,000,000 and 1,000,000,000,000 bytes, respectively, when referring to storage capacity. Accessible capacity is less; up to 3GB is used in service partition. Actual storage capacity will vary based upon many factors and may be less than stated.

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will depend on considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

Maximum internal hard disk and memory capacities may require the replacement of any standard hard drives and/or memory and the population of all hard disk bays and memory slots with the largest currently supported drives available. When referring to variable speed CD-ROMs, CD-Rs, CD-RWs and DVDs, actual playback speed will vary and is often less than the maximum possible.