Technical Report

Evaluating the Performance of FCoE, iSCSI, and FC Using DSS Workloads with Microsoft SQL Server 2008

Wei Liu, NetApp
Kent Swalin, Vinay Kulkarni, IBM
Scott Hinckley, Emulex
June 2010 | TR-3853

ABSTRACT

This technical report describes a series of tests performed by NetApp and IBM with support from Emulex to compare the performance of different storage protocols (FCoE, iSCSI, and Fibre Channel) using Decision Support System (DSS) workloads with Microsoft® SQL Server® 2008 on IBM x3850 X5 server, NetApp® FAS3070 storage systems, and Emulex adapters.

Microsoft GOLD CERTIFIED Partner

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

This paper compares the performance of three different storage protocols running DSS workloads with Microsoft SQL Server 2008:

- Fibre Channel over Ethernet (FCoE)
- iSCSI
- Fibre Channel (FC)

The test results show:

- All three protocols performed comparably, within 10% of each other.
- FCoE displayed the best overall performance.
- iSCSI also showed impressive performance, within 10% of that of FCoE.

The NetApp FAS3070 storage systems selected for the DSS workload testing against SQL Server 2008 provided adequate performance for this workload type. NetApp Snapshot™ technology was used to streamline the tests and significantly shorten the time to return the database to its initial state when switching the different storage protocols.

The IBM x3850 X5 server demonstrated that it embodies the processing power, memory, and I/O bandwidth performance needed to support intensive DSS workloads with SQL Server 2008.

2 INTRODUCTION

The two common storage protocols, Fibre Channel and iSCSI, have been used extensively in the data center. FC is the prevalent technology standard in the storage area network (SAN) data center environment, whereas iSCSI solutions have primarily been used for smaller and midsized 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 in which high data throughput is a requirement.

Another new protocol, Fibre Channel over Ethernet, has emerged. FCoE is not meant to replace FC. Rather, FCoE unifies the Fibre Channel protocol with 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.

To compare the performance of these three protocols, decision support workload scenarios were chosen because these workloads are frequently very I/O and bandwidth intensive. IBM, NetApp, and Emulex wanted to push the bandwidth limit of the interconnects between the storage and the server to see how they compare under intensive loads.

3 TEST ENVIRONMENT

The test environment consists of an IBM system 3850 X5 server and eight NetApp FAS3070 storage controllers. For FC protocol testing, four dual-port Emulex LPe12002-M8 adapters running at 4Gbps and a pair of Brocade 200E 4Gbps FC switches were used for the storage network. For FCoE and iSCSI protocols testing, two dual-port Emulex OneConnect OCe10102 10GbE CNAs and a Cisco Nexus 5010 switch were used to establish the converged network.

Figure 1 illustrates the topology for the 10Gbps FCoE and iSCSI and the 4Gbps FC test runs. The server and storage remained the same for all tests; the storage protocol and network were changed for each test.

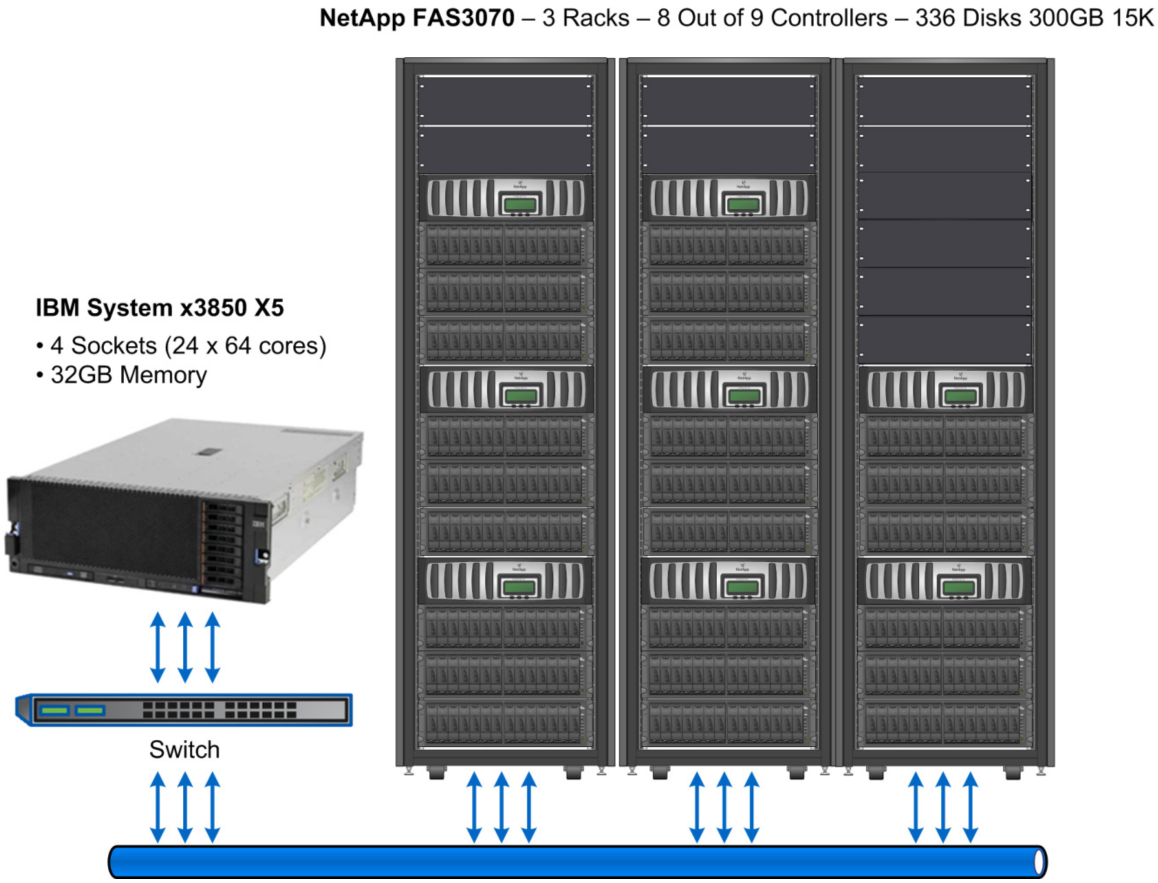

Figure 1) Topology for the FCoE, iSCSI, and FC tests.

3.1 IBM SYSTEM 3850 X5 SERVER

The System x3850 X5 Enterprise server offers:

- Up to 64 logical processors
- Up to 1TB of memory with 64 16GB DIMMs
- 7 Gen2 PCI-E slots
- Up to 8 internal drives

Powered by Intel® Xeon® processors with up to eight cores, the System x3850 X5 is designed to provide superior I/O for high-performance environments. 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 IBM server used in these tests was running Windows® Server 2008 R2 Enterprise Edition and Microsoft SQL Server 2008.

Table 1) IBM System x3850 X5 server configuration.

Component	Description
Processors	4 Intel 6-core 2.66GHz processors (24 cores)
Total Physical Memory	32GB
Operating System	Microsoft Windows Server 2008 R2 Enterprise Edition
Database Server	Microsoft SQL Server 2008 64 bit

Table 2) IBM System x3850 X5 storage interconnect configuration per protocol.

Protocol	Storage Interconnect	Storage Network
FCoE	2 — Dual-Port Emulex OCe10102-F CNA	10Gb Ethernet
iSCSI	2 — Dual-Port Emulex OCe10102-I CNA	10Gb Ethernet
FC	4 — Dual-Port Emulex LPe12002 HBA	4Gb Fibre Channel

3.2 EMULEX LPe12002-M8 HBA

The Emulex LPe12000 series adapters are high-performance 8Gb Fibre Channel host bus adapters (HBAs). As members of the Emulex LightPulse™ HBA family, their highly integrated processor design minimizes onboard components to improve host performance and efficiency. Advanced error-checking features make sure block data integrity is preserved as it traverses the SAN. Emulex's firmware-based architecture enables feature and performance upgrades without costly hardware changes. The unique Service Level Interface (SLITM) allows use of a common driver across all models of Emulex HBAs and CNAs on a given OS platform. Installation and management facilities are designed to minimize server reboots and further simplify deployment.

3.3 EMULEX OCe10102 CNA

The Emulex OCe10102-F, OCe10102-I, and OCe10102-N are high-performance converged network adapters (CNAs). Emulex CNAs allow I/O for iSCSI, FC, and FCoE storage to coexist with network traffic over a common 10-Gigabit Ethernet (10GbE) infrastructure using dedicated bandwidth channels for each. As members of the Emulex OneConnect™ universal converged network adapter (UCNA) family, the OCe10102 series reduces capital expense (capex) for adapters, switches, and cables and reduces operational expense (opex) for power, cooling, and IT administration.

SLITM allows use of a common driver across all models of Emulex HBAs and CNAs on a given OS platform.

3.4 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, as well as for 10-Gigabit 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.

Table 3) Storage configuration.

Storage Entity	Quantity
NetApp FAS3070 controllers	8
300GB 15K 4Gb/sec FC disks	336 (42 per controller)
Aggregates	16 (2 per controller)
FlexVol volumes	16 (2 per controller, 1 per aggregate)
LUNs	32 (4 per controller, 2 per aggregate, 2 per volume)
4Gb/sec FC links	8 (1 per controller)
10GbE FCoE or iSCSI links	8 (1 per controller)

Figure 2 shows the detailed configuration of a FAS3070 storage controller. All eight controllers were configured identically. Four 250GB LUNs were mapped to the server via Port 2a for the FCoE and iSCSI testing. The same four LUNs were mapped to the server via Port 0a (not shown) for the FC testing.

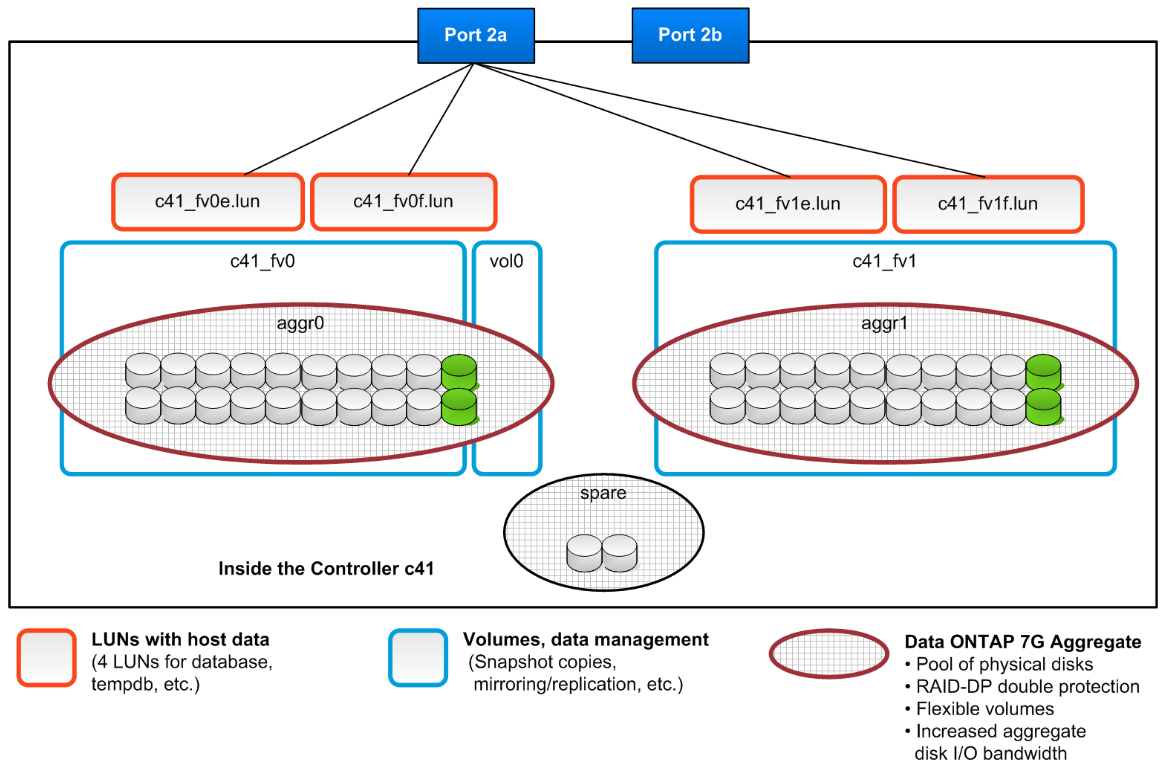

Figure 2) FAS3070 controller configuration.

Each aggregate with flexible volumes has the following characteristics:

- IOPS/disks are available to all flexible volumes.
- Volumes are logical and flexible, not constrained by hardware.
- Each volume can be sized according to capacity requirements.
- Snapshot and volume management is simplified with maximum disk I/O performance.

3.5 MICROSOFT SQL SERVER 2008

Microsoft SQL Server 2008 has been greatly enhanced to improve the performance of DSS workloads, meeting the mission-critical needs of large enterprise customers running data warehouses that measure in terabytes. Here are a few of the many data warehousing performance and scalability enhancements introduced by SQL Server 2008.

For a complete list and more in-depth description of data warehousing improvements in SQL Server Relational Database Management System (DBMS), Integration Services (SSIS), Analysis Services (SSAS), and Reporting Services (SSRS), please refer to this technical white paper by Microsoft: www.microsoft.com/sql/techinfo/whitepapers/SQL2008IntroDW.mspx

STAR JOIN

With dimensionally modeled data warehouses, a big part of the workload typically consists of what are known as “star join” queries. These queries follow a common pattern that joins the fact table with one or several dimension tables. In addition, star join queries usually express filter conditions against the non-key columns of the dimension tables and perform an aggregation on a column of the fact table.

With SQL Server 2008, customers experience significant performance improvements for many star join queries that process a significant fraction of fact table rows. The new star join optimization uses a series of hash joins, building a hash table for each dimension table that participates. As a byproduct of building this hash table, additional information, called a “bitmap filter,” is built. Bitmap filters are pushed down into the scan on the fact table, and they effectively eliminate almost all rows that would be eliminated later by the joins. This eliminates the need to spend CPU time later copying the eliminated rows and probing the hash tables for them.

PARTITION TABLE PARALLELISM

Data warehouse applications typically collect large amounts of historical data in fact tables, which are often partitioned by date. In SQL Server 2005, queries that touch more than one partition use one thread and thus one processor core per partition. This sometimes limits the performance of queries that involve partitioned tables, especially when running on parallel shared memory multiprocessor (SMP) computers with many processor cores. Partitioned table parallelism improves the performance of parallel query plans against partitioned tables by better utilizing the processing power of the existing hardware, regardless of how many partitions a query touches.

FEW OUTER ROW

In some DSS queries, the outer side of a nested loop is a parallel scan with a filter. If the qualifying data is only a few rows and clustered, then they are picked by a single thread, and this thread must do all the work for the nested loop join, even though there are threads idle. SQL Server 2008 introduces an exchange above the outer side of a nested loop join that produces few rows, and, as a result, more evenly redistributes the rows among threads and greatly improves scalability.

COMPRESSION

Database compression is a space-saving feature that helps to compress data on disk and in memory. SQL Server 2008 offers two types of compression: row and page. Row compression compresses data within a row whereas page compression looks for additional opportunities across the rows in a page. Compression helps to reduce the amount of I/O from disk and the in-memory footprint, and thus it helps I/O-intensive workloads. Compression and decompression have a CPU overhead.

BACKUP COMPRESSION

Backup compression is a feature added in SQL Server 2008 that saves both time and space in backups. With this feature, the backup stream is compressed before it is written out to the destination. Compression results highly depend on the data being compressed, but testing on typical customer databases has shown significant savings in space. Since SQL backup is typically I/O bound, the reduction in I/O actually saves time as well. Creation of compressed backups is a feature of the Enterprise Edition SKU, although any SKU can restore a compressed backup.

3.6 DSS WORKLOAD DEFINITION

The type of DSS workload tested is representative of databases found in many customer environments and is designed for tracking sales, customer, supply-chain, and product-lifecycle trends. The central charter of a DSS database is to help organizations increase profitability by analyzing trends and correlations over long periods of time. The decision support database was chosen because much of its I/O activity is uncacheable due to the large tables and the greater percentage of full table scans, and thus presents far more bandwidth demand on the I/O subsystem than an OLTP database system.

The test database was fully normalized and fully indexed on primary and foreign keys. The size of the database, including tables, indexes, and backup, is 2.7TB on disk. The databases, tempdb, and backup files created by SQL Server were evenly divided among the 32 LUNs mounted to the IBM server. The transaction logs were placed on one of the 32 LUNs, since the log files were not used heavily, which is because most of the queries involve reads. As a result, the DSS database was evenly distributed across all eight controllers, available spindles, and interconnects.

4 TEST PROCEDURE

The tests using the 4Gbps FC protocol were run first to establish a baseline. The high I/O requirements of this test were met using:

- NetApp FAS3070 storage systems
- IBM x3850 X5 server:
  - Using four dual-port Emulex LPe12002-M8 adapters running at 4Gbps via a pair of Brocade 200E 4Gbps FC switches
  - Connected with eight 4Gbps FC interconnects for a theoretical maximum bandwidth of 32Gbps, or ~3200MB/sec

For the tests using 10Gbps FCoE and iSCSI, the high I/O requirements were met using:

- NetApp FAS3070 storage systems
- IBM x3850 X5 server:
  - Using two dual-port Emulex OneConnect OCe10102 10GbE adapters via a Cisco Nexus 5010 switch
  - Connected with three 10GbE interconnects to the Cisco switch, while eight 10GbE links were used to connect the switch to the eight storage systems for a theoretical maximum bandwidth of 30Gbps, or ~3000MB/sec

The FCoE configuration was quite involved; the details can be found in “FCoE Setup” on page 14. For the 10GbE iSCSI testing, the iSCSI off-load feature of the OCe10102-I adapters was used. The setup was straightforward; see “iSCSI Setup” on page 23 for details.

The systems tested in this document ran identical DSS workloads against the SQL Server databases using the three different protocols: FCoE, iSCSI, and FC. 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.

After the testbed with each of the three protocols was configured, a quick two-minute IOMeter test was run to calibrate the testbed and validate that the configurations were capable of generating near-wire-speed throughput for each protocol under test. The observed total bandwidths were:

- FC: 3013MB/sec (8 links x 377MB/sec.)
- FCoE: 3261MB/sec (3 links x 1087MB/sec.)
- iSCSI: 2977MB/sec (3 links x 992MB/sec.)

The IOMeter test results for FCoE and iSCSI were in agreement with similar tests conducted independently by Emulex in its labs[1]. Practically, the total bandwidth was kept as close as possible for all tests using these three protocols.

For the series of tests described in this document, NetApp Snapshot technology was used to streamline and speed up the testing process. Figure 3 shows the test procedures. After the initial database creation, NetApp Snapshot copies were taken on all eight storage systems. After the FC and FCoE tests, the Snapshot copies were restored and the databases were returned to their initial state, saving several hours needed for recreating databases and ensuring that all tests were conducted in a fair manner. In all cases, the SnapRestore® operation took less than a minute.

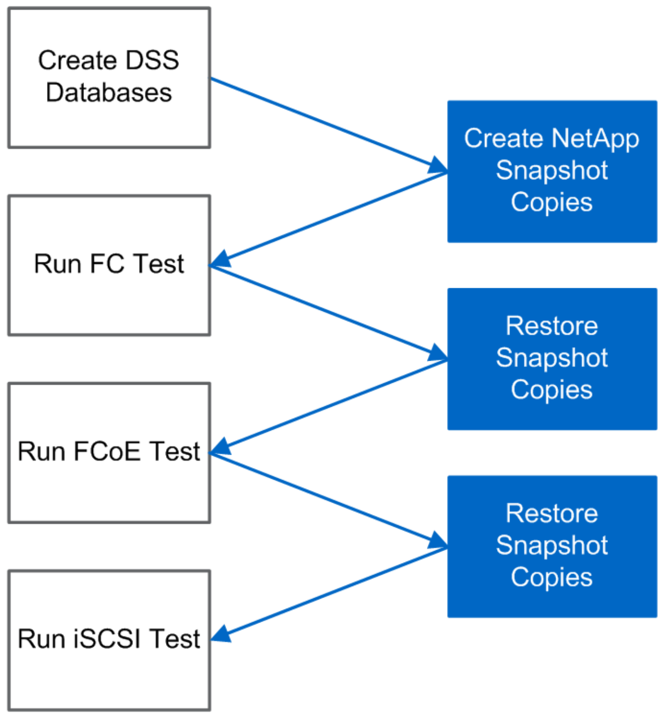

Figure 3) Test procedures and the use of NetApp Snapshot technology.

---

[1] The fact that the measured FCoE bandwidth exceeds that of FC reminds us that it is likely much more difficult to linearly scale up to eight ports than to two or three ports.

5 DSS TESTING QUERIES

A number of typical decision support queries were run to evaluate the performance of the system under stress. These queries were chosen for their complexity and diversity in terms of data access patterns and query parameters and because they access a large proportion of the available data. All queries were run as separate job requests. The set of queries is described in Table 4.

Table 4) DSS testing queries.

Query	Description	Stress Characteristics
Profit Enhancement (PE)	Provides the total increase in profits if certain discounts were not offered on products sold during a specified time period; the PE query assists in determining future product discounts	• I/O intensive • Intensive indexing scan required
Market Share Movement (MS)	Calculates the market share movement for a part in a particular nation in two years	• Complex query • Query plan and optimization critical • Hash joins critical
Shipping Mode (SM)	Determines whether using cheaper shipping modes means delaying delivery of high-priority orders until after the committed date	• Hash joins are critical • Intensive index scans
Late Shipping Suppliers (LS)	Lists suppliers who ship late	• I/O intensive • Intensive scan indexing required

6 RESULTS AND ANALYSIS

This section shows the test results in terms of the following:

- Query execution times
- Read MB/sec per query
- Server CPU consumption per query

6.1 QUERY EXECUTION TIME

Figure 4 shows the relative query execution times for the four typical DSS queries listed in the previous section. Each query's execution times using FCoE and iSCSI were normalized and compared against the execution times using the FC protocol baseline (shown in blue at 100%). In other words, the query execution times using the FC protocol are shown as 100% (baseline) in Figure 4. A query execution time lower than the baseline indicates a faster execution time and better performance.

Overall, FCoE produced the best performance and the lowest query execution times, followed by iSCSI, and then FC. On average, the three protocols performed comparably, with FCoE being 8% faster than FC and iSCSI 2% faster than FC.

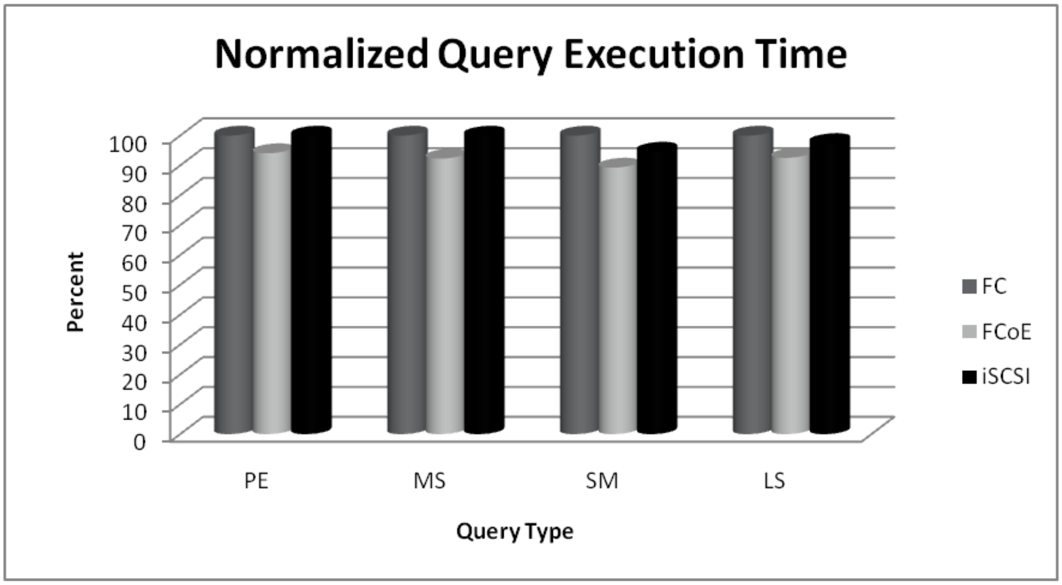

Figure 4) Normalized query execution time (lower is better).

6.2 QUERY CPU CONSUMPTION

Figure 5 shows the relative query server CPU consumption, which is computed by multiplying query execution time and average server processor utilization for each query and then normalized using the FC protocol baseline CPU consumption (shown in blue at 100%).

The FCoE CPU consumption was 8% lower than that of FC. The reduction in processor resources spent on query execution corresponds nicely with the better performance of FCoE. The iSCSI CPU consumption was 13% higher than that of FC (similar to the findings described in TR-3476).

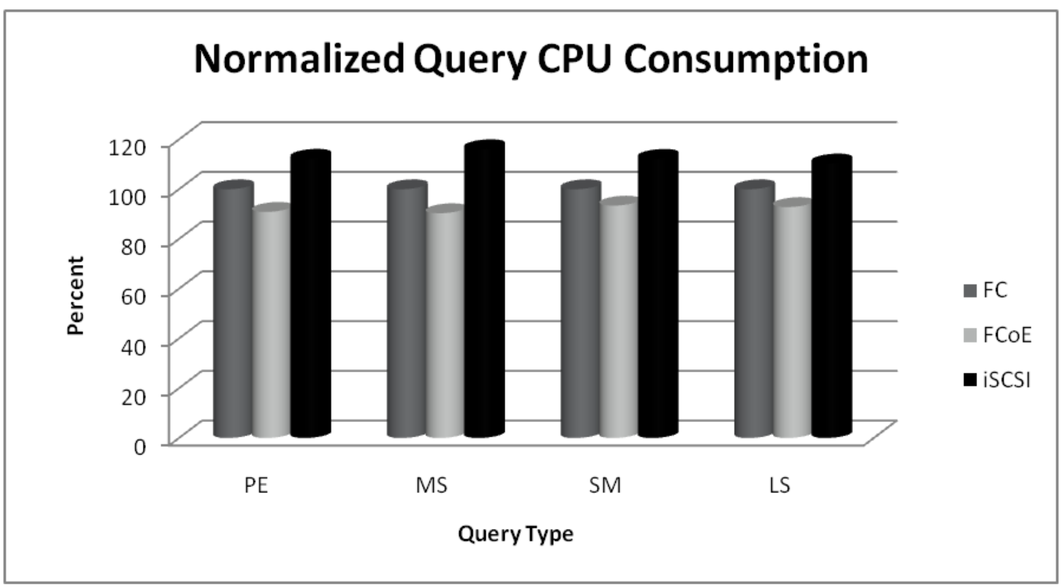

Figure 5) Normalized total query CPU consumption (lower is better).

6.3 QUERY THROUGHPUT

Figure 6 shows the average read throughput per query. FCoE displayed the highest average read throughput, followed by iSCSI, and FC the third. This is in agreement with the query execution time as shown in Figure 4.

Overall, the three protocols performed comparably, with FCoE's average read throughput being 9% higher than FC and iSCSI's being 2% higher than FC.

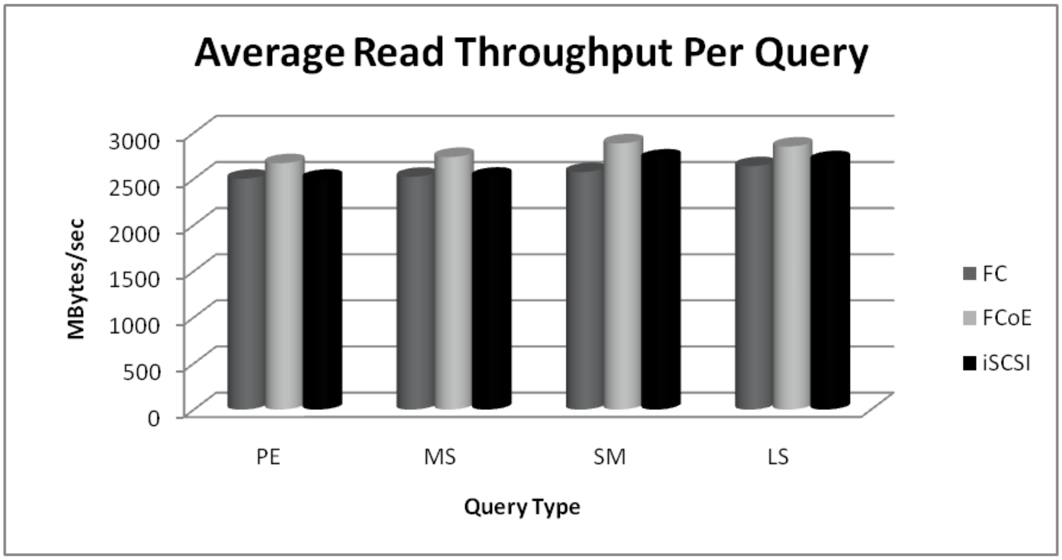

Figure 6) Average throughput (higher is better).

Figure 7 shows the peak read throughput per query. FCoE reached wire-speed throughput, followed closely by iSCSI and FC.

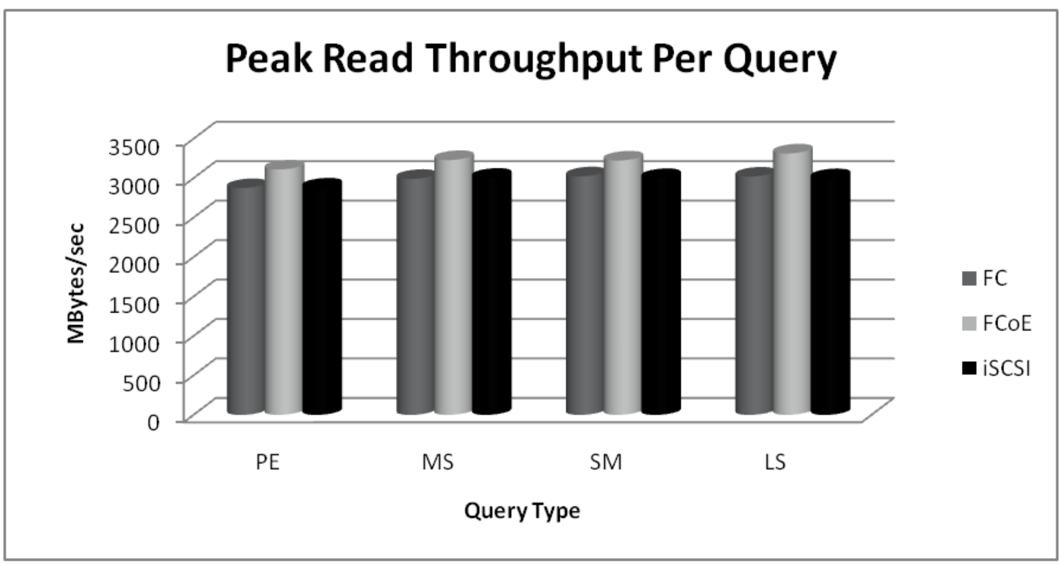

Figure 7) Peak throughput (higher is better).

7 APPENDICES

7.1 FCOE SETUP

CABLING

Figure 8 illustrates the cabling of the physical connections of this particular FCoE testbed for Microsoft SQL Server 2008 testing. The server was connected to the Cisco Nexus switch with Emulex OCe10102 CNAs, and FAS3070s were connected to the switch using NetApp 10GbE Unified Target Adapters (UTA).

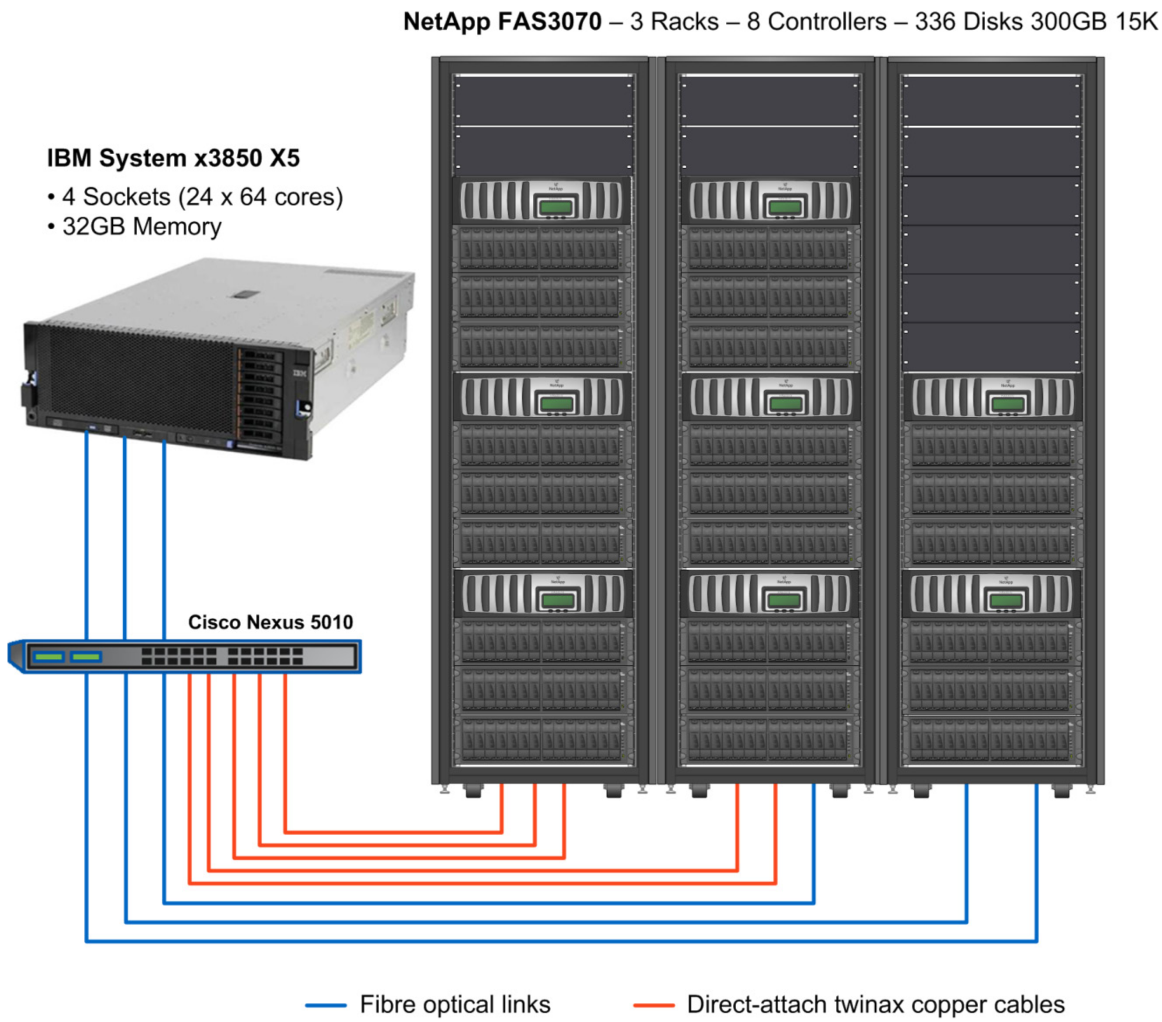

Figure 8) Server and storage are connected via the Cisco Nexus 5010 switch.

SWITCH CONFIGURATION

This section provides a comprehensive tutorial of the Cisco Nexus 5000 switch configuration for FCoE connectivity. It is not exclusive; however, it contains step-by-step instructions to get the end-to-end FCoE configuration to work.

Figure 9 shows logical relationships among the switch entities: Port, Ethernet interface, VFC, VSAN, VLAN, Alias, Zone, and Zoneset.

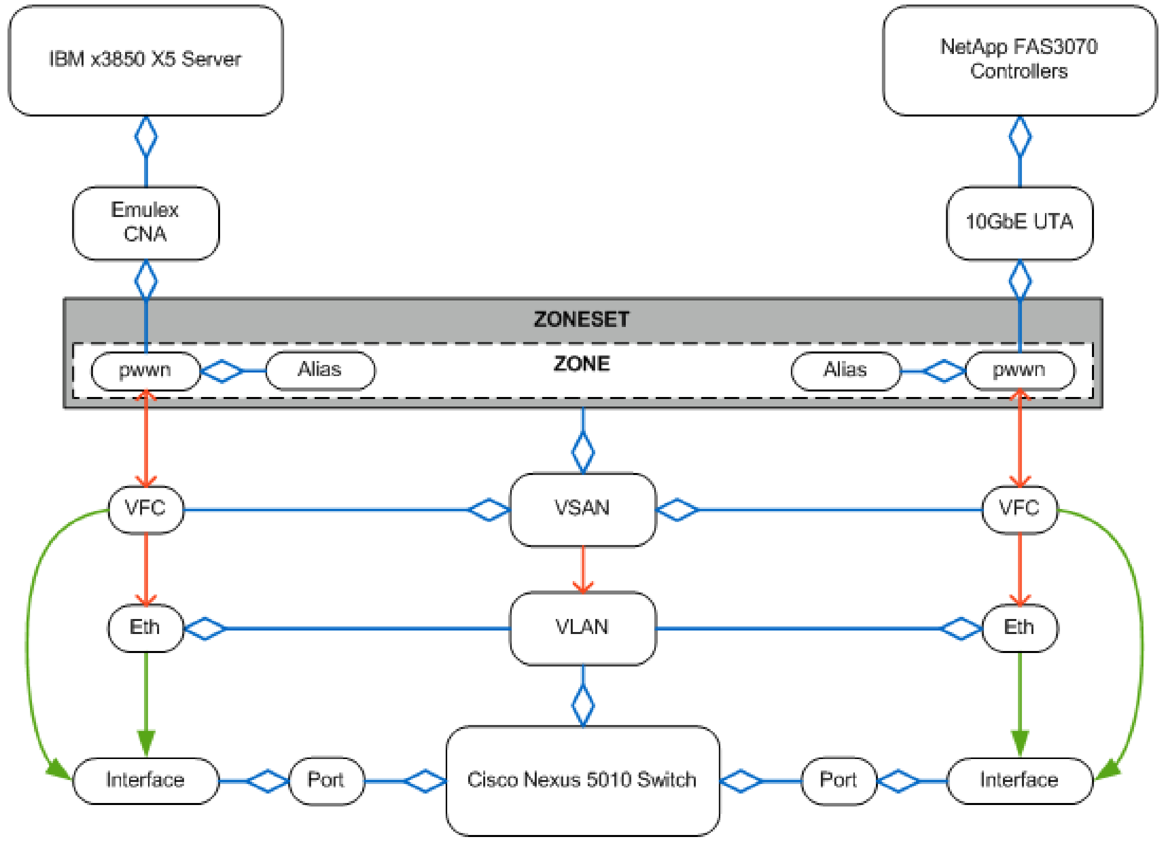

Figure 9) Logical relationships among the entities of the Cisco Nexus 5010 switch.

The switch console was used to carry out all of the configuration steps. As shown in Figure 10, there are nine major steps. Figure 10 also shows the order of these steps; although the order is not strict, it should be followed to streamline the configuration process.

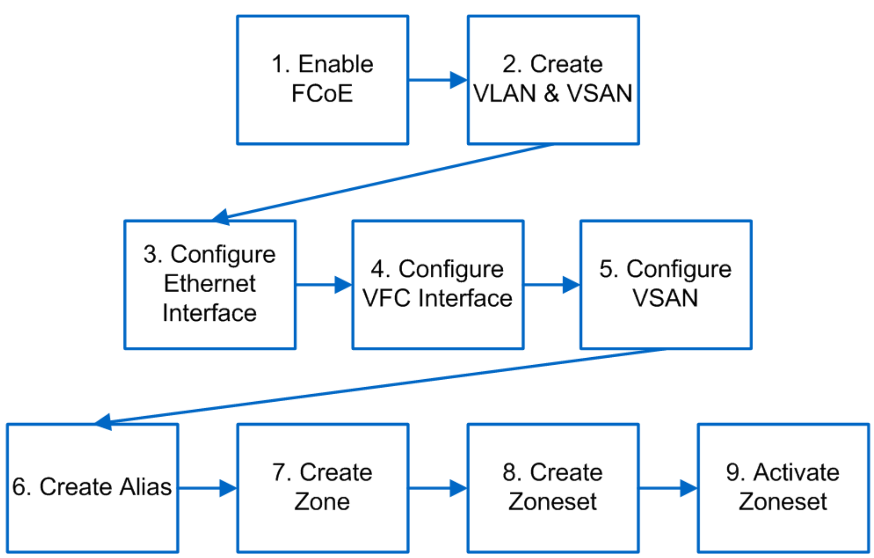

Figure 10) Switch configuration.

Note: For each step, the commands used are shown in the text box (in bold face), and the outcome or the verification of the commands is also included in the same box (highlighted with yellow).

1. Enable FCoE.

   Enable FCoE using the `feature fcoe` command. To verify that FCoE is enabled, use the `show fcoe` command.

```
sisbu-fcoe-5k-1# conf t
Enter configuration commands, one per line. End with CNTL/Z.
sisbu-fcoe-5k-1(config)# feature fcoe

sisbu-fcoe-5k-1(config)# show fcoe
Global FCF details
        FCF-MAC is 00:0d:ec:b1:3a:80
        FC-MAP is 0e:fc:00
        FCF Priority is 128
        FKA Advertisement period for FCF is 8 seconds
```

2. Create the VLAN and the VSAN.

   FCoE enables Fibre Channel and Ethernet traffic to be carried on the same physical Ethernet infrastructure. The FC traffic can be transmitted across a Virtual Fabric (VSAN) via a Virtual Fibre Channel (VFC) interface. A unique dedicated FCoE VLAN (virtual LAN) must be created to carry traffic for each VSAN, which needs to be mapped to the VLAN.

   Use these steps to create a VSAN. Use the `sh vsan` command to verify that the VSAN was created correctly.

```
sisbu-fcoe-5k-1# conf t
Enter configuration commands, one per line.  End with CNTL/Z.
sisbu-fcoe-5k-1(config)# vsan database
sisbu-fcoe-5k-1(config-vsan-db)# vsan 100
sisbu-fcoe-5k-1(config-vsan-db)# exit

sisbu-fcoe-5k-1(config)# sh vsan
vsan 1 information
         name:VSAN0001  state:active
         interoperability mode:default
         loadbalancing:src-id/dst-id/oxid
         operational state:down

vsan 100 information
         name:VSAN0100  state:active
         interoperability mode:default
         loadbalancing:src-id/dst-id/oxid
         operational state:down
...
```

   The following steps create a VLAN. The `fcoe vsan 100` command enables the VLAN for FCoE and maps VSAN 100 to VLAN 100. The `sh vlan fcoe` command verifies that the VLAN 100 was created and mapped to the VSAN 100 correctly.

```
SISBU-FCOE-5K-1(CONFIG)# vlan 100
sisbu-fcoe-5k-1(config-vlan)# fcoe vsan 100
sisbu-fcoe-5k-1(config-vlan)# exit

sisbu-fcoe-5k-1(config)# sh vlan fcoe
VLAN       VSAN       Status
--------   --------   --------
100        100        Operational
```

3. Configure the Ethernet interface.

   The Ethernet links that carry both LAN and SAN traffic are called Unified Fabric (UF) links. Cisco Nexus 5000 best practices for directly connected CNAs state that UF links must be configured as **trunk ports** and **spanning-tree edge ports**, and that all FCoE VLANs must be configured as members of the UF links. Therefore, the Ethernet interface (links or ports) carrying FCoE traffic should be configured as shown here.

   The `switchport trunk allowed vlan 1, 100` command associates VLAN 1 (default) and VLAN 100 (for FCoE) to the Ethernet port `eth 1/1`. To verify the state of the Ethernet interface, use the command `sh int eth 1/1` and then use the command `sh int eth 1/1 switchport` to verify its configuration.

```
sisbu-fcoe-5k-1(config)# int eth 1/1
sisbu-fcoe-5k-1(config-if)# spanning-tree port type edge trunk
sisbu-fcoe-5k-1(config-if)# switchport mode trunk
sisbu-fcoe-5k-1(config-if)# switchport trunk allowed vlan 1, 100
Warning: Edge port type (portfast) should only be enabled on ports
 connected to a single host. Connecting hubs, concentrators, switches,
 bridges, etc... to this interface  when edge port type (portfast) is
 enabled, can cause temporary bridging loops.
 Use with CAUTION

sisbu-fcoe-5k-1(config-if)# sh int eth 1/1
Ethernet1/1 is up
  Hardware: 1000/10000 Ethernet, address: 000d.ecb1.3a88 (bia
000d.ecb1.3a88)
  MTU 1500 bytes, BW 10000000 Kbit, DLY 10 usec,
     reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation ARPA
  Port mode is trunk
  full-duplex, 10 Gb/s, media type is 10g
...

sisbu-fcoe-5k-1(config-if)# sh int eth 1/1 switchport
Name: Ethernet1/1
  Switchport: Enabled
  Switchport Monitor: Not enabled
  Operational Mode: trunk
  Access Mode VLAN: 1 (default)
  Trunking Native Mode VLAN: 1 (default)
  Trunking VLANs Enabled: 1,100
...

sisbu-fcoe-5k-1(config-if)# exit
```

   The Ethernet interface configuration steps outlined here should be repeated for all Ethernet ports (eth 1/1 to eth 1/11 in this case) associated with the FCoE VSAN.

4. Configure VFC Interface.

   In order to carry FCoE traffic, create VFCs and bind VFCs to physical Ethernet interfaces that are configured for carrying FCoE traffic, as shown here.

   a. The `int vfc5` command creates the VFC interface called vfc5.
   b. The `no shutdown` command enables the VFC interface.
   c. The `bind int eth 1/5` command binds vfc5 to Ethernet 1/5.

```
sisbu-fcoe-5k-1(config)# int vfc5
sisbu-fcoe-5k-1(config-if)# no shutdown
sisbu-fcoe-5k-1(config-if)# bind int eth 1/5
sisbu-fcoe-5k-1(config-if)# sh int vfc5
vfc5 is down
    Bound interface is Ethernet1/5
    FCF priority is 128
    Hardware is Virtual Fibre Channel
    Port WWN is 20:04:00:0d:ec:b1:3a:bf
    Admin port mode is F, trunk mode is on
    snmp link state traps are enabled
    Port vsan is 1
...
```

   The VFC interface configuration steps outlined here should be repeated for all Ethernet ports (eth 1/1 to eth 1/11) associated with the FCoE VSAN, one VFC per Ethernet port.

   To view the state of vfc5, use the `sh int vfc5` command. As shown here, vfc5 is bound to Ethernet 1/5. You will notice that it is still down because it has not joined the VSAN yet.

5. Configure the VSAN.

   All the VFCs created in Step 4 need to join VSAN 100, the FCoE VSAN created for carrying FCoE traffic in Step 2.

   Use the `vsan database` command to enter the VSAN configuration mode. The next command, `vsan 100 interface vfc5`, adds (or associates) vfc5 to VSAN 100. Use the `sh int vfc5` command to verify the state of vfc5. This time, vfc5 is up. It is bound to Ethernet 1/5, and it belongs to VSAN 100.

```
sisbu-fcoe-5k-1(config-if)# vsan database
sisbu-fcoe-5k-1(config-vsan-db)# vsan 100 interface vfc5

sisbu-fcoe-5k-1(config-vsan-db)# sh int vfc5
vfc5 is up
    Bound interface is Ethernet1/5
    FCF priority is 128
    Hardware is Virtual Fibre Channel
    Port WWN is 20:04:00:0d:ec:b1:3a:bf
    Admin port mode is F, trunk mode is on
    snmp link state traps are enabled
    Port mode is F, FCID is 0x990004
    Port vsan is 100
...
```

   The fabric login (FLOGI) database and name server (FCNS) database are two of many features provided in Cisco Nexus 5000 switches. The `sh flogi database vsan 100` command shows storage devices logged in to VSAN 100, and the `sh fcns database vsan 100` command displays information about VSAN 100. This example shows the information about VSAN 100 after five VFCs (vfc1-5) joined VSAN 100.

```
sisbu-fcoe-5k-1(config-vsan-db)# sh flogi database vsan 100
--------------------------------------------------------------------------------
INTERFACE        VSAN    FCID           PORT NAME               NODE NAME
--------------------------------------------------------------------------------
vfc1             100    0x990000  10:00:00:00:c9:93:f9:97 20:00:00:00:c9:93:f9:97
vfc2             100    0x990001  10:00:00:00:c9:93:f8:f5 20:00:00:00:c9:93:f8:f5
vfc3             100    0x990002  10:00:00:00:c9:93:f8:f7 20:00:00:00:c9:93:f8:f7
vfc4             100    0x990003  50:0a:09:85:87:59:54:17 50:0a:09:80:87:59:54:17
vfc5             100    0x990004  50:0a:09:85:87:19:54:4d 50:0a:09:80:87:19:54:4d

Total number of flogi = 5.

sisbu-fcoe-5k-1(config-vsan-db)# sh fcns database vsan 100

VSAN 100:
--------------------------------------------------------------------------
FCID        TYPE  PWWN                    (VENDOR)        FC4-TYPE:FEATURE
--------------------------------------------------------------------------
0x990000    N     10:00:00:00:c9:93:f9:97 (Emulex)        ipfc scsi-fcp:init
0x990001    N     10:00:00:00:c9:93:f8:f5 (Emulex)        ipfc scsi-fcp:init
0x990002    N     10:00:00:00:c9:93:f8:f7 (Emulex)        ipfc scsi-fcp:init
0x990003    N     50:0a:09:85:87:59:54:17 (NetApp)        scsi-fcp:target
0x990004    N     50:0a:09:85:87:19:54:4d (NetApp)        scsi-fcp:target

Total number of entries = 5

sisbu-fcoe-5k-1(config-vsan-db)# exit
```

The Port Name reflects the connected CNA port's PWWN (Port World Wide Name). In this example, vfc1, vfc2, and vfc3 are bound to eth 1/1, eth1/2, and eth 1/3, respectively. And switch ports eth 1/1, eth 1/2, and eth 1/3 are connected to three ports of two dual-port Emulex CNAs, which reside in the IBM x3850 X5 server. By the same token, vfc4 and vfc5 are associated with two NetApp FAS3170 storage controllers.

At this point, the server's Emulex CNA FCoE interfaces are up, but they cannot access target devices yet (Figure 11).

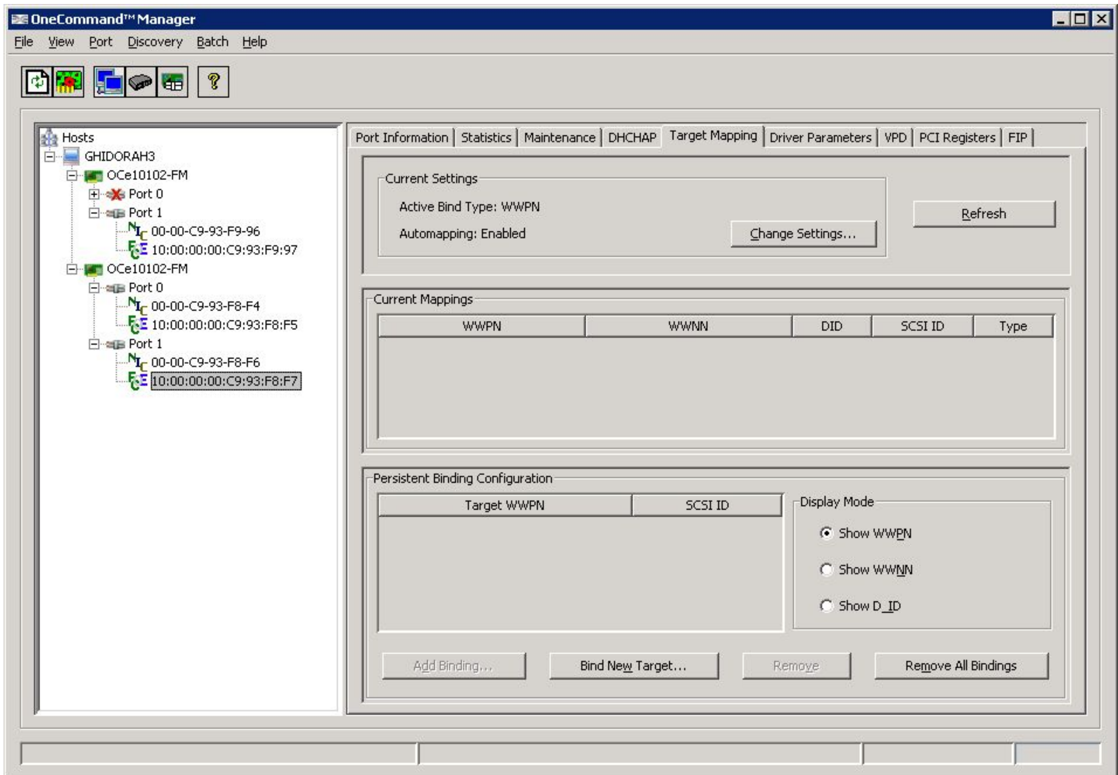

Figure 11) OneCommand GUI shows Emulex CNAs are up but without targets.

Zoning is required to enable access control among storage devices (see steps 6–9).

6. Create an alias.

   PWWNs are not easy to read and remember. So, **device aliases** can be created for PWWNs to make them more user friendly. This example shows how to create a device alias for the storage controller c42, whose PWWN is 50:0a:09:85:87:19:54:4d.

```
sisbu-fcoe-5k-1(config)# fcalias name c42 vsan 100
sisbu-fcoe-5k-1(config-fcalias)# mem pwwn 50:0a:09:85:87:19:54:4d

sisbu-fcoe-5k-1(config-fcalias)# sh fcalias
fcalias name c42 vsan 100
  pwwn 50:0a:09:85:87:19:54:4d

sisbu-fcoe-5k-1(config-fcalias)# exit
```

   The `sh fcalias` command verifies that the alias just created is as intended. This step is repeated for all VFCs created in Step 4. For the three Emulex CNA ports, the aliases ghidorah3_p1, ghidorah3_p2, and ghidorah3_p3 were used. For the eight controllers, the aliases c42, c41, and so on were used.

7. Create a zone.

   A zone can have multiple zone members. The members in a zone can access each other. Use the command `zone name` **`zone_ghidorah3_3070s`** `vsan 100` to create a zone named **zone_ghidorah3_3070s** for VSAN 100. The next command, `mem fcalias c41`, adds controller c41 to the zone. Note that the alias for the controller was used instead of its PWWN. This command was repeated to add all eight controllers and three Emulex CNA ports to the same zone, allowing each host port to access all eight controllers to balance the load and achieve the maximum bandwidth.

Verify the state of the zone with the `sh zone` command. This example shows the state of `zone_ghidorah3_3070s vsan 100` after three host ports and two storage controllers were added to the zone.

```
sisbu-fcoe-5k-1(config)# zone name zone_ghidorah3_3070s vsan 100
sisbu-fcoe-5k-1(config-zone)# mem fcalias c41

sisbu-fcoe-5k-1(config-zone)# sh zone
zone name zone_ghidorah3_3070s vsan 100
  fcalias name c41 vsan 100
    pwwn 50:0a:09:85:87:59:54:17

  fcalias name ghidorah3_p1 vsan 100
    pwwn 10:00:00:00:c9:93:f9:97

  fcalias name ghidorah3_p2 vsan 100
    pwwn 10:00:00:00:c9:93:f8:f5

  fcalias name ghidorah3_p3 vsan 100
    pwwn 10:00:00:00:c9:93:f8:f7

  fcalias name c42 vsan 100
    pwwn 50:0a:09:85:87:19:54:4d

sisbu-fcoe-5k-1(config-zone)# exit
```

8. Create a zoneset.

   Zones are organized into zonesets. A zoneset can have one or more zones. Use the command `zoneset name` **`zset_tpch_fcoe`** `vsan 100` to create a zoneset for VSAN 100. The next command, `mem zone_ghidorah3_3070s`, adds **zone_ghidorah3_3070s** to the zoneset just created. The `sh zoneset` command verifies the members of the zoneset.

```
sisbu-fcoe-5k-1(config)# zoneset name zset_tpch_fcoe vsan 100
sisbu-fcoe-5k-1(config-zoneset)# mem zone_ghidorah3_3070s

sisbu-fcoe-5k-1(config-zoneset)# sh zoneset
zoneset name zset_tpch_fcoe vsan 100
  zone name zone_ghidorah3_3070s vsan 100
    fcalias name c41 vsan 100
      pwwn 50:0a:09:85:87:59:54:17
...

sisbu-fcoe-5k-1(config-zoneset)# exit
```

9. Activate the zoneset.

   Each VSAN can have one or more zonesets. Only one zoneset can be activated at any given time. The final step is to activate the zoneset `zset_tpch_fcoe` as shown here.

```
sisbu-fcoe-5k-1(config)# zoneset activate name zset_tpch_fcoe vsan 100
Zoneset activation initiated. check zone status

sisbu-fcoe-5k-1(config)# copy running-config startup-config
 [########################################] 100%

sisbu-fcoe-5k-1(config)# exit
```

The `copy running-config startup-config` command makes the configuration persistent across switch reboot or power cycle.

After zoning setup and activating the zoneset, Emulex CNAs can discover the storage controllers. Figure 12 shows that, after one storage controller was added to the zone, the Emulex CNAs were able to discover the target storage device.

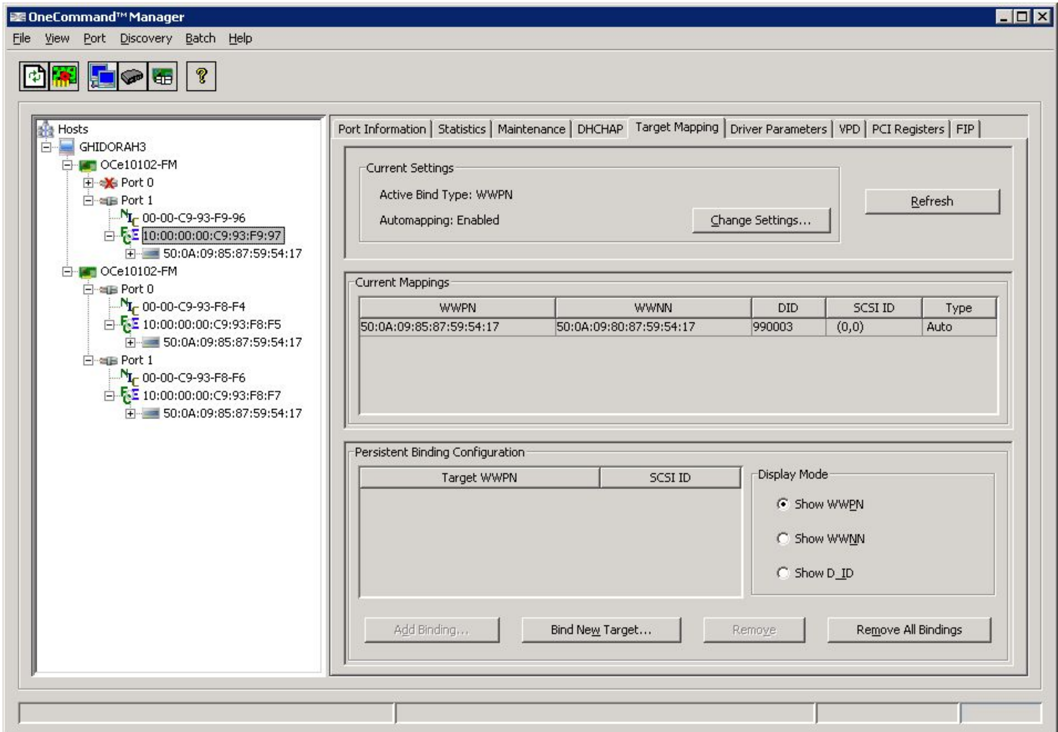

Figure 12) OneCommand GUI shows Emulex CNAs discovered storage controller c41.

FAS3070 CONFIGURATION

FCoE configuration on NetApp storage controllers is done the same way as on a standard FCP testbed. For example, on controller c41, check the initiators using `fcp show initiator` via c41's telnet interface.

```
netapp-e4-2*> fcp show initiator
...
Initiators connected on adapter 2a:
Portname                  Group
--------                  -----
10:00:00:00:c9:93:f9:97
10:00:00:00:c9:93:f8:f5
10:00:00:00:c9:93:f8:f7
...
```

The three initiators from the host's Emulex CNAs are connected. At this point, the next steps are creation of an **igroup**, with these three initiators as its members, and mapping LUNs (four LUNs per controller) to the igroup. This step is repeated for each of the controllers. Then, all 8 controllers and their LUNs (32 total) are visible and accessible to the server.

7.2 ISCSI SETUP

CABLING

Cabling for iSCSI is the same as for FCoE for Microsoft SQL Server 2008 testing (see Figure 8 on page 14).

OCE10102-I CONFIGURATION

To configure iSCSI using OneCommand:

1. Set the iSCSI port IP address.
   a. In the OneCommand GUI, select the Port you want to configure.
   b. Switch to the **iSCSI Port Info** tab.
   c. Click the **Modify…** button.
   d. Enter the IP address for this initiator port.
   e. Repeat this step for all initiator ports in the tests.
2. Add the target and login.
   a. From the **iSCSI Target Discovery** tab, click the **Manually Add Target…** button to add an iSCSI target (FAS3070 controller).
   b. Enter the IP address of the controller and click **OK**.
   c. Click the **Target Login…** button to log in to the selected target.
   d. Repeat this step for all controllers.

FAS3070 CONFIGURATION

After the iSCSI connections are established, the next steps are:

1. Create an igroup with the three iSCSI initiators as its members.
2. Map the LUNs (four LUNs per controller) to the igroup. This step is repeated for each of the controllers.

All 8 controllers and their LUNs (32 total) are now visible and accessible to the server.

7.3 FC SETUP

CABLING

Cabling for 4Gbps FC was different in that two Brocade 200E switches were used. Figure 13 illustrates the cabling of the physical connections of this particular FC testbed for Microsoft SQL Server 2008 testing.

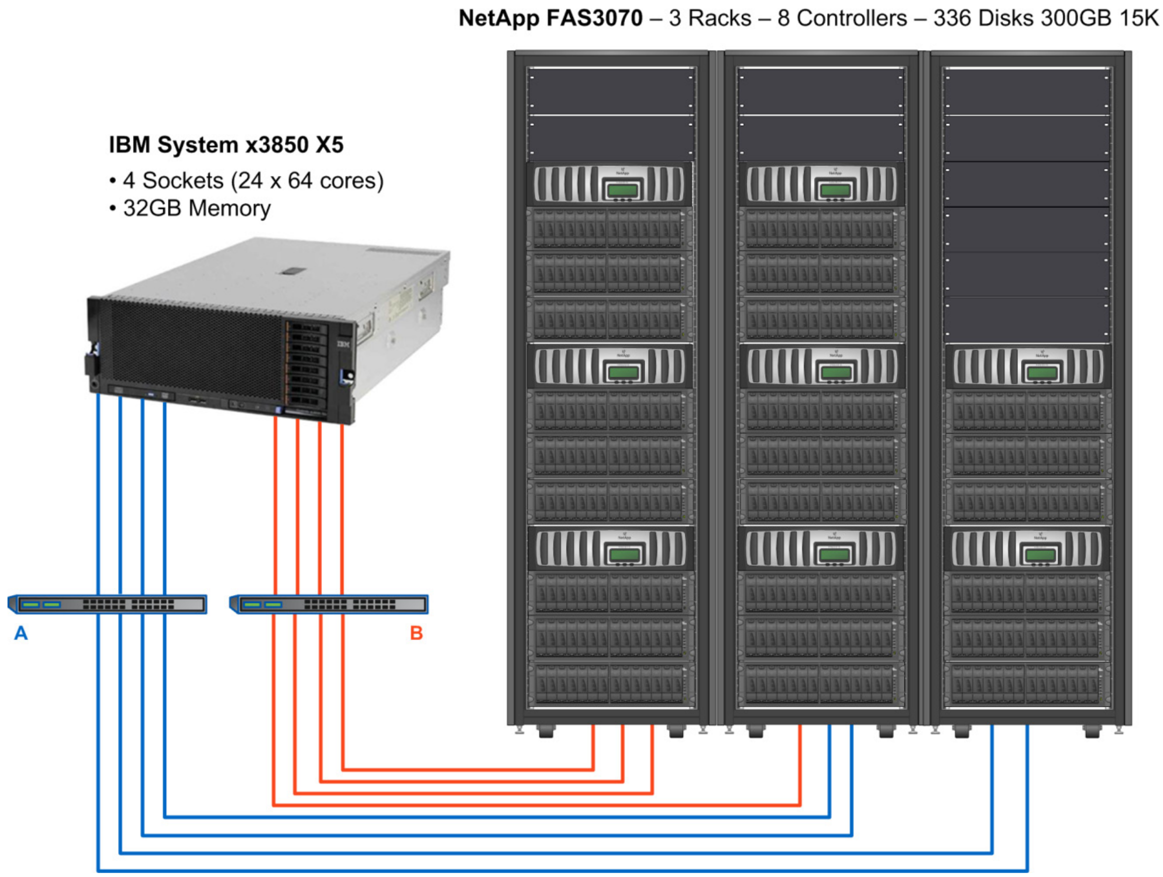

Figure 13) Pair of Brocade 200E switches connecting server and storage.

FAS3070 CONFIGURATION

On the pair of Brocade switches, no zoning was set up (open access by default). Each of the eight host ports were mapped to one of the eight controllers by creating an igroup with the particular host port’s WWN, and then by mapping the LUNs (four LUNs per controller) to the igroup. This step is repeated for each of the controllers. Then, all 8 controllers and their LUNs (32 total) are visible and accessible to the server.

7.4 SQL SERVER 2008 TUNING OPTIONS

SQL Server 2008 automatically performs most of the necessary tuning and dynamically configures its parameters based on usage and availability of system resources. Other than enabling large-page support, all these tests were run using the default tuning options.

8 REFERENCES

- An Introduction to Fibre Channel over Ethernet and Fibre Channel over Convergence Enhanced Ethernet
  www.redbooks.ibm.com/abstracts/redp4493.html?Open
- FCoE: The Future of Fibre Channel?
  http://netapp.com/us/communities/tech-ontap/tot-fcoe.html
- Protocols
  http://netapp.com/us/products/protocols/
- Microsoft SQL Server 2008: New Compression Technology Enhances Decision Support Workload Performance
  http://media.netapp.com/documents/tr-3719.pdf
- Fibre Channel and iSCSI Performance Comparison for DSS Workloads Using SQL Server 2005
  http://media.netapp.com/documents/tr-3476.pdf
- Using Star Join and Few-Outer-Row Optimizations to Improve Data Warehousing Queries
  http://msdn.microsoft.com/en-us/library/ee410012.aspx
- Emulex Drivers for Windows
  www-dl.emulex.com/support/windows/windows/230016/windows_drivers.pdf
- Cisco Nexus 5000 Series NX-OS Fibre Channel over Ethernet Configuration Guide
  www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/fcoe/b_Cisco_Nexus_5000_Series_NX-OS_Fibre_Channel_over_Ethernet_Configuration_Guide_.html
- Cisco Nexus 5000 Series Switch CLI Software Configuration Guide
  www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/configuration/guide/cli/CLIConfigurationGuide.html
- A Standard for Bootstrapping Clients Using the iSCSI Protocol
  http://ietf.org/proceedings/49/I-D/draft-ietf-ips-iscsi-boot-00.txt

9 ACKNOWLEDGEMENTS

The authors want to thank Richio Aikawa of Emulex; Shawn Dutton, Richard Preston, and Chris Lemmons of NetApp; and the Microsoft SQL Server Performance team for assistance in getting this project completed.