Enterprise Storage

HITACHI
DATA SYSTEMS

Using Hitachi Dynamic Link Manager™ Software for a Highly Available Microsoft® Exchange 2000 Configuration

A Performance Brief

by Joe Carlisle

December 2002

Application Integration Lab
Technical White Paper Series

Executive Summary

In supporting the Hitachi TrueNorth™ vision, Hitachi Data Systems strives to provide not only high-quality, intelligent storage and an open, standards-based storage management framework, but also a collaborative business model that leverages the strengths of our partners.

Hitachi Data Systems enacts the TrueNorth vision by embracing the Microsoft® Windows® 2000 platform, an increasingly vital component to many enterprise IT efforts. The Hitachi Data Systems Application Integration Lab recently used the LoadSim benchmark to test and measure an increasing I/O load on the Microsoft Exchange 2000 system, while using Hitachi Dynamic Link Manager™ software for a highly available Exchange 2000 configuration.

Many large and rapidly growing enterprises are now shifting mission-critical workloads from mainframe (IBM® S/390®) and large UNIX® platforms to Wintel-type servers. As Windows 2000 takes on additional mission-critical applications, it is becoming obvious that these applications will have to provide the same (or even better) quality of services (QoS) as the ones provided by the S/390 world—including high availability and data protection.

The fact that Microsoft has renamed its top-of-the-line Windows 2000 server software "Windows 2000 Data Center" signals the desire and commitment of Microsoft to play a major role in large data centers. Most Windows 2000-based mission-critical applications are now using the Microsoft Cluster Server (MSCS) to provide the nonstop availability required by users. Major applications and relational database management systems (RDBMSs) using MSCS include:

- Microsoft Exchange 2000
- Microsoft SQL Server™ 2000
- Oracle9*i*™
- SAP® R/3®

Data protection against various types of failure is also key. The end user must be protected against hardware/media failures, logical failures, and disaster. This white paper reveals the ways in which Dynamic Link Manager software helps to address these issues for the Microsoft Exchange 2000 application. As an integral component of the Hitachi Data Systems Performance Enhancement Suite, Dynamic Link Manager software automates I/O load balancing, path failover, and recovery capabilities in the event a single path breaks down. In a unique balancing act, Dynamic Link Manager software ensures that no single path becomes congested and overworked while another is underutilized. By automatically allocating data to an alternate path, it removes the threat of application failure.

This paper also discusses how two types of Dynamic Link Manager software configurations can be implemented and identifies which of the two is a better practice for overall failover performance in a clustered environment.

Acknowledgment

The Application Integration Lab would like to acknowledge the following people for their contributions to this technical white paper:

- Tarny Hou
- Wayne Validum
- Derek Schene
- Les Boynton
- Joe Ortiz

Contents

Using Hitachi Dynamic Link Manager™ Software for a Highly Available Microsoft® Exchange 2000 Configuration

A Performance Brief

by Joe Carlisle

Introduction

Organizations today experience ever-increasing requirements for higher availability of and accessibility to their data. The Hitachi Dynamic Link Manager™ path failover and I/O balancing software provides an incremental level of higher availability for components that provide connectivity between the server and the Hitachi Freedom Storage™ system, such as host bus adapters (HBAs), HBA drivers, fabric switches, switch firmware, Gigabit Ethernet Interface Cards (GBICs), and fiber cables. The Dynamic Link Manager software program manages access to storage paths. It provides functions for distributing the load across paths and for switching to another path in the case of path failure, thus increasing availability and improving the reliability of the system. These attributes, plus the capability of Dynamic Link Manager software to leverage the strengths of our partners, support the Hitachi TrueNorth™ vision to provide high-quality, intelligent storage and an open, standards-based storage management framework within a collaborative business model.

While server-clustering technology provides high availability, products such as Dynamic Link Manager software perform an important path failover function that clustering products cannot handle. In a clustered environment, if a path should fail and there is no path failover solution available, the cluster middleware will fail over the entire server to a backup server. Server failover is a lengthy process and seriously impacts the customer's ability to access data. This paper will show how to potentially avoid server failovers by properly configuring Dynamic Link Manager software to fail over at the HBA level, when I/O connectivity is lost at the server or storage edge.

Configuring a Highly Available Microsoft® Exchange 2000 Server

Exchange 2000 is increasingly becoming a mission-critical application that requires minimal or no downtime and is an ideal candidate for highly available systems. Microsoft Cluster Server (MSCS) is designed to keep user disruption to a minimum in the event of a server failure, but even with the high availability of a clustered environment there is potential to increase availability by avoiding cluster failover altogether. This can be accomplished by using Dynamic Link Manager software to avoid server failover at the HBA level. Dynamic Link Manager software can be configured to fail one HBA to another when I/O connectivity is lost at the server or storage edge. By defining Dynamic Link Manager software in this way, a cluster server failover can be avoided.

Since the Loadsim MMB2 workload is used with Exchange 2000 in an Active/Passive Microsoft Clustered Server environment, we will first describe the Loadsim configuration. We will then illustrate Exchange Server configuration used for these measurements and describe the two Dynamic Link Manager software configurations used. Finally, we present test results, which demonstrate how a server failover can potentially be prevented when using the correct Dynamic Link Manager software configuration.

Appendix A contains a detailed description of how the Emulex® LP8000 HBAs were configured for these tests.

LoadSim 2000

LoadSim MMB2 Industry Benchmark

The measured workload for this study simulates up to 6000 users and uses the industry standard LoadSim benchmark with the MMB2 profile. MMB2 is a benchmark for a significantly increased workload that employs increased message sizes and more frequent and diverse workload actions compared to MMB, which was used with versions of Exchange prior to Exchange 2000. MMB2 more accurately reflects the increased workload seen by customers as they increasingly rely on messaging services deployed in today's heavy corporate e-mail environments.

LoadSim Client Specifications

Table 1 details the hardware and software configuration of the LoadSim client used to simulate multiple Microsoft Exchange users generating the MMB2.

LoadSim Clients	Configuration
Model	HP® (Compaq) ProLiant ML370
Client CPU types and speeds	(2) 500Mhz Intel® Pentium® III
Number of clients	10 clients with 1GB RAM (600 users for each)
Network topology (100Base T, Token Ring, etc.)	100 Base-TX
Network controllers	HP (Compaq) 10/100 TX
Client network software name and version	Microsoft Windows® 2000 Professional, SP2
LoadSim version	2000

Table 1: Configuration of LoadSim Client.

Exchange 2000 System

Server Specifications

The Exchange Server is configured in a MSCS Active/Passive environment to reflect a highly available system. The server is set up so that 6,000 users can be simulated using the LoadSim MMB2 workload. LoadSim 95th percentile response time is sub-second using this configuration.

Table 2 describes the configuration of the Exchange Server. The cluster is configured in Active/Passive mode as a two-node cluster. Allocated space on the Hitachi Freedom Storage Lightning 9900™ Series, Lightning 9960™ system for Exchange system and database files is shown.

***ProLiant* 6400R**
Four Pentium III Xeon™ 550 Processors—2MB L2 cache per processor
4GB RAM
One HP (Compaq) SMART Array 3200 Controller with 64MB cache; one SMART 4200 Array Controller Exchange Log Files: four Lightning 9960 LUNs—RAID 1+ Exchange Information Store Files: four Lightning 9960 LUNs—RAID 1+
Windows 2000 Advanced Server Build 2195 SP2—Active/Passive Cluster Exchange Server Version 2000—Enterprise Edition with Service Pack 2
Emulex HBAs with following parameters; • Emulex LP-8000 (4 per node) • Firmware: 3.82A0 • Driver: elxsli2.sys v5-2.11a2 (full port) • Configuration: elxcfg.exe v1.4.1.2
Connected into a Lightning 9960 SAN via McDATA® ED-6064 with Open-L LUNS
McDATA Directors and Switch Version—ED-6064 v 01.03.00 35

Table 2: Configuration of Exchange Server.

Note that in the Appendix there are recommended parameter settings for the Emulex LP-8000 HBA. These settings are recommended to provide optimal connectivity to Lightining 9960 storage systems.

Software Specifications

The Exchange 2000 configuration includes Dynamic Link Manager software running on Windows 2000 Advanced Server. Table 3 provides details.

Software Specifications
Dynamic Link Manager software Version 3.0.2
Operating System details; • Server OS: Windows 2000 Advanced Server • Service Pack 2 • Hotfix Q296441 • Hotfix Q300972 • Hotfix Q309304 • Hotfix Q313450 • Hotfix Q313829 • Hotfix Q314147 • Windows 2000 Security Rollup

Table 3: Software Specifications.

Lightning 9960 Storage System

Storage Specifications

The storage type is a Lightning 9960 system at microcode level 01-17-83, with 16GB of cache installed. The Exchange databases are defined as RAID-1+, OPEN-L*2 LUNs.Storage utilization for the four Exchange databases used for test measurements is depicted in Table 4.

Storage Utilization by Host	
Host	**Capacity in Use**
MSEMAIL1	539GB
Host: **MSEMAIL1**	

CL1-B/CL2-B							**Allocated Capacity in Group**	**% Used**
Home / LUN SCAN /							73GB	
Subsystem					Ports		LUN Security	
					CL1-B			
					CL2-B			
LDEV	LUN	Port	Host	Device Name	Mount Point	SCSI ID	Size	% Used
0:00	01	CL1-B	MSEMAIL1	SG1-DB1	n:\	15	72.9GB	54
0:02	02	CL1-B	MSEMAIL1	SG2-DB2	o:\	15	72.9GB	54

CL1-F/CL2-F							**Allocated Capacity in Group**	**% Used**
Home / LUN SCAN /							146GB	
Subsystem					Ports		LUN Security	
					CL1-F			
					CL2-F			
LDEV	LUN	Port	Host	Device Name	Mount Point	SCSI ID	Size	% Used
0:10	03	CL1-F	MSEMAIL1	SG3-DB3	u:\	15	72.9GB	55
0:12	04	CL1-F	MSEMAIL1	SG4-DB4	v:\	15	72.9GB	54

Table 4: Storage Utilization by Host.

HBA pathing is controlled by Dynamic Link Manager software for SG1-DB1 and SG2-DB2 on ports 1B and 2B. HBA pathing is controlled for SG3-DB3 and SG4-DB4 on ports 1F and 2F. Notice the LUN numbers for the databases. SG1-DB1 is defined on LUN 1, SG2-DB2 is defined on LUN 2, SG3-DB3 is defined on LUN 3, and SG4-DB4 is defined on LUN 4. The output of the Dynamic Link Manager software VIEW PATH command displays statistics by LUN number, which will be shown in Tables 5 and 6 in the Dynamic Link Manager Software Connectivity section.

Introduction to Hitachi Dynamic Link Manager Software

Dynamic Link Manager Software Configuration Basics

In a large-scale storage system, fibre cables or SCSI cables connect a host and storage. The cable port on the host is a host bus adapter (HBA). The cable port on storage is a port on a disk controller. A logical unit (LU) contained in storage is the target of input/output to and from the host. The LU can also be divided into devices. The connection path between the host and device is simply called a path. In the description of the operation of, and procedures for, Dynamic Link Manager software, the connection path between the host and LU is called a *physical pat*h. Dynamic Link Manager software manages the paths.

This performance brief demonstrates how Dynamic Link Manager software can be configured to provide HBA failover for a clustered Exchange 2000 system when connectivity is lost at either the server or the storage side by a controlled or uncontrolled pull of fibre. Two configurations are discussed. The first provides HBA failover when the server I/O connectivity is lost; however a cluster server failover occurs when storage I/O connectivity is lost. The second describes the proper setup for Dynamic Link Manager software to provide HBA failover when either server or storage I/O connectivity is lost. Performance measurements are included to show timings of cluster and HBA failover. The timings for the work-loads include no load on the Exchange Server and up to 6,000 simulated users running a LoadSim MMB2 workload.

Dynamic Link Manager Software Configuration 1

Dynamic Link Manager software configuration 1 (see Figures 1 and 2) is set up with four HBAs per node. For ease of illustration, only the Exchange databases are described for the Dynamic Link Manager software configuration. (See Table 4 for the Storage Utilization by Host.) On node 1, two HBAs are controlled by Dynamic Link Manager software to connect to devices behind port 1B and two HBAs are controlled by Dynamic Link Manager software to connect to devices behind port 1F (all ports are on storage cluster 1). On node 2, two HBAs are controlled by Dynamic Link Manager software to connect to devices behind port 2B and two HBAs are controlled by Dynamic Link Manager software to connect to devices behind port 2F (all ports are on storage cluster 2). If there is fibre loss at the server edge, an HBA failover results. If there is fibre loss at the storage edge, a cluster failover results.

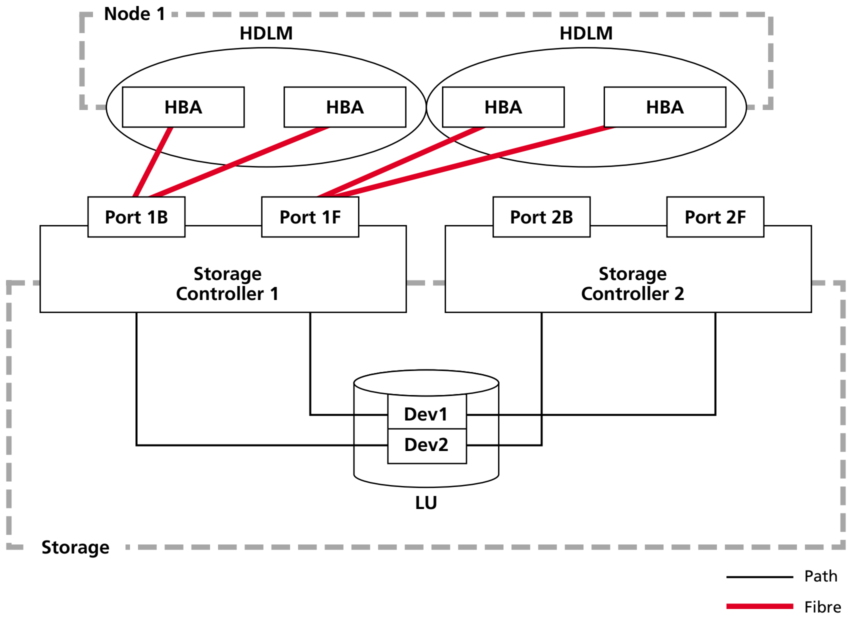


Figure 1: Node 1 in Dynamic Link Manager Software Configuration 1.

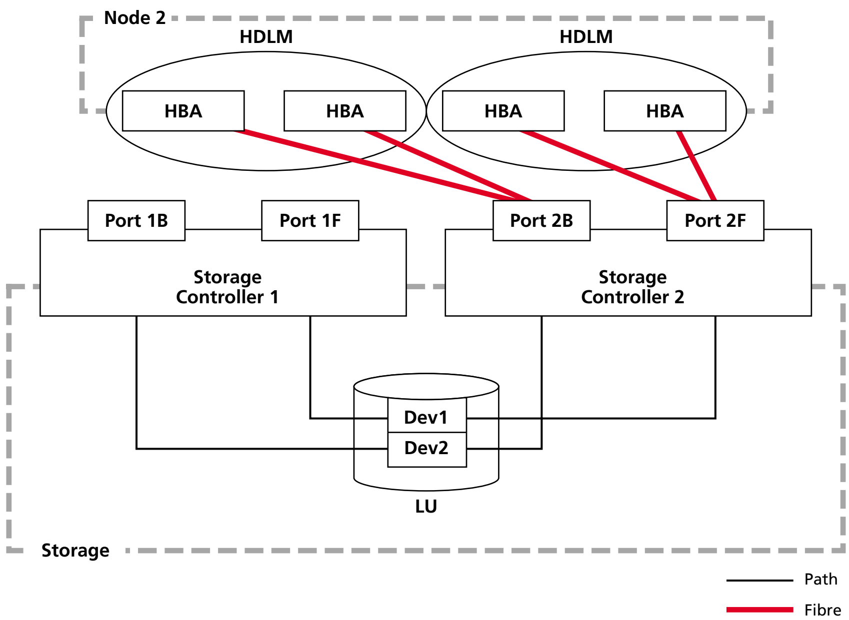


Figure 2: Node 2 in Dynamic Link Manager Software Configuration 1.

Dynamic Link Manager Software Configuration 2

Dynamic Link Manager Software configuration 2 (see Figures 3 and 4) is set up with four HBAs per node. For ease of illustration only, the Exchange databases are described for the Dynamic Link Manager software configuration. On node 1, two HBAs are controlled by Dynamic Link Manager software to connect to devices behind port 1B and port 2B, and two HBAs are controlled by Dynamic Link Manager software to connect to devices behind port 1F and port 2F (ports are spread between storage cluster 1 and storage cluster 2). The same configuration model is also set up for node 2. If there is fibre loss at the server edge, an HBA failover results. If there is fibre loss at the storage edge, an HBA failover results.

Note that the Dynamic Link Manager software configuration 2 does not cause a cluster failover but rather an HBA failover when fibre is lost at the server or storage edge. This results in overall improvement in path recovery times.

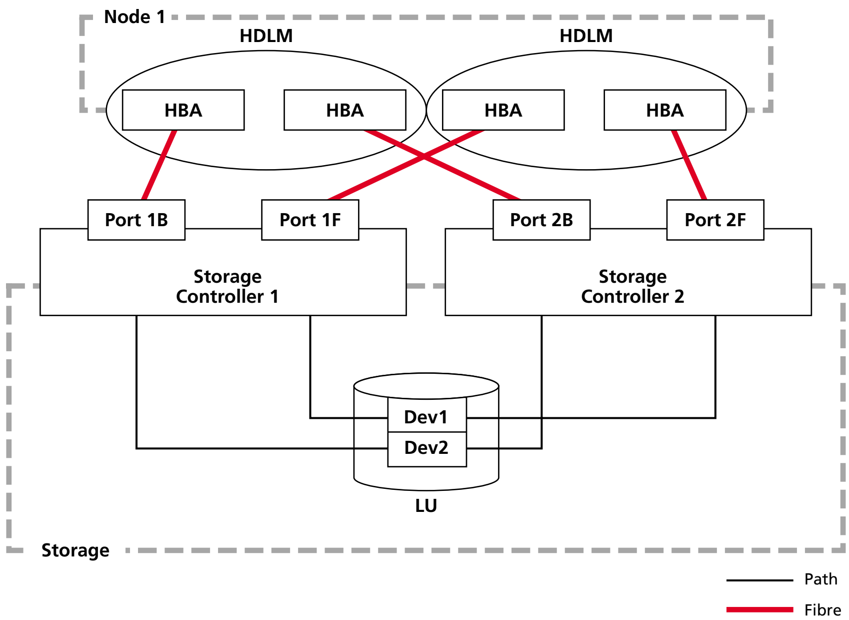


Figure 3: Node 1 in Dynamic Link Manager Software Configuration 2.

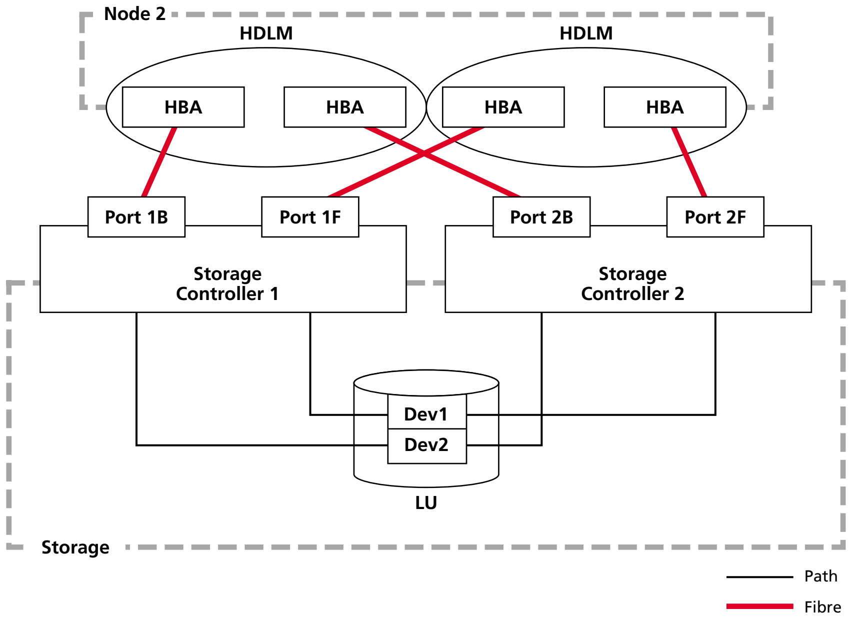


Figure 4: Node 2 in Dynamic Link Manager Software Configuration 2.

Dynamic Link Manager Software Connectivity

Configuration 1 Output

Table 5 depicts output from the DLNKMGR VIEW –PATH command for Dynamic Link Manager software configuration 1.

```
Paths:00008 OnlinePaths:00008
PathStatus  IO-Counts  IO-Errors
Online      29988      0

PathID PathName             DskName                          iLU  DkcPort
Status     Type I/O Counts  I/O-Errors  DNum HDevName
000000 004.0000.00001.0001 HITACHI .OPEN-L*2       .76A7        0000 1B
Online     Own       9255         0    0 M
000001 004.0000.00001.0002 HITACHI .OPEN-L*2       .76A7        0010 1B
Online     Own       8930         0    0 N
000002 005.0000.00001.0001 HITACHI .OPEN-L*2       .76A7        0000 1B
Online     Own          0         0    0 M
000003 005.0000.00001.0002 HITACHI .OPEN-L*2       .76A7        0010 1B
Online     Own          0         0    0 N
000004 006.0000.00005.0003 HITACHI .OPEN-L*2       .76A7        0002 1F
Online     Own       7288         0    0 O
000005 006.0000.00005.0004 HITACHI .OPEN-L*2       .76A7        0012 1F
Online     Own       4515         0    0 P
000006 007.0000.00005.0003 HITACHI .OPEN-L*2       .76A7        0002 1F
Online     Own          0         0    0 O
000007 007.0000.00005.0004 HITACHI .OPEN-L*2       .76A7        0012 1F
Online     Own          0         0    0 P
KAPL01001-I The Dynamic Link Manager command completed normally.
Operation name = view
```

Table 5: Dynamic Link Manager Software, Configuration 1.

Since four LUNs are used for the Exchange databases, four paths that show I/O counts are displayed in Table 5. Four HBAs are also installed on each of the nodes in the cluster. Node 1 HBAs are mapped as HBA 4, 5, 6, and 7. Of the four HBAs installed on the node, only two are active at any time. Much like a cluster configuration, the online HBAs can be considered active/passive. Dynamic Link Manager software controls the use of the active/passive HBAs by failing over from one HBA to the other after detecting a loss in I/O connectivity. When installing Dynamic Link Manager software, the administrator should first configure the HBAs to have identical LUN mappings from both. What controls the Dynamic Link Manager software configuration is how the port mappings are set up for the HBAs. In the case of Dynamic Link Manager software configuration 1, HBAs 4 and 5 use port 1B and are mapped to point to the same set of LUNs 1 and 2. HBAs 6 and 7 use port 1F and are mapped to point to the same set of LUNs 3 and 4.

The LUNs used for the Exchange databases are mapped as follows for HBAs 4 and 6 on node 1 of the cluster

- SG1-DB1, which is designated as PathName 004.000.00001.0001, (LUN 1)
- SG2-DB2, which is designated as PathName 004.000.00001.0002, (LUN 2)
- SG3-DB3, which is designated as PathName 006.000.00005.0003, (LUN 3)
- SG4-DB4, which is designated as PathName 006.000.00005.0004, (LUN 4)

The LUNs used for the Exchange databases are mapped as follows for HBAs 5 and 7 on node 1 of the cluster

- SG1-DB1, which is designated as PathName 005.000.00001.0001, (LUN 1)
- •SG2-DB2, which is designated as PathName 005.000.00001.0002, (LUN 2)
- SG3-DB3, which is designated as PathName 007.000.00005.0003, (LUN 3)
- SG4-DB4, which is designated as PathName 007.000.00005.0004, (LUN 4)

The PathName is indicated by a string of up to 19 one-byte characters. A path name consists of the port number for the HBA, bus number, target ID, and host LU number, each of which is separated by a period.

Notice that only the LUNs for HBAs 4 and 6 on node 1 are active in Table 5. This is indicated by the number of I/O counts. Also notice that SG1-DB1 and SG2-DB2 use port 1B and SG3-DB3 and SG4-DB4 use port 1F on node 1.

If there is a server-edge I/O failure for HBA 4 it will fail over to HBA 5 and still use port 1B. If there is a server-edge I/O failure for HBA 6 it will fail over to HBA 7 and still use port 1F.

If there is a storage-edge I/O failure for port 1B the cluster will fail over from node 1 to node 2 and use port 2B. If there is a storage-edge I/O failure for port 1F the cluster will fail over from node 1 to node 2 and use port 2F.

Since (in configuration 1) a storage-edge I/O failure will cause a cluster server failover rather than a HBA failover, let's now consider Dynamic Link Manager software configuration 2.

Configuration 2 Output

Output from the DLNKMGR VIEW –PATH command is shown for Dynamic Link Manager software configuration 2 in Table 6.

```
Paths:00008 OnlinePaths:00008
PathStatus  I/O Counts  I/O Errors
Online      45110       0

PathID PathName           DskName                       iLU  DkcPort
Status     Type I/O Counts  I/O Errors  DNum HDevName
000000 004.0000.00001.0001 HITACHI .OPEN-L*2       .76A7       0000 1B
Online     Own      13112          0    0 M
000001 004.0000.00001.0002 HITACHI .OPEN-L*2       .76A7       0010 1B
Online     Own      12270          0    0 N
000002 005.0000.00011.0001 HITACHI .OPEN-L*2       .76A7       0000 2B
Online     Own          0          0    0 M
000003 005.0000.00011.0002 HITACHI .OPEN-L*2       .76A7       0010 2B
Online     Own          0          0    0 N
000004 006.0000.00005.0003 HITACHI .OPEN-L*2       .76A7       0002 1F
Online     Own       8820          0    0 O
000005 006.0000.00005.0004 HITACHI .OPEN-L*2       .76A7       0012 1F
Online     Own      10908          0    0 P
000006 007.0000.00015.0003 HITACHI .OPEN-L*2       .76A7       0002 2F
Online     Own          0          0    0 O
000007 007.0000.00015.0004 HITACHI .OPEN-L*2       .76A7       0012 2F
Online     Own          0          0    0 P
KAPL01001-I The Dynamic Link Manager command completed normally.
Operation name = view
```

Table 6: Dynamic Link Manager Configuration 2 Output.

Since four LUNs are used for the Exchange databases, four paths that show I/O counts are displayed in Table 6. Four HBAs are also installed on each of the nodes in the cluster. Node 1 HBAs are mapped as HBA 4, 5, 6 and 7. Of the four HBAs installed on the node, only two are active at any time. Much like a cluster configuration, the online HBAs can be considered active/passive. Dynamic Link Manager software controls the use of the active/passive HBAs by failing over from one HBA to the other after detecting a loss in I/O connectivity. When installing Dynamic Link Manager software, the administrator should first configure HBAs to have identical LUN mappings from both. What controls the Dynamic Link Manager software configuration is how the port mappings are set up for the HBAs. In the case of Dynamic Link Manager software configuration 2, HBA 4 uses port 1B and HBA 5 uses port 2B. Both HBA 4 and 5 are mapped to point to the same set of LUNs 1 and 2. HBA 6 uses port 1F and HBA 7 uses port 2F. Both HBA 6 and 7 are mapped to point to the same set of LUNs 3 and 4.

The LUNs used for the Exchange databases are mapped as follows for HBAs 4 and 6 on node 1 of the cluster

- SG1-DB1, which is designated as PathName 004.000.00001.0001, (LUN 1)
- SG2-DB2, which is designated as PathName 004.000.00001.0002, (LUN 2)
- SG3-DB3, which is designated as PathName 006.000.00005.0003, (LUN 3)
- SG4-DB4, which is designated as PathName 006.000.00005.0004, (LUN 4)

The LUNs used for the Exchange databases are mapped as follows for HBAs 5 and 7 on node 1 of the cluster

- SG1-DB1, which is designated as PathName 005.000.00011.0001, (LUN 1)
- SG2-DB2, which is designated as PathName 005.000.00011.0002, (LUN 2)
- SG3-DB3, which is designated as PathName 007.000.00015.0003, (LUN 3)
- SG4-DB4, which is designated as PathName 007.000.00015.0004, (LUN 4)

The PathName is indicated by a string of up to 19 one-byte characters. A path name consists of the port number for the HBA, bus number, target ID, and host LU number, each of which is separated by a period.

Notice that only the LUNs for HBAs 4 and 6 on node 1 are active in Table 6. This is indicated by the number of I/O counts. Also notice that SG1-DB1 and SG2-DB2 use port 1B and SG3-DB3, and SG4-DB4 use port 1F on node 1.

If there is a server-edge I/O failure for HBA 4 it will fail over to HBA 5 and use port 2B. If there is a server-edge I/O failure for HBA 6 it will fail over to HBA 7 and use port 2F.

If there is a storage-edge I/O failure for port 1B then HBA 4 will fail over to HBA 5 and use port 2B. If there is a storage-edge I/O failure for port 1F then HBA 6 will fail over to HBA 7 and use port 2F.

This is an improvement from Dynamic Link Manager software configuration 1 since a storage-edge I/O failure will now cause an HBA failover rather than a cluster server failover.

Output language from the DLNKMGR VIEW -PATH command is explained in Table 7.

Item	Description
Paths	Sum of the number of displayed paths, indicated by up to five decimal numbers
OnlinePaths	Number of available paths in the displayed paths, indicated by up to five decimal numbers
PathState	Status of the displayed paths Online: All paths are available. Shrunken: Some of the paths are available. Offline: No paths are available.
I/O counts	Total I/O count for the displayed paths, indicated by up to ten decimal numbers
I/O errors	Total I/O error count for the displayed paths, indicated by up to ten decimal numbers
PathID	AutoPATH_ID indicated by up to six decimal numbers
PathName	Path name indicated by a string of up to 19 one-byte characters. A path name consists of the port number for the host bus adapter, bus number, target ID, and host LU number, each of which is separated by a period.
DskName	Storage system name indicated by a string of up to 38 one-byte characters. A storage system name consists of the vendor ID, product ID, and serial number, each of which is separated by a period.
iLU	LU number of the storage system, indicated by a string of up to four one-byte characters
DkcPort	Port number of the disk controller, indicated by a string of two one-byte characters.
State ***Note 1***	Status of the path Online: Online Offline (C): Offline by a command Offline (E): Offline due to an error
Type	Attribute of the path Own: Owner path Non: Non-owner path
I/O counts	Total I/O count for the path, indicated by up to ten decimal numbers
I/O errors	Total I/O error count for the path, indicated by up to ten decimal numbers
DNum	Device number, indicated by up to three decimal numbers. In Windows, the device number 0 is displayed. This number indicates the entire LU, but device 0 usually is not used.
HDevName	Host device name. In Windows systems, a single one-byte character indicates the drive letter. If no drive letter has been assigned, a hyphen (-) is displayed. In Windows 2000, the drive letter is displayed in the **HDevName** field even if the value of the **DNum** field is 0. ***Note***: the displayed drive letter is the drive letter for one of the devices included in the LU.

Table 7: Dynamic Link Manager Software Output Names Defined.

Note 1: A path that has the same **PathName** as a path whose state is **offline(E)** may be in an error status even though its state is **online**.

Test Description

Test measurements for failover timings are made for no load (no simulated users active) and for 3000, 4500, and 6000 LoadSim users.

Controlled tests are defined as simulating fibre loss by blocking I/O at the McDATA director using EFC Manager. The storage-edge test has I/O blocked for the storage ports. The server-edge test has I/O blocked for the HBAs.

Uncontrolled tests are defined as simulating fibre loss by physically pulling the fibre. The storage-edge test has fibre pulled from the McDATA director. The server-edge test has fibre pulled from the HBAs.

The HBA timing for failover can be described as how long it takes from the moment the fibre is pulled to when we get a response from the DLNKMGR command. The measurement timing uses a Windows 2000 resource kit command called TIMETHIS to measure elapsed time for the DLNKMGR command. The command is initiated by starting a batch file that issues the command every second. When fibre is pulled, the DLNKMGR command does not respond until the HBA has failed over. By looking at DLNKMGR timings, one can determine exactly how long the failover took.

Cluster failover timings are stopwatch.

Test Results

Notice that the timings include HBA or cluster for the type of Dynamic Link Manager software configuration used. Depending on the Dynamic Link Manager software configuration, the storage-edge tests result in a cluster failover for configuration 1 or an HBA failover for configuration 2. Interestingly, the HBA failover timings are faster for the server-edge test when using Dynamic Link Manager software configuration 2.

Using Dynamic Link Manager software configuration 1 with 6000 Loadsim users active the uncontrolled storage-edge failure test resulted in a corrupted MTA resource in the Exchange cluster. This may be due to the excessive amount of time it took for the server to do a cluster failover from cluster node 1 to cluster node 2 while under load. This needs to be further investigated to determine if there are timing settings for the Exchange MTA cluster resource that can be modified to improve the failover process and thereby prevent a corrupt MTA. In any event, the MTA corruption did not occur with Dynamic Link Manager software configuration 2 because the storage-edge failure test caused an HBA rather than cluster failover, resulting in faster recovery of I/O, which the Exchange system was able to handle without problem.

	Controlled Storage-edge I/O Failure Test	Controlled Server-edge I/O Failure Test	Uncontrolled Storage-edge I/O Failure Test	Uncontrolled Server-edge I/O Failure Test
Dynamic Link Manager Software Config 1				
No load	Cluster–35 secs	HBA–35 secs	Cluster–35 secs	HBA–35 secs
3,000 Users	Cluster–180 secs	HBA–90 secs	Cluster–180 secs	HBA–90 secs
4,500 Users	Cluster–180 secs	HBA–120 secs	Cluster–180 secs	HBA–120 secs
6,000 Users	Cluster–MTA error	HBA–140 secs	Cluster–MTA error	HBA–140 secs
Dynamic Link Manager Software Config 2				
No load	Not measured	Not measured	Not measured	Not measured
3,000 Users	HBA–45 secs	HBA–45 secs	HBA–45 secs	HBA–45 secs
4,500 Users	HBA–90 secs	HBA–90 secs	HBA–90 secs	HBA–90 secs
6,000 Users	HBA–120 secs	HBA–120 secs	HBA–120 secs	HBA–120 secs

Table 8: Test Results.

Without a proper Dynamic Link Manager software configuration, it is possible to unnecessarily fail over a server when I/O connectivity is lost at the storage edge. With a server failover that takes a significant amount of time, we've seen this can affect the startup of the Exchange MTA cluster resource on the second node of a cluster due to a corrupted MTA.

In the event of storage-edge I/O connectivity loss, Dynamic Link Manager software configuration 2 is the preferable method to use because the failover occurs at the HBA level and a server failover is avoided.

Summary

As described above, The Hitachi Data Systems Application Integration Lab's LoadSim benchmark testing has revealed a best choice configuration of Dynamic Link Manager software for a highly available Exchange 2000 configuration. In the event of storage-edge I/O connectivity loss this configuration is preferable, because the failover occurs at the HBA level and a server failover is avoided.

By evaluating collaborative options that Hitachi Data Systems enacts with its partners, we add proven, tested solutions to the high-quality, intelligent storage, open, standards-based storage management framework, and collaborative business model of our TrueNorth vision and strategy for the future of storage.

Appendix

Driver Parameters

The following are Hitachi Data Systems recommended driver levels, firmware levels, and parameter settings for the Emulex LP-8000 HBA covering:

- Firmware _Ver3.82a0
- BootBIOS _Ver1.53A1
- Full Port Driver _Ver5-2.11a2

For more information on any of these parameters, please contact the Hitachi Data Systems Technical Response Center at 1-800-348-4357.

	Initial Defaults	FC-AL Hitachi	Fabric Hitachi	Setting	Note
Link control button					
Point-to-Point			X		
Arbitrated Loop	X	X			
Link Speed	AUTO	AUTO	AUTO		
Adapter controls					
Automatically map SCSI devices	X	X	X		Note 1
Query name server for all N-Ports	X	X	X		
Allow multiple paths to SCSI targets	0	X	X		
Register for state change	X	X	X		
Use report LUNs	X	X	X		
Use Name Server after RSCN	X	0	X		
LUN mapping	0	0	X		
Scan in device ID order	X	X	X		
Enable Class 2 for SCSI devices	0	0	0		
Report unknown SCSI devices	X	X	X		
Look for disappearing devices	0	0	0		
Translate queue full to busy	0	0	0		
Use Bus Reset Status for retries	0	0	0		
Retry unit attention	0	0	0		
Retry PLOGI open failures	0	0	0		
Enable FCP-2 recovery	X	X	X		
Maximum number of LUNs	32	255	255		
Maximum queue depth	32	128	128		Note 2
Static poll destination address	BLANK	BLANK	BLANK		
Link timer	30	30	30		
Retries	64	64	64		
E_D_TOV	2000	2000	2000		
AL_TOV	15	15	15		
Wait ready timer	45	-1	-1		Note 3
Retry timer	2000	2000	2000		
R_A_TOV	2	2	2		
ARB_TOV	1000	1000	1000		
Registry parameters					
QuickFail	0	1	1		Note 4
NameServerDelay	0	800	800		

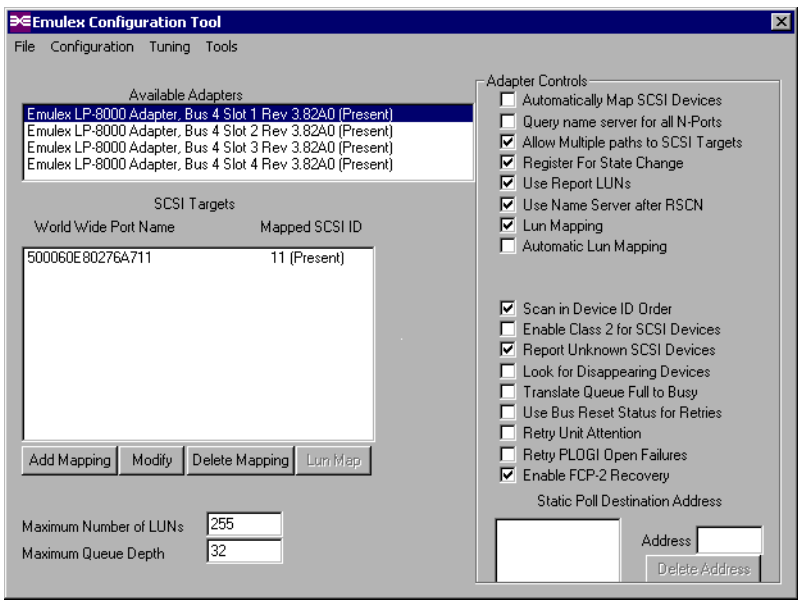
* The shading parts are needed to change the default value.

Appendix Table 1: Driver Parameters.

Note 1: X: Marked. 0: No marked.
Note 2: We changed to 128.
Note 3: If the timer values are set to -1, then the driver will use 2*R_A_TOV as the timer value. If the timer is set to 0, then the driver will wait indefinitely for devices to become available.
Note 4: This parameter needs to set for each adapter installed in the system.
QuickFail parameter should be set on each HBA entry and global entry of registry.

Configuration Tool

Use elxcfg.exe to set the value.
elxcfg.exe is the Emulex Utility attached Port Driver.



Appendix Figure 1: Sample Emulex Configuration Tool Screen.

QuikFail Keys

QuickFail is useful in redundant path setups. When QuickFail is nonzero, then I/O is returned with an error as soon as the WaitReadyTimeVal expires the first time. Also, if a node is not present in the Name Server response after an RSCN, the node is immediately marked off-line and I/O is returned. If set to 0, the driver will send a PLOGI after the initial WaitReadyTimeVal expiration and wait one additional WaitReadyTimeVal before returning I/O with an error.

Default value: 0
Type: REG_DWORD

If QuickFail is 1, the wait ready timer will only have to expire once before I/Os are returned to the upper levels. Also, when QuickFail is 1, any devices not present in the name server data obtained after an RSCN will be immediately marked as failed. This causes failover to occur more quickly. If QuickFail is 1 and the adapter is connected to a multiswitch fabric, the NameServerDelay value, see above, should be set to at least 800 milliseconds (ms).

Procedure for adding the QuickFail keys to the registry

Be very careful when editing the registry. Every change you make will be in real-time and wrong entries can cause damage to the system.

1. Click Start, Run and type regedt32.exe.
2. Select the HKEY_LOCAL_MACHINE hive.
3. For the QuickFail key there will 3 keys in total for a W2K machine with 2 HBAs. One key in the global entry of the registry (HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\elxsli2 and one for each HBA. (HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\elxsli2\ **BusXSlotX**).
4. To enter a key: Follow the structure as above for the location of the key in the HKEY_LOCAL_MACHINE hive. Select Edit from Taskbar, select "add key," enter the name of the key **QuickFail**, leave Class as blank.
5. Select the newly created key: click edit on taskbar and select "Add Value" For "Value name" enter the name of the key (**QuickFail** in this case) and for Data Type select "REG_DWORD" and hit OK.
6. Select the Radix as Decimal and enter the data value of 1 and hit OK.
7. Do the same for the 2 HBA card enties (BusXSlotX).
8. Just a note: Check in the Emulex configuration (elxcfg.exe) to verify in which Bus and Slot the HBAs are configured in. If the HBAs were removed and not replaced in original slot there may be more BusXSlotX entries than there are HBAs installed in the system.

Procedure for adding the NameServerDelay keys to the registry

1. The procedure is the same except you will only have 2 keys, one for every BusXSlotX key.
2. Click Start, Run and type regedt32.exe.
3. Select the HKEY_LOCAL_MACHINE hive.
4. (HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\elxsli2\ **BusXSlotX**)
5. To enter a key: Follow the structure as above for the location of the key in the HKEY_LOCAL_MACHINE hive. Select Edit from Taskbar, select "add key," enter the name of the key as **NameServerDelay**, leave Class as blank.
6. Select the newly created key: click edit on taskbar and select "Add Value." For "Value name" enter the name of the key (**NameServerDelay** in this case) and for "Data Type" select "REG_DWORD" and hit OK.
7. Select the Radix as Decimal and enter the data value of 800 and hit OK.
8. Do the same for the other HBA.

Hitachi Data Systems

www.hds.com

storage@hds.com

Corporate Headquarters

750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
(408) 970-1000
info@hds.com

Asia Headquarters

Suite 3301-6, Shell Tower
Times Square, 1 Matheson Street
Causeway Bay
Hong Kong
2525-2385
infoasia@hds.com

Australia/New Zealand Headquarters

Level 3
82 Waterloo Rd.
North Ryde, NSW 2113
Australia
+61-2-9325-3300
info.australia@hds.com

Canada Headquarters

2550 Victoria Park Avenue
Suite 601
Toronto, Ontario M2J 5A9
(416) 494-4114
www.hds.com
Cananda.Sales@hds.com

Europe Headquarters

Sefton Park
Stoke Poges
Buckinghamshire SL2 4HD
United Kingdom
+44 (0) 1753-618000
info.eu@hds.com
www.eu.hds.com

Latin America Headquarters

750 Central Expressway, MS 3468
Santa Clara, California 95050-2627
U.S.A.
(408) 970-7447
infolatin@hds.com

U.S. Headquarters

750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
(408) 970-1001
ussalesinfo@hds.com

WHP-131-01
December 2002