IBM

IBM System Storage™ Digital Media Storage Solution Installation Guide

Selling and Deploying DS3000/DS4000/DS5000 in Apple and StorNext Environments

Con Rice
Modular Storage Product Specialist
LSI Corporation
12007 Sunrise Valley Drive, Suite 325
Reston, VA 20191
Cell: 703-867-0012
Tel: 703-262-5418
e-mail: con.rice@lsi.com

Digital Media Storage Solution Installation Guide
February 25, 2010
© Copyright 2010, IBM Corporation

Contents

INTRODUCTION

There has long been a desire to support Apple servers and work stations with IBM DS series storage, and now we have that capability. But due to confusion in the past, we will be approving a very specific set of known, quality configurations. There are two basic aspects of these solutions: One is basic hardware connectivity, the second is the support for file sharing. It's important not to confuse the two. Basic hardware connectivity to Apple is supported via RPQ. File sharing is also supported, but because of the added complexity requires very specific configuration and a more restrictive RPQ process.

File Sharing Environments

Apple installations typically use a shared file system called Xsan. Xsan is in reality a lite version of Quantum's StorNext File System. StorNext is very popular in the media market because it allows shared block-level file system access while offering integrated archive management. However, Apple offers very limited support outside very specific configurations. A 100% apple configuration must be used.

There are three components to the StorNext solution: Clients, Meta Data Controller (MDCs), and the storage. There are also ancillary components such as the SAN, the TCP/IP network, and archive storage such as tape libraries. However, for selling and support purposes the components are software, support, and implementations services.
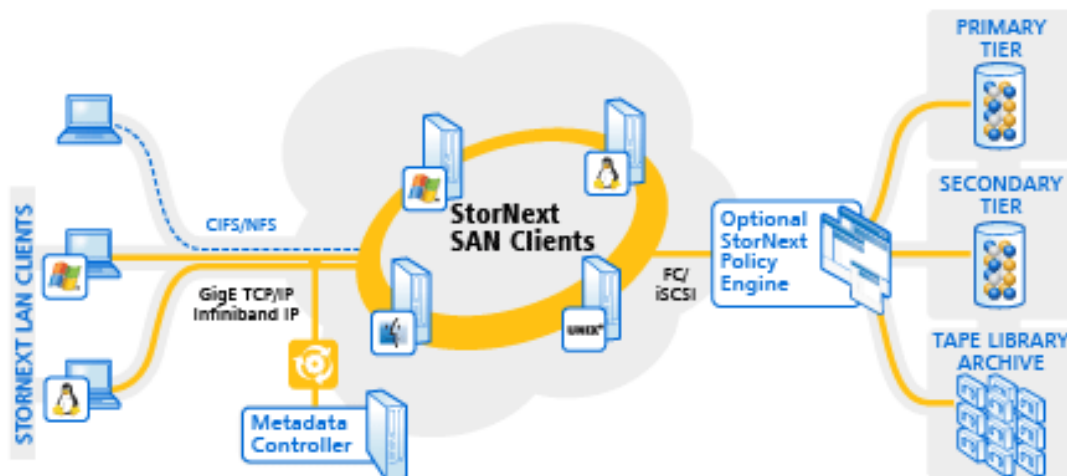
StorNext Environment

The announced support for the IBM Digital Media Storage Solution is the result of efforts between ATTO (HBA's), Quantum (StorNext), LSI (Storage Subsystems), and IBM. We have identified a set of hardware, firmware, and configuration parameters that will be supported. The set of configurations will be expanded in a logical manner over the next few months as additional choices are available and tested.

THREE CURRENT SOLUTIONS

Apple Homogenous

Apple Xsan Clients with Linux MDC (Meta Data Controller)
(Apple can also be connected without Xsan/StorNext support)
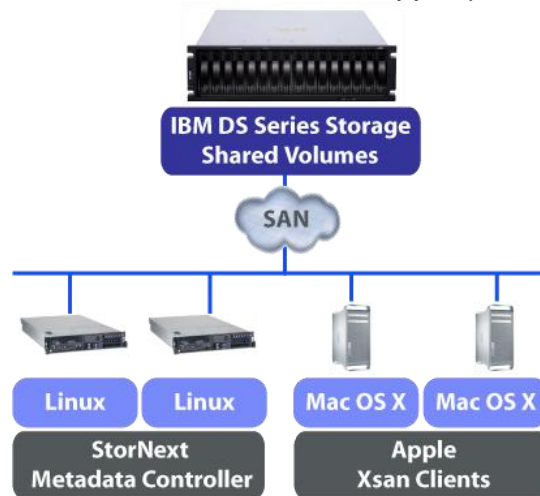
Linux Homogenous
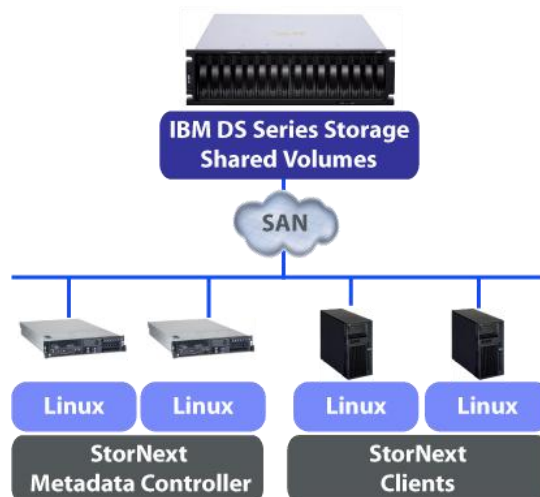
Linux StorNext Clients with Linux MDC

Heterogeneous

Apple Xsan Clients, Linux StorNext Clients, Windows Clients (June) and Linux MDC
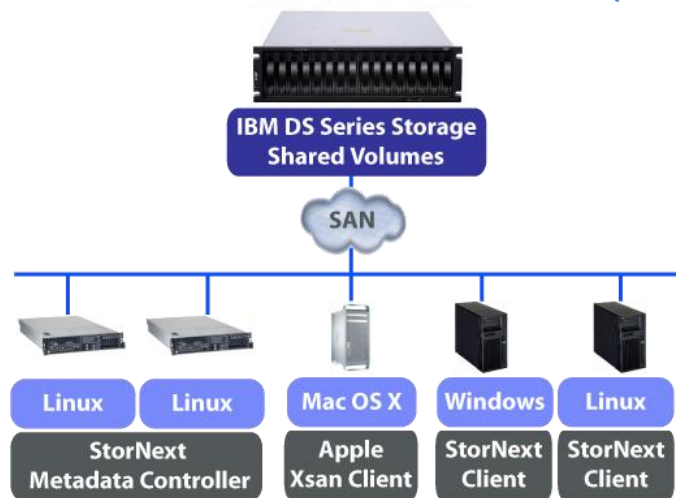
COMPONENT DESCRIPTIONS

Apple Clients
- Apple connection without StorNext/Xsan support, or
- May use Apple Xsan Client software
  - If you use the Xsan clients and Xsan Meta Data Controller (MDC), then your support comes from Apple. However, if you use the StorNext MDC from Quantum, then Quantum will support the entire solution including the Xsan clients loaded on the Mac's. As you can see, this is a much more flexible solution.
- Must use Linux StorNext Meta Data Controllers
- Must use ATTO HBA and Multipath driver
  - ATTO Celerity 8GB FC HBAs:
    - FC-84EN [Quad Port]
    - FC-82EN [Dual Port]
    - FC-81EN [Single Port]

Linux StorNext Clients and MDC's
- Must buy Quantum StorNext software
- Must buy Installation services from StorNext
- Must buy support contract from Quantum
- Supported Qlogic or Emulex HBA's
- RDAC MPP Multipath diver
- Servers must run approved versions of SLES or RHEL

SAN Architecture must meet RPQ criteria
- Limit number of paths for Xsan/StorNext clients to four
  - This is to limit both complexity and the amount of time required to initialize clients on the SAN
- Redundant Meshed Fibre Channel SAN (cross-connected)
  - RDAC MPP supports this configuration
  - Some IBM documentation may seem to preclude this, but those documents are based on an old non-"mpp" version of RDAC
  - Limits cause for failover to Controller failure
- Specific Firmware, software, and Operating System Levels

IBM Storage
- DS5100/DS5300
- DS5020
- DS4700
- DS3400
- Future

	DS3400	DS4700	DS5020	DS5100 base performance	DS5300 enhanced performance
Burst I/O rate cache reads (IOPS)	110,000	120,000	200,000	650,000	700,000
Sustained I/O rate disk reads	21,000	39,250	50,000	76,000	170,000
Sustained I/O rate disk writes	4,500	9,350	9,500	20,000	38,000
Drives	48 SAS/SATA	112 FC/SATA	112 FC/SATA/FDE	256 FC/SATA/FDE	448 FC/SATA/FDE /SSD* (480) SATA*
Burst throughput cache read	1,600	1,550MB/s	3,700 MB/s	1,600MB/s	6,400 MB/s
Sustained throughput disk read	925	980MB/s	1,600 MB/s	1,600MB/s	6,400 MB/s
Sustained throughput disk write	720	850 MB/s	1,400 MB/s	1,300 MB/s	5,300 MB/s
Host ports	4	4	4	8	16

PLANNED SUPPORT ROLLOUT

There are very specific capabilities that are being added and they are tied to specific hardware, software, and driver availability. Support will be added in this order:

Apple Homogenous

Apple Xsan Clients with Linux MDC (Meta Data Controller)
*(Apple can also be connected without Xsan/StorNext support)*
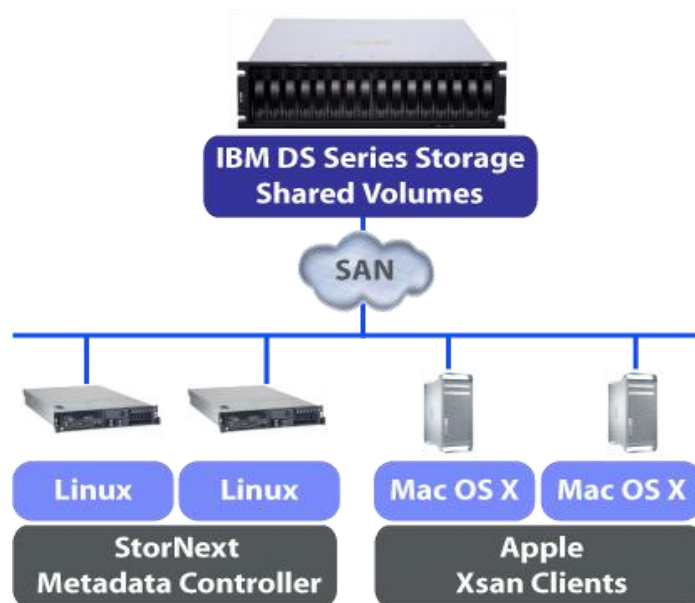
1. Single Host Client connected to DS3000/DS4000/DS5000 Storage (NOW)
   a. Apple Leopard and Snow Leopard
      i. Using ATTO HBA's and multipathing driver
   b. Without StorNext or Xsan support
   c. This is basic host connectivity support.

2. Multiple Apple Host Client connected to DS3000/DS4000/DS5000 Storage (NOW)
   a. Apple Leopard and Snow Leopard using Xsan Client software
      i. Using ATTO HBA's and multipathing driver
   b. With StorNext MDC's running on RHEL/SLES
      i. Using previously supported QLogic and/or Emulex HBA's

Linux Homogenous

Linux StorNext Clients with Linux MDC
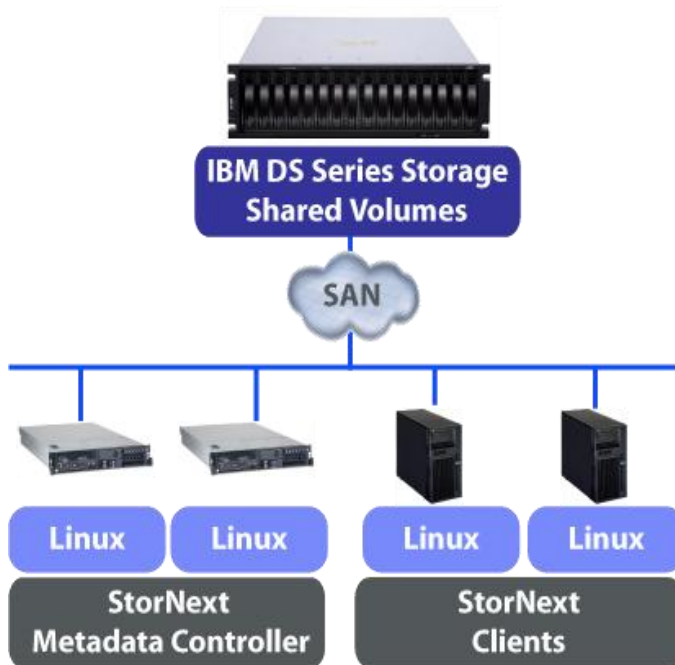
3. Multiple Linux Host Client connected to DS3000/DS4000/DS5000 Storage (NOW)
    a. RHEL/SLES using StorNext Client software
        i. Using previously supported QLogic and/or Emulex HBA's
    b. With StorNext MDC's running on RHEL/SLES
        i. Using previously supported QLogic and/or Emulex HBA's

Heterogeneous

Apple Xsan Clients, Linux StorNext Clients, Windows Clients (June) and Linux MDC
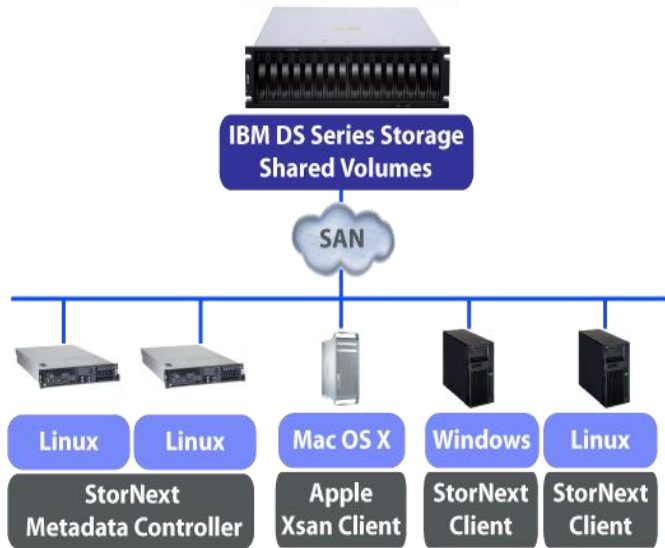
4. Mixed Host Apple and Linux Client connected to DS3000/DS4000/DS5000 Storage (NOW)
   a. Apple Leopard and Snow Leopard using Xsan Client software
      i. Using ATTO HBA's and multipathing driver
   b. RHEL/SLES systems using StorNext CLIENT software
      i. Using previously supported QLogic and/or Emulex HBA's
   c. With StorNext MDC's running on RHEL/SLES
      i. Using previously supported QLogic and/or Emulex HBA's

5. Single Host Windows Client connected to DS3000/DS4000/DS5000 Storage
   a. Windows client (XP, Vista, Windows 7) (June)
      i. Using ATTO HBA's and multipathing driver
   b. Without StorNext or XSAN support
   c. This is basic host connectivity support.

6. Add Windows Host Client in mixed environments support connected to DS3000/DS4000/DS5000 Storage
   a. Windows client (XP, Vista, Windows 7) (June)
      i. Using ATTO HBA's and multipathing driver
   b. With StorNext MDC's running on RHEL/SLES
      i. Using previously supported QLogic and/or Emulex HBA's
      ii. Using ATTO HBA's with either RDAC MPP or ATTO multipathing driver

Planned Future Enhancements

Additional Linux HBA choices and Windows Server 200x Meta data Controller

7. Expand RHEL/SLES connectivity choices (Planned)
    a. Add ATTO HBA's for either MDC's and/or clients using RDAC MPP
        i.   This step is not tied to any other steps and may occur at any time after step 3
    b. Add ATTO HBA's for either MDC's and/or clients using ATTO multipathing driver
        i.   This step is not tied to any other steps and may occur at any time after step 3

8. Add Windows Server 200x as either StorNext clients or as StorNext MDC's (Planned)
    a. Windows client Server 200x
        i.   Using previously supported QLogic and/or Emulex HBA's
        ii. Using ATTO HBA's with either RDAC MPP or ATTO multipathing driver
    b. We will NOT add this until AFTER the Windows Clients (XP, Vista, Windows 7) are added to reduce confusion

APPLE, WINDOWS, and STORNEXT SUPPORT MATRIX

	O/S	Architecture	HBA	MPF	Supported
Meta Data Controller					
Linux	RHEL 5.3	x64	*As currently supported (Qlogics/Emulex)*	RDAC (MPP)	Yes
Linux	RHEL 5.3	x64	ATTO	ATTO	Planned
Linux	SLES 10.2	x64	*As currently supported (Qlogics/Emulex)*	RDAC (MPP)	Yes
Linux	RHEL 5.3	x64	ATTO	ATTO	Planned
Client					
Apple OSX	Leopard	x64	ATTO	ATTO	Yes
Apple OSX	Snow Leopard	x64	ATTO	ATTO	Yes
Linux	RHEL 5.3	x86, x64	*As currently supported (Qlogics/Emulex)*	RDAC (MPP)	Yes
Linux	RHEL 5.3	x86, x64	ATTO	ATTO	Yes
Linux	SLES 10.2	x86, x64	*As currently supported (Qlogics/Emulex)*	RDAC (MPP)	Yes
Linux	SLES 10.2	x86, x64	ATTO	ATTO	Yes
Windows	2003 SP2	x86, x64	*As currently supported (Qlogics/Emulex)*	MPIO/DSM	Planned
Windows	2008 SP2	x86, x64	*As currently supported (Qlogics/Emulex)*	MPIO/DSM	Planned
Windows	2008 R2	x86, x64	*As currently supported (Qlogics/Emulex)*	MPIO/DSM	Planned
Windows	Windows 7	x86	ATTO	ATTO	Planned
Windows	Vista	x86	ATTO	ATTO	Planned
Windows	XP	x86	ATTO	ATTO	Planned
Mixed			*As appropriate above*	*As appropriate above*	

SOLUTION RESTRICTIONS AND RECOMMENDATIONS

There are several important aspects to the configuring of the SAN that must be understood. The first is that the SAN design is architected to avoid logical drive failover in the storage subsystem. This is accomplished by employing a meshed fabric where each HBA on the server can see both controllers on the storage subsystem. The second aspect has to deal with access to the shared logical drives. There are three types of shared logical drives; Metadata, Journal, and Data. All systems (MDC's and Clients) must be able to see the Data LUNs. The MDC's must also be able to both see the Metadata and Journal LUNs. It is important that the clients do not have access to the Metadata or Journal LUNs.

When configuring the storage partitioning within the storage subsystem, the methodology is as follows:

- All HBA's in the MDC's will be defined within a single host definition (called MDCs?). This host definition will have the Metadata and Journal LUNs directly mapped to this host definition.
- All Clients have their own host definitions. If desired, all clients that have the same OS can have their HBA's combined within a single host definition.
- All Clients and the MDC's will be assigned to a single host group (called StorNext?). All of the Data LUNs will be assigned to that host group.

	HBA's	Connections	Zone
MDC	Single or multiple (typical)	Up to four	All ports see Controller A and Controller B on DSxxxx
MDC - Limit number of defined paths to optimize initialization time. There are few MDC's in comparison to clients, so this choice should not be as restrictive as below. This is not a hard requirement, but is recommended. Most configurations consist of two MDC's for high availability. Both MDC's must have the same configuration; OS, HBA's, Driver Levels, etc			
Client	Single (typical) or multiple	Up to two	All ports see Controller A and Controller B on DSxxxx
Client – Limit number of defined paths to four to optimize initialization time. This is not a hard requirement, but is recommended.			
Storage	Controller **A**	Up to 8	All ports see all MDC and client ports
	Controller **B**	Up to 8	All ports see all MDC and client ports

RECOMMENDED RAID CONFIGURATIONS

	RAID level	RAID Array size	Segment size	Disk type
Journal	1 or 10	As required	256	FC or SAS recommended
Meta Data	1 or 10	As required	256	FC or SAS recommended

- Journal and Meta Data may reside in separate LUN's carved from the same RAID Array if required in small implementations. However, the ideal condition is separate LUN's tied to separate RAID groups for Journal and Meta Data. If possible, these two LUN's should be on a different storage system than the data LUN's.
- StorNext can stripe across multiple LUN's. It is recommended that multiple RAID 1 Groups be used for Journal and Meta data rather than similarly sized RAID 10's. This allows multiple threads to be talking to disk offering enhanced performance.

	RAID level	RAID Array size	Segment size	Disk type
Data	5	4+1	256	As required to meet performance criteria
	5	8+1	128	As required to meet performance criteria

- Data – Based upon StorNext file system header block size of 1MB, it is important to keep this in mind when selecting the segment size. RAID Array (also called Group) size and appropriate segment size should result in stripe size of 1MB. This will result in optimizing write throughput speeds. Ex. RAID5 4+1 (4 data) drives x 256K (seg size) = 1 MB Write stripe & RAID5 8+1x128K (seg size) = 1 MB stripe.
- Note – this rule does not apply to all deployments and could be dramatically different for oil and gas industry deployments among others. It is very important to understand the I/O block size of the applications sitting atop the StorNext File system.

	RAID level	RAID Array size	Segment size	Disk type
Data	6	4+2	256	As required to meet performance criteria
	6	8+2	128	As required to meet performance criteria

- Some customer applications may specify SATA drives and/or require additional RAID protection. The DS3000/DS4700/DS5000 storage servers offer additional levels of RAID protection if desired.

SOLUTION INSTALLATION

1. Use a meshed fabric with each HBA zoned to see both controllers but controllers unable to see each other.

2. Use RDAC or MPIO failover drivers as appropriate for your operating systems (ATTO for Apple). Do not use the StorNext failover driver

3. For Apple Clients

   a. Set host type **LNXCLVMWARE**
   b. Set **TPGS** On
   c. See script later in this document
   d. MAC's must use ATTO Celerity FC81-EN, FC82-EN, or FC-84-EN HBA's using the **1.3.1b12** driver

      Atto site
      http://www.attotech.com/

      Driver downloads,
      https://www.attotech.com/register/index.php?FROM=http://www.attotech.com/downloads.html

      The user is required to register and create a free account. Then navigate to,
      Fibre Channel HBA's for Storage Partners.
      IBM DS Series Celerity FC-8xEN

4. When connecting to MAC systems
   a. Switched environments preferred
   b. In non switched environments, may need to reboot MAC if it doesn't see storage
   c. To eliminate this issue - configure the ATTO host adapter and DS controller for a "Point to Point" connection instead of allowing them to auto negotiate."

5. For Linux Clients mixed with Apple Clients

   a. If Apple servers or clients are sharing physical connection to a storage system, it is important to use the proper host type definitions.
   b. Since we are using **LNXCLVMWARE** for apple, use **LINUX** for Linux servers and clients.

6. How to handle mixing multiple server and client O/S types need to access same LUN's:
   a. The MDC's in the approved solutions will always be Linux (Windows may also eventually be supported)
   b. Clients can be Apple or Linux (and also eventually Windows)
   c. To allow disparate O/S host connections:
      1. Create a Host Group with no Host type associated
      2. Assign multiple host types to this group.

7. Set **AVT Off** for all Host Types.
   a. See script later in this document

8. Recommended cache settings: Set pre-fetch value to any number other than 0 (zero) to enable automatic prefetch.

9. For clients, limit the number of defined paths to four. This gives you the redundancy you need while minimizing boot time. During boot, StorNext tries EVERY defined path.

10. RAID 1 or 10 for Journal and Meta Data
    a. Multiple RAID 1 Groups would be preferable to a similarly size RAID 10
    b. Placing Meta Data and Journal on a storage system separate from the data is preferable.

11. It is desirable to separate Meta data and Journal LUN's and access from the Data LUN's.
    a. In small installations, Meta data, Journal and Data can be different LUN's on the same RAID group
    b. However, it is best to separate MDC and Journal into their own RAID 1 RAID group.
    c. When Data gets to be large (approximately 250,000 files or more), it is best to completely separate Meta Data and Journal into their own storage system.
    d. A typical segment size for the MDC RAID group(s) is 256. Though this may vary with various applications. It is critical to understand the I/O demands of the applications being deployed.

12. The size and RAID type used for Data are really dependent upon customer requirements. But these levels are recommended for typical media applications:
    a. RAID 5 4+1 or RAID 5 8+1
    b. Explanation: RAID 5 4+1 means a RAID group using RAID 5 with 4 data drives and one parity drive. Even though in actuality, the data and parity would be striped across all drives.

13. Some applications (especially those using SATA drives) use different RAID types. Many scientific applications might keep very large data stores on relatively inexpensive SATA disks. In these cases, RAID 6 may be desired.
    a. RAID 6 4+2 or RAID 6 8+2

14. Typically, the media applications using StorNext write in 1M block. So:
    a. Using the formula Block size / number of data disks
    b. RAID 5 **4**+1 using 256K segment size
    c. RAID 6 **4**+2 using 256K segment size
    d. RAID 5 **8**+1 using 128K segment size
    e. RAID 6 **8**+2 using 128K segment size
    f. These segment sizes are representative of the 1 meg I/O size typical in most media industry applications. However, some applications (notably those in the oil and gas industry) may require different segment sizes. It is critical to understand the I/O demands of the applications being deployed.

15. Apple Clients HBA's: ATTO HBA's use default settings

16. Linux Clients HBA's: Emulex and QLogic HBA's use default settings.

17. Linux MDC and Clients HBA's: Emulex and QLogic HBA's use default settings.

18. Use standard performance tuning documentation as regards to disk and RAID group layouts etc.

QUANTUM / APPLE JOINT PRODUCT SUPPORT PROGRAM

Quantum / Apple Joint Product Support Program

Overview

QUANTUM and Apple have agreed to a cooperative support relationship for QUANTUM's StorNext and Apple's Xsan products. This means joint customers are assured that both companies will work together to resolve any technical issues with our respective products.

Xsan Controlled SAN File System

In the case of the joint customer having an Xsan controlled SAN file System, (i.e. Xsan runs the main server or Meta Data Controller), then Apple should receive the first call. If QUANTUM does receive the first call QUANTUM will start a case report and try to determine if the issue is related to the QUANTUM portion of the install. If after some investigation it is determined that an Apple case remedy is required, then QUANTUM shall notify the Apple Technical Support Contact, and shall provide the customer's information. Apple shall contact the joint customer in accordance with that party's customer support and/or maintenance agreement with the joint customer. QUANTUM shall continue to be available to provide information and/or assistance, as reasonably necessary.

StorNext Controlled SAN File System

In the case of the joint customer having an StorNext controlled SAN file system (i.e. StorNext runs the main server or Meta Data Controller), then QUANTUM should receive the first call. If Apple does receive the first call Apple will start a case report and try to determine if the issue is related to the Apple portion of the install. If after some investigation it is determined that an QUANTUM case remedy is required, then Apple shall notify the QUANTUM Technical Support Contact, and shall provide the customer's information. QUANTUM shall contact the joint customer in accordance with that party's customer support and/or maintenance agreement with the joint customer. Apple shall continue to be available to provide information and/or assistance, as reasonably necessary.

Cooperation Effort Between Apple and QUANTUM

If, after making reasonable but unsuccessful efforts to provide a case remedy, the parties shall then endeavor to share information and use cooperative efforts to determine and/or isolate the cause of the case report. The party in whose product the cause is isolated shall deliver or implement a timely case remedy; the other party agrees to use reasonable efforts to supply information, assist, and confirm the case remedy. Apple and QUANTUM are responsible for support and maintenance of its own products, and are not authorized to support or maintain the products of the other party.

Closing a Case

The consent of a joint customer shall be required to close a case.

MIGRATING FROM XSAN MDC TO STORNEXT MDC

Q. Are there new features in Xsan 2.2 that make it harder to swap out XServe MDCs and replace them with StorNext Linux MDCs?

A: "NamedStreams" is the Apple Xsan feature that causes some problems with StorNext. If you don't turn it on, you're fine. By default, "NamedStreams" is turned on. NamedStreams will make things like the basic StorNext command line copy command, cvcp, not work. Before Xsan 2.2, you could have it copy the "._" files, but now you can't, so you're stuck doing a move with the Apple Finder (which is basically as good as cvcp anyway). Other than that, it shouldn't be much of an issue since it doesn't affect clients that don't know about it, and by definition Windows/Linux clients don't know about resource forks.

Q. Does a customer need to buy A special level of StorNext support to support StorNext in Xsan environments? Is this true whether you are using SN MDCs or Apple MDCs?

A: No, You don't need to buy any special support to get support for StorNext in an Xsan environment. See next page.

ATTO

ATTO Technology, Inc.
Value Proposition

ATTO Technology is a leader in storage and connectivity solutions. In business since 1988, ATTO has maintained a consistent strategy of technical innovation by continually creating new and exciting Host Bus Adapter (HBA) products that are on the leading edge of the storage market. Throughout the years, ATTO has maintained a reputation for providing high-performance products in a variety of storage connectivity technologies for a wide variety of market segments.

ATTO's customer base for host bus adapter solutions continues to grow each year from both our OEM and channel partners through various market segments.   ATTO has been manufacturing and shipping Fibre Channel host adapters since 1997 with an installed based of over 1,000,000 Fibre Channel ports. With a solid customer base and comprehensive line of HBA products which are augmented by complimentary storage-centric products, customers can be assured that ATTO is a capable business partner fully versed in the HBA market.

ATTO bases all host adapter products on a common architecture with the goal of delivering consistent, industry leading performance solutions to market.

- The current 8-Gb Celerity line is the 6th generation of host adapters.

- ATTO's Optimization Cores are a collection of internally developed, industry-tested solutions designed for the low-latency, high-bandwidth demands of high-performance streaming storage applications.

- ATTO's Exclusive Advanced Data Streaming (ADS™) Technology provides the ultimate in I/O performance
    - Controlled acceleration of large and small blocks of data provides faster, more efficient data transfer Accelerated PCI Bus management
    - Efficient Full Block Buffering
    - Optimized SCSI-3 Algorithms for Streaming Media applications
    - Improved Data Interleave Architecture

- ATTO designs are based upon extensive knowledge and experience on what it takes to move data for the video market. ATTO firmware and drivers are written with the specific goal of managing latency so that customers are able to move multiple streams of uncompressed and compressed video with alpha channels and audio.

The storage industry continues to advance at an unparalleled pace, and solutions are becoming more complex and more intelligent. ATTO is an industry leader at incorporating the solutions for tomorrow's storage needs in today's products.

Product Features	Celerity HBA's		
	FC84EN	FC82EN	FC81EN
Fibre Channel Ports	4	2	1
Maximum Data Rate	8Gb	8Gb	8Gb
Maximum Transfer Rate (Full Duplex)	6.4 GB/s	3.2 GB/s	1.6 GB/s
Maximum Transfer Rate (Half Duplex)	3.2 GB/s	1.6 GB/s	800 MB/s
Bus Type	PCIe 2.0	PCIe 2.0	PCIe 2.0
Bus Characteristics	X8	X8	X8
Optical Interface	SPF+LC	SPF+LC	SPF+LC
Maximum Cable Length	300m-2GB 150m-4GB 50m-8GB	300m-2GB 150m-4GB 50m-8GB	300m-2GB 150m-4GB 50m-8GB
Low Profile Form Factor	NO	YES	YES
Advanced Data Streaming (ADS)	YES	YES	YES
Software RAID Support	YES	YES	YES
Developers Kit (Target Mode and API)	YES	YES	YES
Windows (Server)	YES	YES	YES
Windows (work station / client)	PLANNED	PLANNED	PLANNED
Linux (Red Hat, SUSE)	YES	YES	YES
MAC OSX	YES	YES	YES
VMWare ESX	YES	YES	YES
RoHS Compliant	YES	YES	YES

http://www.attotech.com/selectioncharts/Celerity_Selection_Chart.pdf

ATTO HBA INSTALLATION

ATTO Technology and IBM have teamed up to offer a high performance Fibre Channel multipathing connection into the IBM DS Series for MAC OS X, Windows and Linux.

Multipathing = failover + load balancing

Key Components:
- *ATTO Celerity 8Gb FC HBAs (FC-81EN, FC-82EN, FC-84EN)*
- *ATTO custom Multipathing driver for IBM DS Series Storage*
- *ATTO Configuration Tool*
- *IBM DS Series Storage*

Product Features	FC Speed	# of FC Ports	Form Factor	Max Xfer Rate (1/2 duplex)	Bus Type
Celerity FC-81EN	8-Gb	1	Low Profile	800 MBPS	PCIe 2.0 x8
Celerity FC-82EN	8-Gb	2	Low Profile	1600 MBPS	PCIe 2.0 x8
Celerity FC-84EN	8-Gb	4	Full Height	3200 MBPS	PCIe 2.0 x8

Backward compatible with 4-Gb or 2-Gb storage

- Multipathing is the ability to send I/O over multiple paths to the same LUN
- Failover and load balancing done on a LUN basis
- Monitors I/O paths – identifies changes in path status
- Automatic re-routing of I/O upon detection of an interrupted connection
- Failover/failback transparent to the OS or application
- Increase performance by sending data over multiple paths

ATTO Config Tool Features
- Allows simple Setup for the user. This is critical as IT staffs in the digital video industry are limited.
- Shows the number of paths
- Shows the status of each connection
- Displays valuable FC Target and Initiator info   (ex. WWPN, WWNN, etc)
- Shows path data flow information in bytes
- "Locate" button to blink the LED of the Fibre port associated with the path selected.

ATTO OS X Driver Installation
- Download the latest driver file from www.attotech.com/soultions/ibm
- *** Driver does not ship with standard ATTO product ***
- Double-Click the osx_drv_celerity8_131MP.dmg file
- (This will create and launch the disk image)
- Click continue to agree to the "Software License Agreement"
- Select the "Destination drive (boot drive)" and select continue
- The driver will now automatically install
- Reboot the machine

## ATTO OS X Configuration Tool Installation

- Download the latest ATTO Config Tool from www.attotech.com/solutions/ibm
- Double-click the osx_app_configtool_328.dmg file (This will create and launch the disk image)
- Double-click on CfgTool_328
- Click next, to agree to the "Software License Agreement"
- Select the "Default Folder" and click next
- Choose "Full Installation" and click next
- The ATTO Configuration Tool will now install
- When complete the following screen will appear, select "Done"
   Note: The default path is: **/Applications/ATTO Configuration Tool**

Viewing multipathing information

- Launch the ATTO Configuration Tool (double click the Configuration Tool link located in the Applications folder).
- Expand the tree in the Device Listing window on the left until you see your device (DS System) and virtual disks (LUNs).
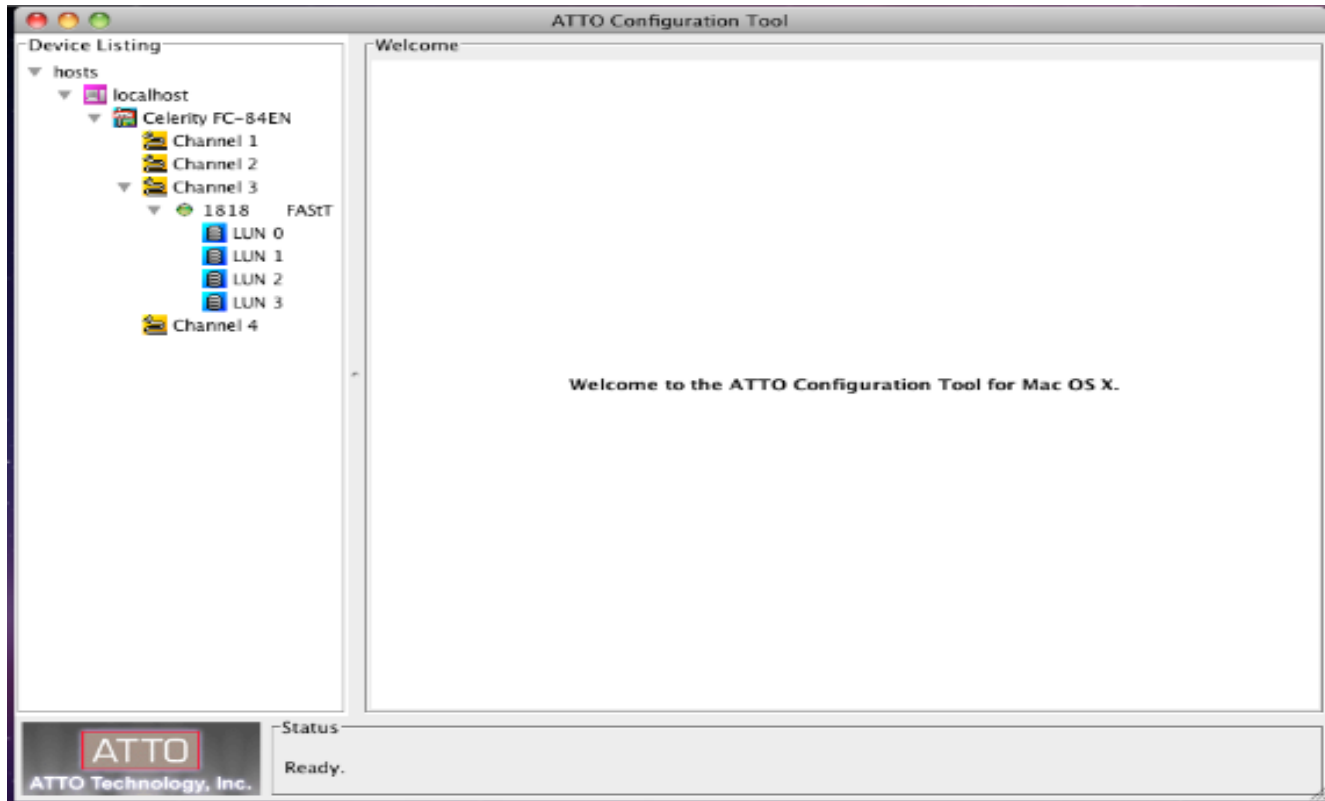
Paths Tab

The Paths tab in the multipathing window displays path information. The information displayed is based on a per Target or per LUN basis, depending on what device is selected to the left of the Configuration Tool.
- The upper half of the window displays status information for all paths.
- The lower half of the window displays information regarding the path selected in the upper window.

Paths (Target Base)
Click on the Target in the device tree to display the Target multipathing window on the right

Paths (LUN Base)
Click on a LUN below the device to display the LUN multipathing window to the right

Path Status at <u>Target</u> Level

The icon next to the device indicates the multipathing status for the target (not the LUN):

 A single path is connected to your device

 Multiple paths are connected to your device and are functioning normally

 Multiple paths **were** connected to your device, but one or more have failed but not all

 Multiple paths **were** connected to your device, but all have failed

Path Status at <u>LUN</u> Level

 Path connected to selected LUN is functioning normally

 Path connected to the selected LUN was working normally but has failed

 Path to the selected LUN is set to Alternate for Read and Write functionality

 Represents a Logical Unit
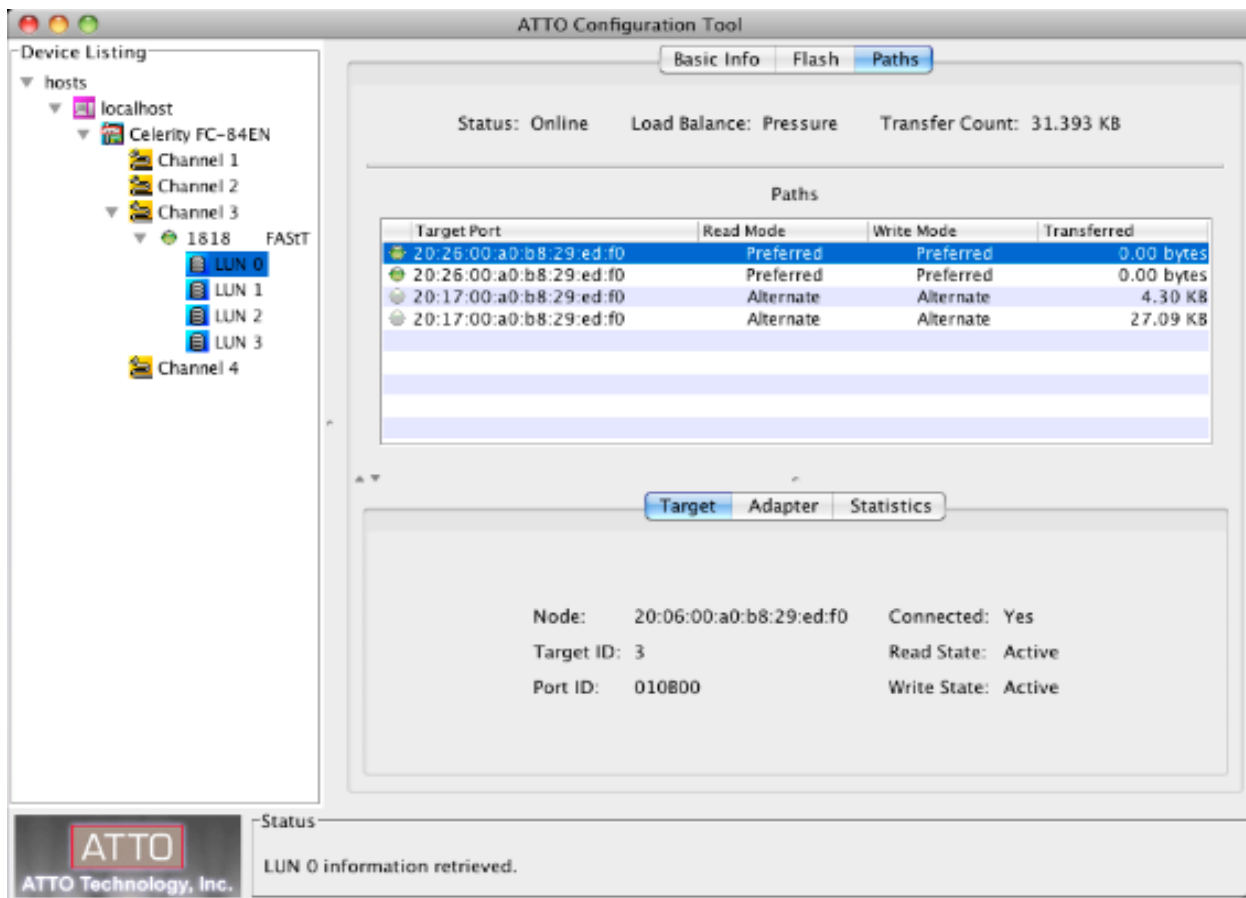
Additional Path Information

- Status – the overall status of all paths.
- *"Online" indicates all paths are connected.*
- *"Degraded" indicates one or more paths are not connected.*
- *"Offline" indicates all paths have failed*
- Load Balance – load balancing policy chosen (Round Robin, Queue Depth, Pressure)
- Transfer Count – the total number of bytes transferred on all paths.
- Target Port – the DS Storage WWPN (World Wide Port Name).
- Read Mode – indicates the selected paths read policy (Preferred, Alternate or Disabled)
- Write Mode – indicates the selected paths write policy (Preferred, Alternate or Disabled)
- Transferred – the total number of bytes transferred on a particular path.
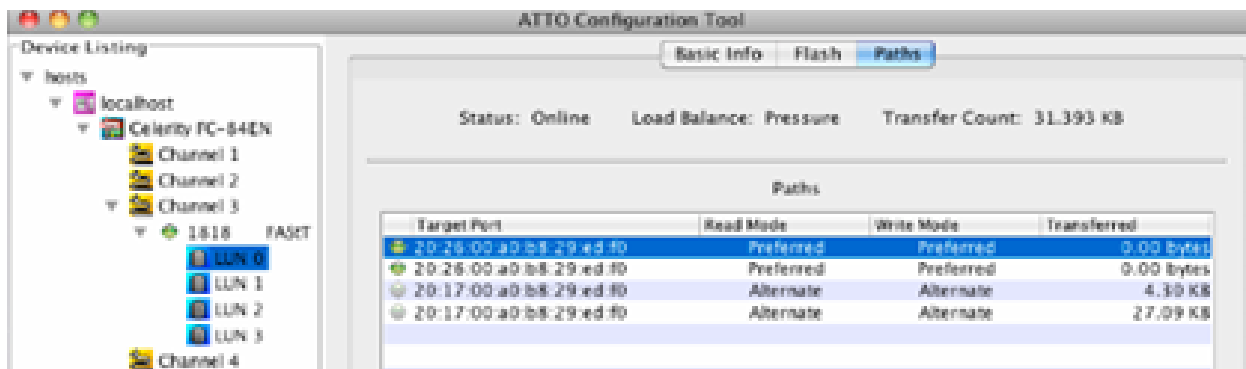
Detailed Info

Click on a path in the table to display detailed information for the path in the lower half of the window. Information is available in three tabs:

- Target – Displays information for the target port that is connected to the path
- Adapter - Displays information for the Celerity host adapter channel that is connected to the path
- Statistics – Displays information regarding data transferred down paths
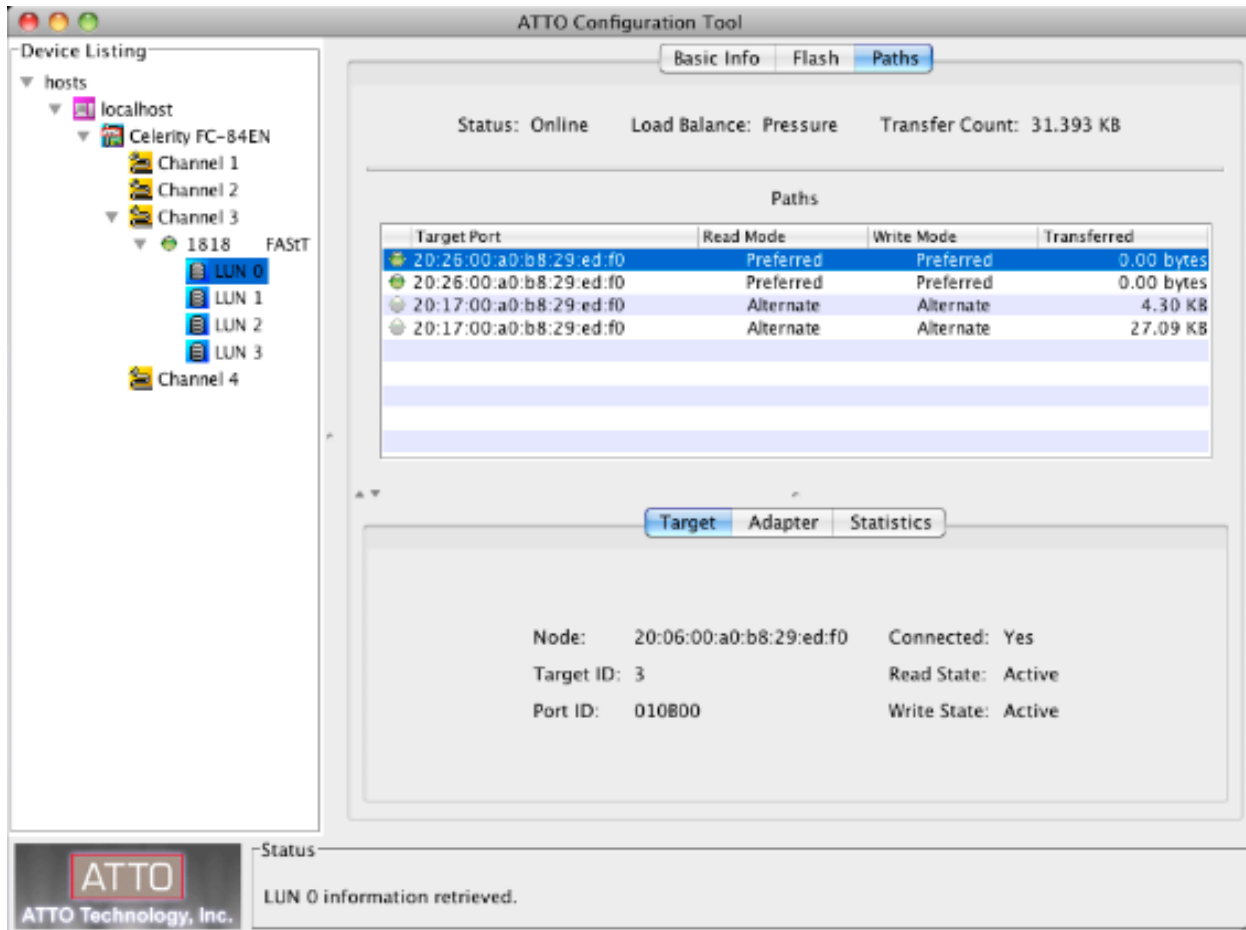
Target (DS Storage) Details
– Node – World Wide Node Name of the Target
– Target ID – Target ID presented to OS X
– Port ID – Fibre Channel Port ID of the DS Storage
– Connected – Link Status (Yes, No)
– Read State – Status of Read Policy (Active, Alternate, Disabled)
– Write State – Status of Write Policy (Active, Alternate, Disabled)
Detailed Adapter Info
– Adapter Details
    – Node – World Wide Node Name of the Celerity FC Adapter
    – Port – World Wide Port Name of the Celerity FC Adapter
    – Port ID – Fibre Channel Port ID of the Celerity FC Adapter
    – Topology – Fibre Channel Topology (PTP, AL)
    – Link – Status of Fibre link (up, down)
    – Link Speed – Fibre Channel Data Rate (8Gb, 4Gb, 2Gb)

Detailed Statistics Info
–   Statistics (Shown in Average and Total)
    –   Read Rate / Data Read – Read count in bytes for a particular path
    –   Write Rate / Data Written – Write count in bytes for a particular path
    –   Command Rate / Command Count – Number of commands sent down a particular path
    –   Error Rate / Error Count – Number of errors for a particular path

Load Balancing Policies
Pressure
>   The path with the fewest bytes will be used
>   This option provides the best performance for the digital video market (large transfers)

Queue Depth
>   The path with the fewest commands will be used

Round Robin
>   The least used path is selected

Configuring Multipathing
- *When new devices (Targets and LUNs) are presented to the ATTO HBA, multipathing must be configured and/or saved using the the ATTO Configuration Tool.*
- *User's can choose between default settings or custom settings (Load Balancing Policies).*
- *Default Settings for DS Storage*
- Devices are configured using the following policies:
- *Read/Write modes will be set automatically*
- *Load Balancing Policy for all targets, LUN's and paths is set to Pressure*
- *Saving Default Settings*
- Navigate to the Paths tab in the Apple toolbar
- Choose "Save Configuration"
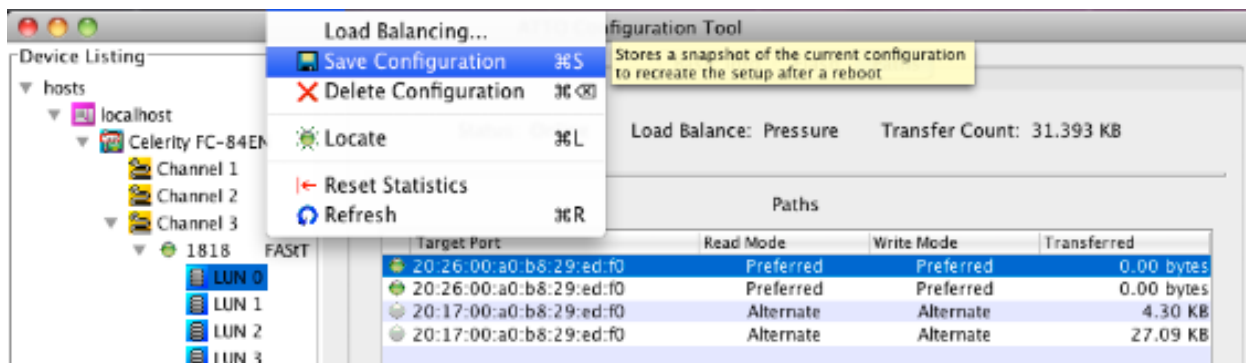
Custom Settings
- Load Balancing can be configured on a per LUN or per Target basis (depending on which is highlighted in the left pane)
- If the Target is highlighted, then **all** LUN's controlled by it will be configured the same.
- Highlight either the Taget or LUN to be configured
- *Navigate to the Paths tab in the Apple toolbar*
- *Choose "Load Balancing"*
- *Notice the configurable options in the lower right pane*
- *Select the Load Balancing Policy for the highlighted path*

Choose the load balancing policy for this LUN. Click **Finish** to use these settings or **Next** to configure the mode of each path.

Load Balance Policy:

| Round Robin |
| √ Pressure |
| Queue Depth |

(Back) (Next) (Finish) (Cancel)
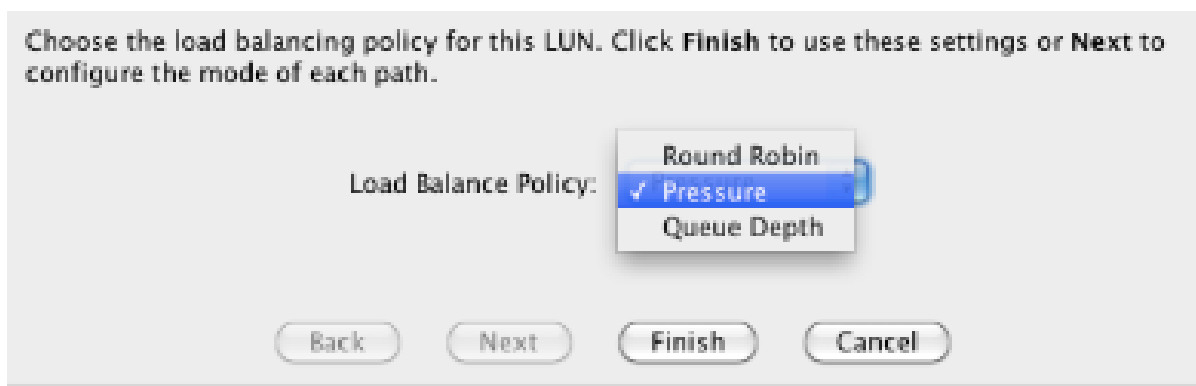
- Saving the Configuration
    - *When done, select "Finish"*
    - *Navigate to the "Paths" menu and choose "Save Configuration" followed by "Yes" when prompted*

LUN 4	● 50:05:07:68:01:40:2e:13	Preferred	Preferred	78.00 KB
LUN 5	● 50:05:07:68:01:20:2e:13	Preferred	Preferred	78.00 KB
LUN 6	● 50:05:07:68:01:10:2e:13	Preferred	Preferred	840.14 KB
LUN 7				255.28 KB
LUN 8				59.51 KB

Save Settings

Saving will allow the current configuration for this target to be displayed even after a link failure.

Would you like to continue?

(No) (Yes)

Node:	50:05:07:68:01:00:2e:0e	Connected:	Yes
Target ID:	0	Read State:	Active
Port ID:	010A00	Write State:	Active

If the connection status for a path changes, the Configuration Tool will automatically refresh the display
The Configuration Tool will periodically (~30 seconds) refresh the path statistics for all devices
Paths can be manually refreshed by choosing "refresh" in the "Paths" menu
Statistics can be reset to '0' by choosing "Reset Statistics" in the "Paths" menu
The "locate" button will blink the particular FC adapters channels LED (very useful troubleshooting tool)

Collecting Celerity Host Adapter Info
Click on the adapter under device listing.
- *Basic Info tab shows:*
    - HBA model number   - PCI slot number the adapter is located in
    - HBA driver version   - Location and name of the Celerity driver
- The flash tab is used to view flash information and update the flash version of the HBA
- Flash Tab shows:
    - HBA flash version - Flash update progress bar
    - Flash update information

FC Host Adapter Configuration
Click on the adapter channel under device listing.
Under the NVRAM tab you can:
- View / Change the Frame Size.
- View / Change the Connection Mode.
- View / Change the Data Rate.
- Load the HBA defaults.

Digital Media Storage Solution Installation Guide   27
© Copyright 2010, IBM Corporation. All rights reserved.

Changes to the adapters NVRAM settings must be changed on each channel of the HBA followed by the commit button and a restart

Useful OS X Applications

TextEdit – Text editor
Can be found in: /Applications/TextEdit
Disk Utility – OS X disk repair / formatting tool
Can be found in: /Applications/Utilities/DiskUtility
Console – Allows user to see application logs in real-time
Can be found in: /Applications/Utilities/Console
Finder – Standard file viewer (similar to Windows Explorer)
Terminal – Terminal application which allows a user to send Unix based commands *Can be found in: /Applications/Utilities/Terminal*
Kextstat | grep ATTO *– Shows which ATTO Fibre Channel driver is loaded into memory (key info)*
Apple System Profiler – System utility which will output a file with useful system info (processor type, RAM installed, OS version, driver versions, application versions, logs, etc.)
Can be found in: Apple Menu → About this Mac → More Info

ATTO Problems?

Escalation Checklist

1. *Computer Model:*
2. *Operating System:*
3. *OS Patch Level:*
4. *PCI or PCIe slot # and type (x4, x8, …):*
5. *ATTO driver version (from the Configuration Tool):*
6. *List of all of the devices attached to the ATTO HBA:*
7. *Did this configuration ever work?*
8. *Is this a new error that just started happening, or is this an error that has been around since the first use.*
9. *How duplicable is the error? Does the error occur sporadically/randomly, or can it be reproduced each and every time?*

Useful Logs to collect and escalate:

- Apple System Profiler Output (ASP)
- IOREG Output (from IOREG Command)
- Dump of ATTOCelerityFCLOG and ATTOCelerityFCMPLog Utility when the issue occurs
- Terminal output with the following commands:
- *Kextstat | grep ATTO*
- *Java – version*

APPENDIX: CONFIGURATION SCRIPTS

Setting TPGS On

This script turns on TPGS for host type LNXCLVMWARE (which is what we are using for OSX).
Set TPGS on by running the script below from the Management GUI:

```
show controller[a] HostNVSRAMbyte[13,0x28];
set controller[a] HostNVSRAMbyte[13,0x28]=0x02;
show controller[a] HostNVSRAMbyte[13,0x28];

show controller[b] HostNVSRAMbyte[13,0x28];
set controller[b] HostNVSRAMbyte[13,0x28]=0x02;
show controller[b] HostNVSRAMbyte[13,0x28];

reset controller[a];
reset controller[b];
```

Turning AVT Off

This script turns off AVT for all host types.
Set AVT off for all HOST REGIONS by running the scripts below from the Management GUI:

```
/* Disable AVT in all host regions */
set controller[a] HostNVSRAMByte[0x00, 0x24]=0x00;   /* 0x01 is enable AVT */
set controller[a] HostNVSRAMByte[0x01, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x02, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x03, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x04, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x05, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x06, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x07, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x08, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x09, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x0a, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x0b, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x0c, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x0d, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x0e, 0x24]=0x00;
set controller[a] HostNVSRAMByte[0x0f, 0x24]=0x00;

set controller[b] HostNVSRAMByte[0x00, 0x24]=0x00;   /* 0x01 is enable AVT */
set controller[b] HostNVSRAMByte[0x01, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x02, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x03, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x04, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x05, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x06, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x07, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x08, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x09, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x0a, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x0b, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x0c, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x0d, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x0e, 0x24]=0x00;
set controller[b] HostNVSRAMByte[0x0f, 0x24]=0x00;
show storageArray hostTypeTable;
```

APPENDIX: EXAMPLE CONFIGURATIONS

Large Heterogeneous StorNext configuration
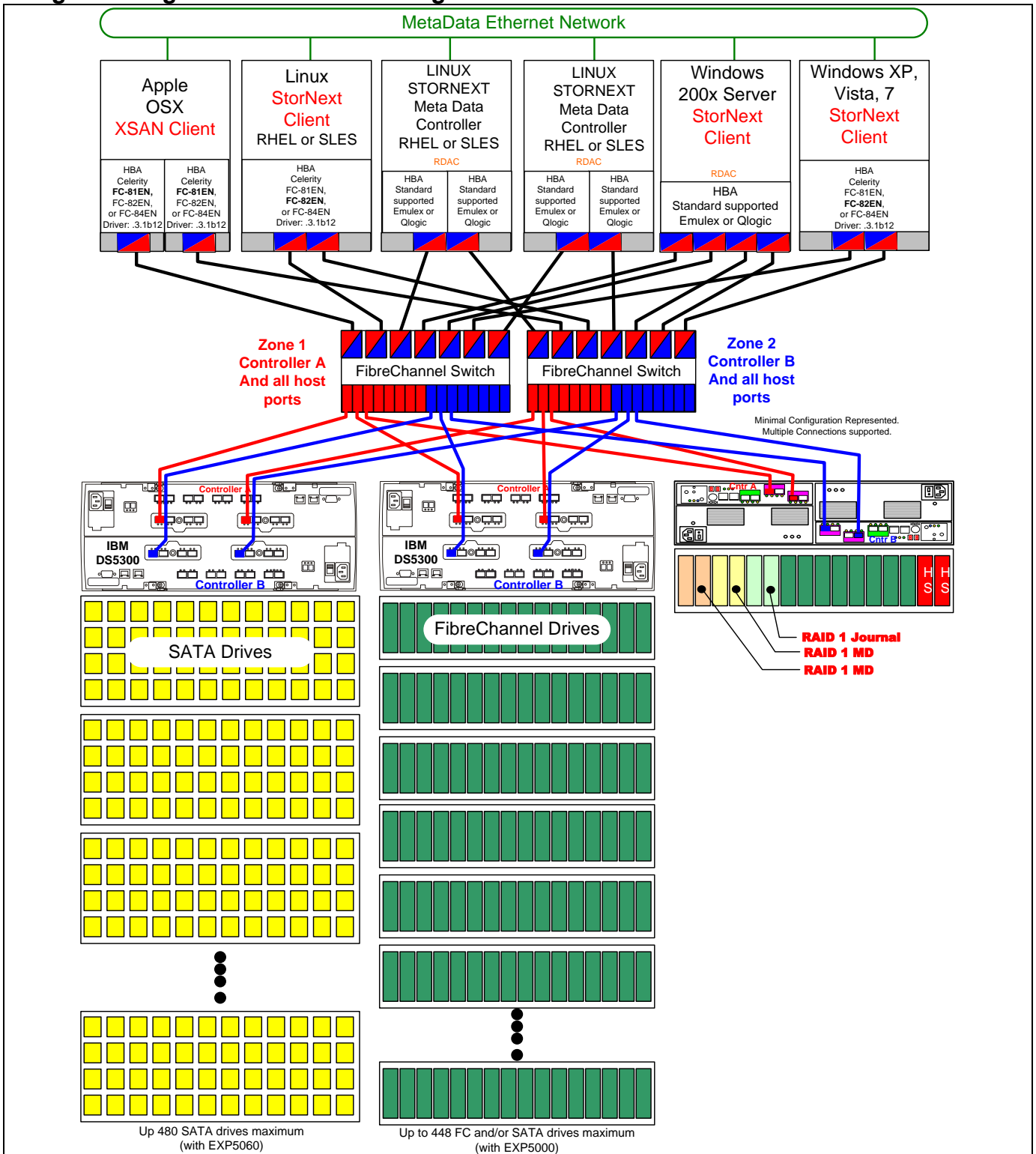
Example 1.Drawing depicts large and expandable SAN with heterogeneous client connectivity. Multiple DS5300 connected to meshed SAN fabric with a minimum of two Fibre Channel switches. For DS5100/DS5300 configurations, a minimum of two switches is recommended. Drawing shows multiple DS5300 for Data storage and a separate DS5020 for Meta Data.

Detailed explanation of example 1.
Example Large StorNext configuration showing DS5300's with Fibre Channel and/or SATA disk for Data and a separate DS5020 for Meta Data and Journal. These examples are representative only and should not be seen as restrictive. The idea is to use our knowledge of these products to craft solutions with the proper performance and resiliency characteristics to meet customer needs and budget concerns.

DS5020 for Journal and Meta Data with FC drives
- This is an example configuration only
- RAID 1 for StorNext Meta Data
- RAID 1 for StorNext Journal
- Add additional RAID 1's rather than growing to RAID 10 for performance reasons.
- Controller A can see all client and MDC host ports but not Controller B
- Controller B can see all client and MDC host ports but not Controller A

DS5300's with FC and/or SATA for Data
- This is an example configuration only
- Multiple RAID 5 (both 4+1 and 8+1) RAID Arrays for StorNext data
- Multiple RAID 6 (both 4+2 and 8+2) RAID Arrays for StorNext data may also be desirable in some situations
- Controller A can see all client and MDC host ports but not Controller B
- Controller B can see all client and MDC host ports but not Controller A

Mixed Clients – SAN Zoning
- Clients have single dual ported HBA's with connections into a redundant switched network and zoned to be able to see both Controller A and B
- MDC's have dual dual ported HBA's with connections into a redundant switched network and zoned to be able to see both Controller A and B

**All Storage Configurations for StorNext environments**
- AVT should be set OFF for all partitions and host types
- For **MAC Clients** There is not currently a host type definition for MAC Clients (there will be in the future). Because of this a work-around has been developed for use until the MAC host type is available.
    1. Set **host type LNXCLVMWARE**
    2. Set **TPGS on**(via a provided script)
- For **Linux Clients (and VMware servers)**
    1. If no MAC clients exist, then the host type LNXCLVMWARE should be used, as this host type has AVT disabled.
    2. If MAC hosts exist in the environment, Linux clients (and VMware servers) should use the host type LINUX,
    3. Insure that AVT is turned off for all host regions, including LINUX (via a provided script). This can be done by checking the bottom of the complete (all tab) system Profile.
- For mixed environments:
    1. Create a Host Group for all MDC's and Clients that will access the StorNext file system. There is no Host type associated with Host Groups (only individual hosts)
    2. Assign multiple host types to this group.

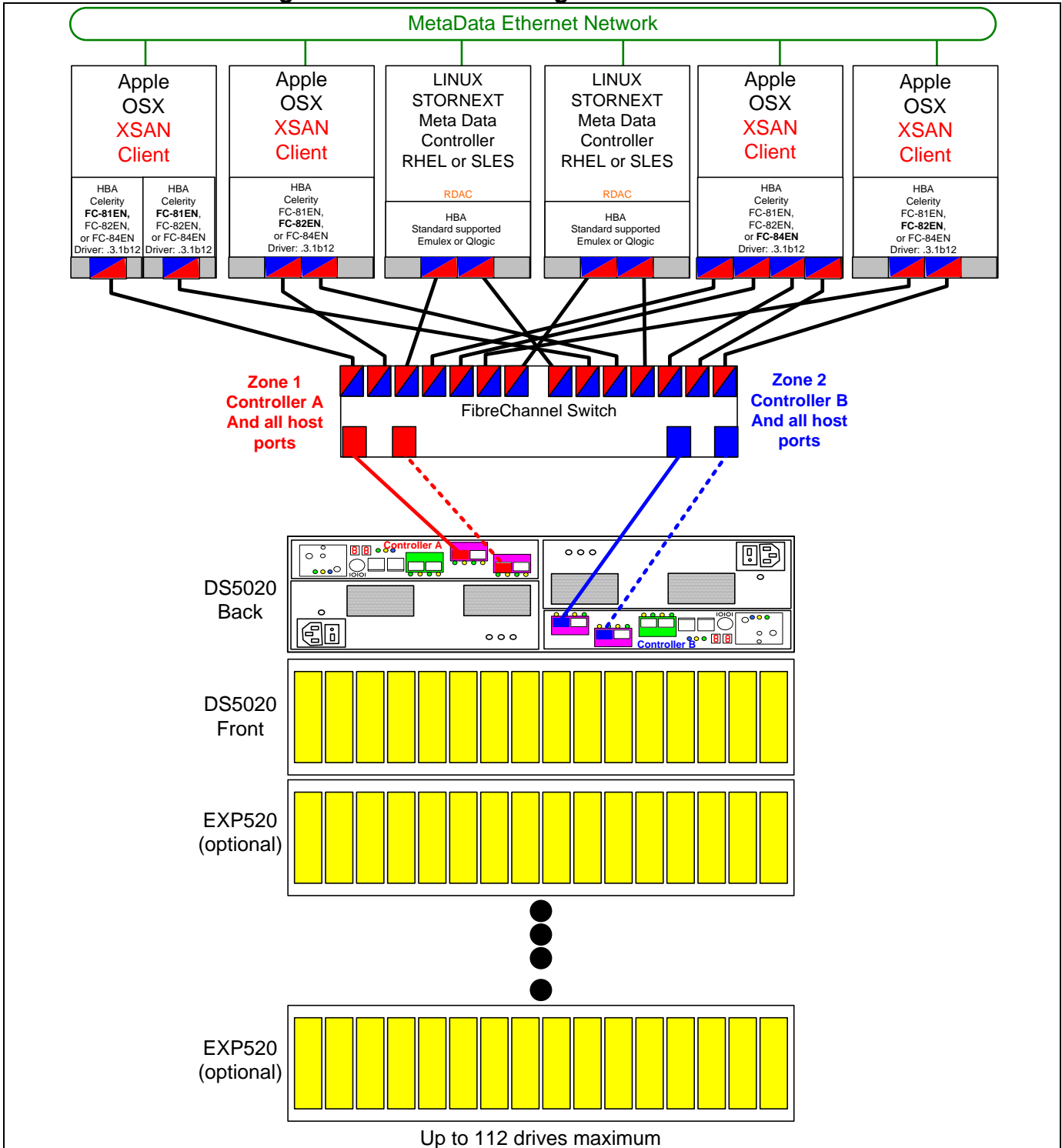## Mid to large -sized "MAC Homogenous" StorNext configuration
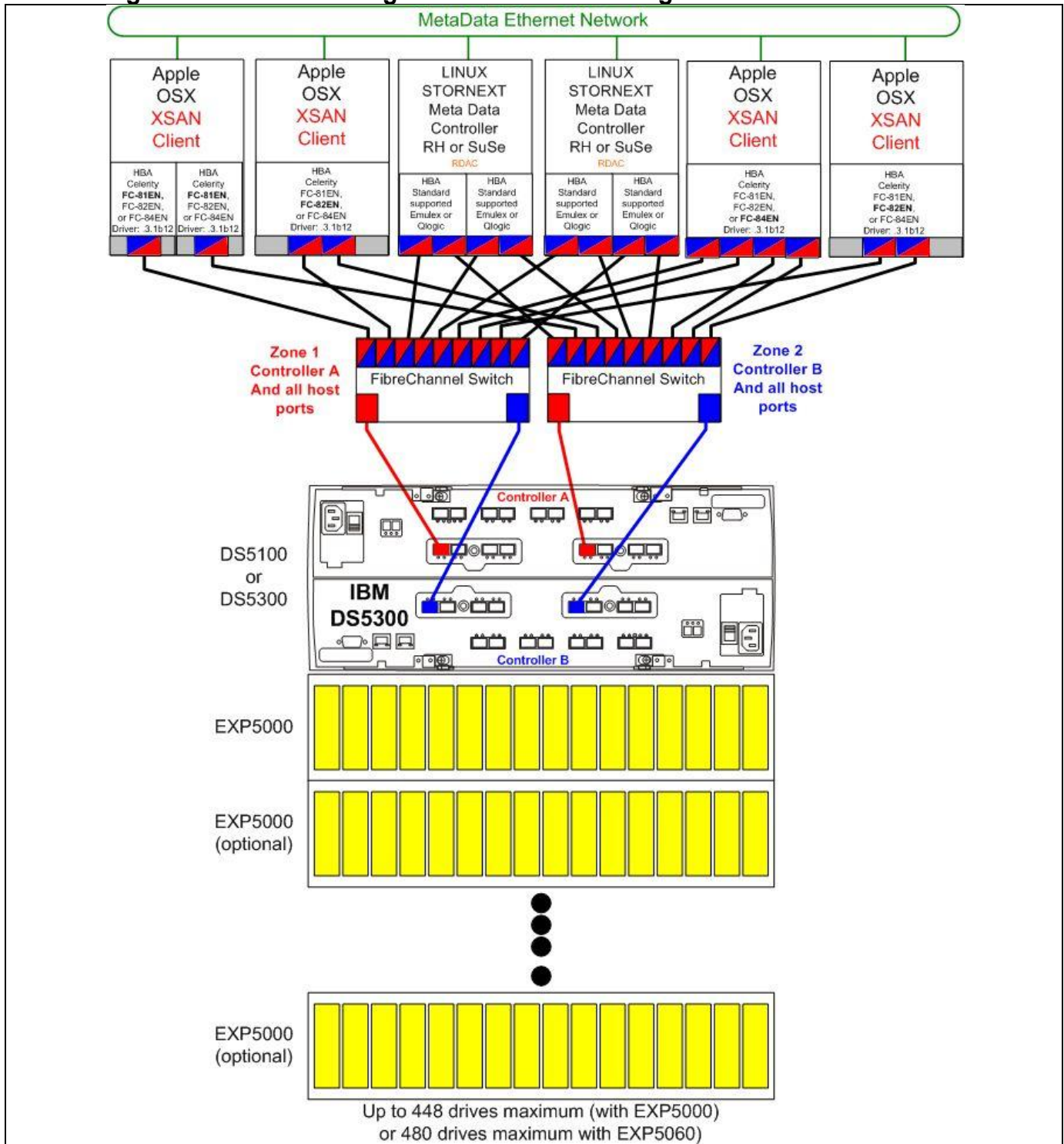
Example 2. In this example diagram Linux MDC Servers are shown, but only MAC Xsan Clients are shown. DS5300 connected to meshed SAN fabric with a minimum of two Fibre Channel switches. For DS5100/DS5300 configurations, a minimum of two switches is recommended. Drawing shows minimal SAN to DS5300 connections to simplify the drawing. However, up to 16 connections is supported. In this example, the Meta data Controllers each have a two dual ported HBA. Multiple multi-ported HBA's is not required, but it is recommended. Also in this example multiple HIC cards on the DS5300 are used to increase resiliency.

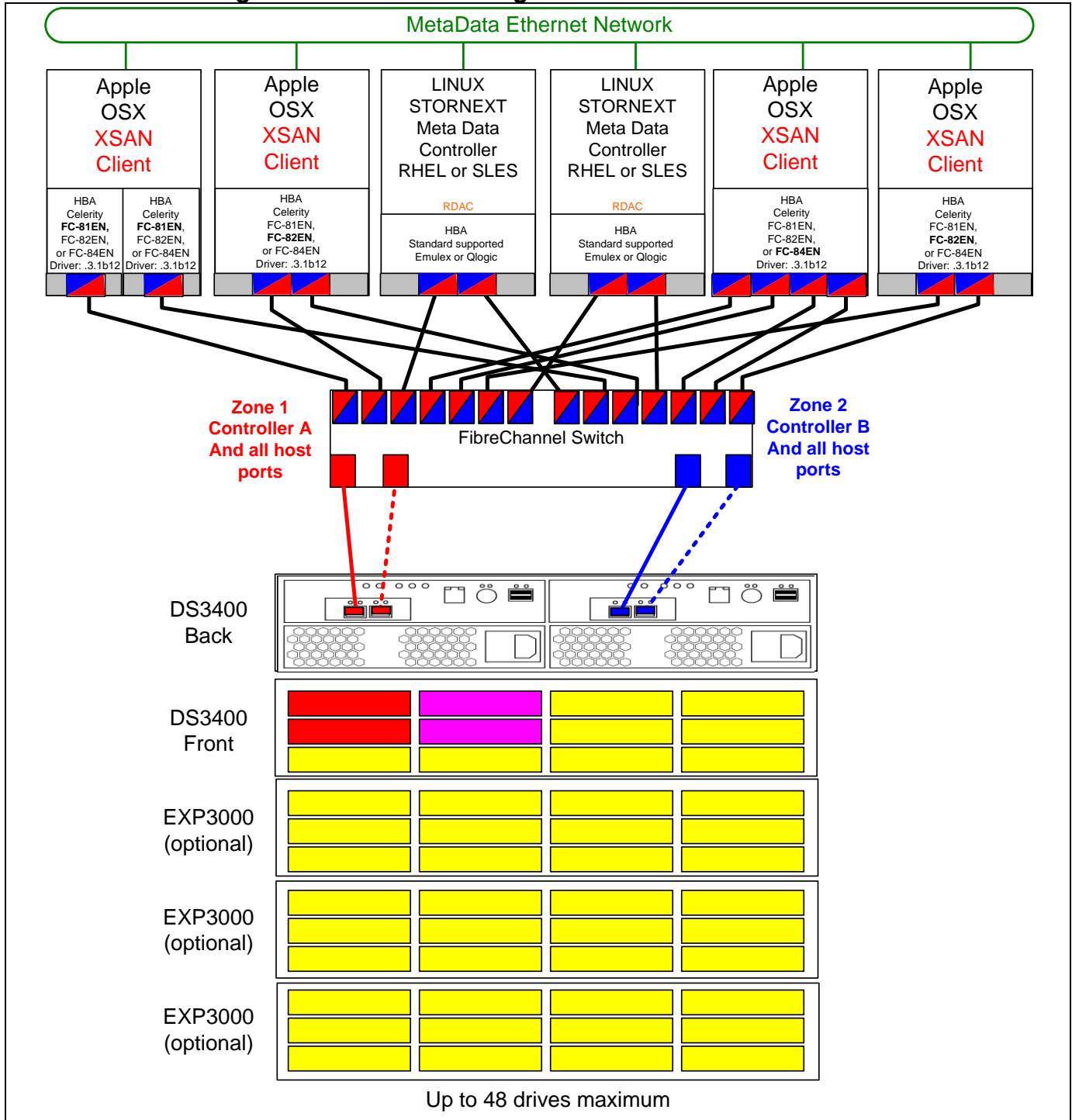## Mid-sized "MAC Homogenous" StorNext configuration
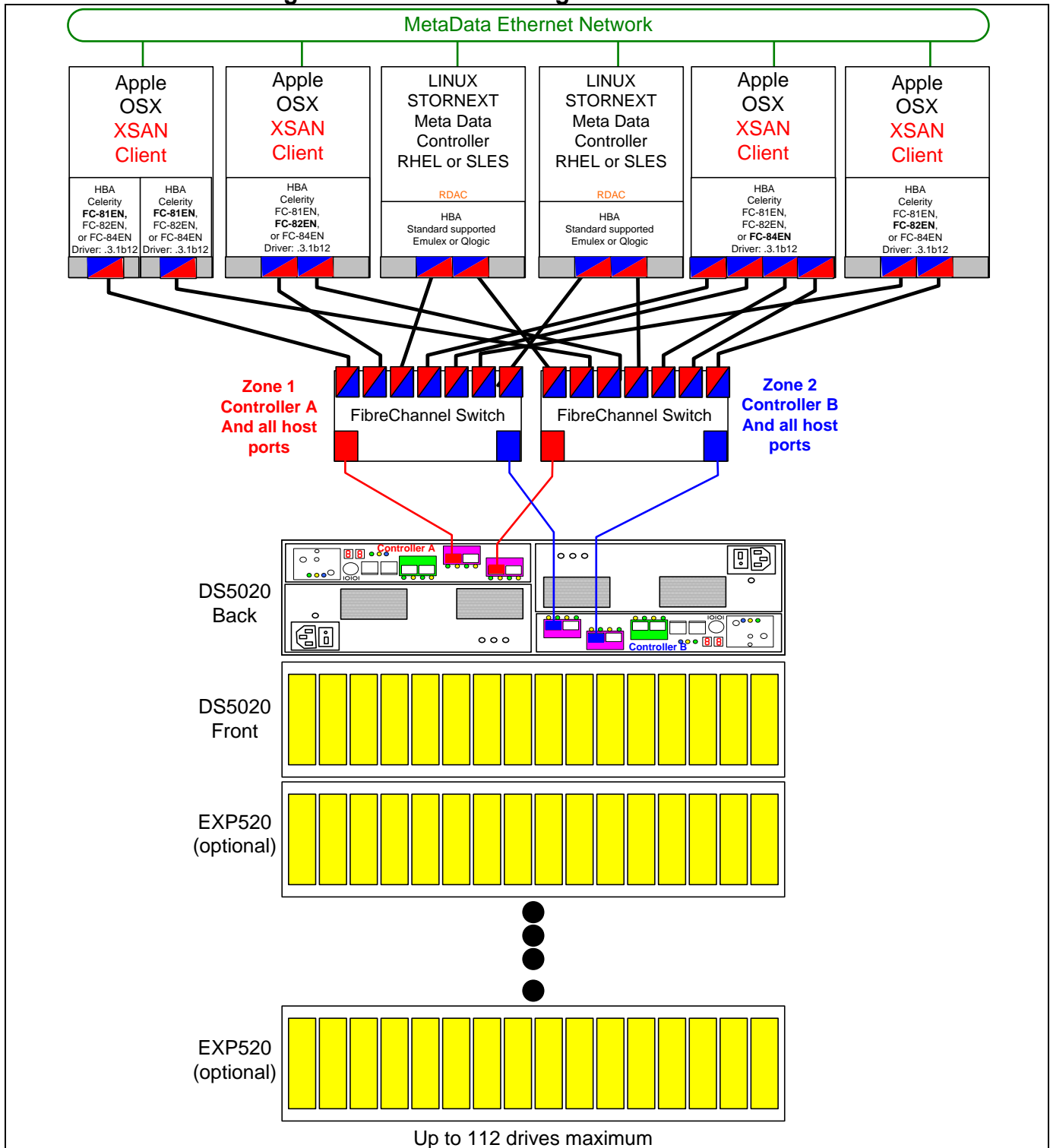
Example 3. In this example diagram Linux MDC Servers are shown, but only MAC Xsan Clients are shown. DS5020 connected to meshed SAN fabric with a minimum of two Fibre Channel switches. For DS5020 configurations, Multiple Fibre Channel switches are not required, but they are recommended. Drawing shows minimal SAN to DS5020 connections to simplify the drawing. However, up to 8 connections are supported. In this example, the Meta data Controllers each have a single dual ported HBA. Also in this example multiple HIC cards on the DS5020 are used to increase resiliency.

Mid-sized "MAC Homogenous" StorNext configuration

Example 4. In this example diagram Linux MDC Servers are shown, but only MAC Xsan Clients are shown. DS5020 connected to meshed SAN fabric with a single Fibre Channel switch. For DS5020 configurations, Multiple Fibre Channel switches are not required, but they are recommended. Drawing shows minimal SAN to DS5020 connections to simplify the drawing. However, up to 8 connections are supported. In this example, the Meta data Controllers each have a single dual ported HBA. Also in this example multiple HIC cards on the DS5020 are used to increase resiliency.

Small "MAC Homogenous" StorNext configuration

Example 5. In this example diagram Linux MDC Servers are shown, but only MAC Xsan Clients are shown. DS3400 connected to meshed SAN fabric with a single Fibre Channel switch. For DS3400 configurations, Multiple Fibre Channel switches are not required, but they are recommended. Drawing shows minimal SAN to DS3400 connections to simplify the drawing. However, up to 4 connections are supported. In this example, the Meta data Controllers each have a single dual ported HBA.

RAID 1 pairs for Meta Data and Journal are represented. Global Hot spares would be assigned as appropriate. While SAS drives would be recommended for Meta Data and Journal, SAS and/or SATA drives could be used for data as appropriate.

Mid to large -sized Heterogeneous StorNext configuration
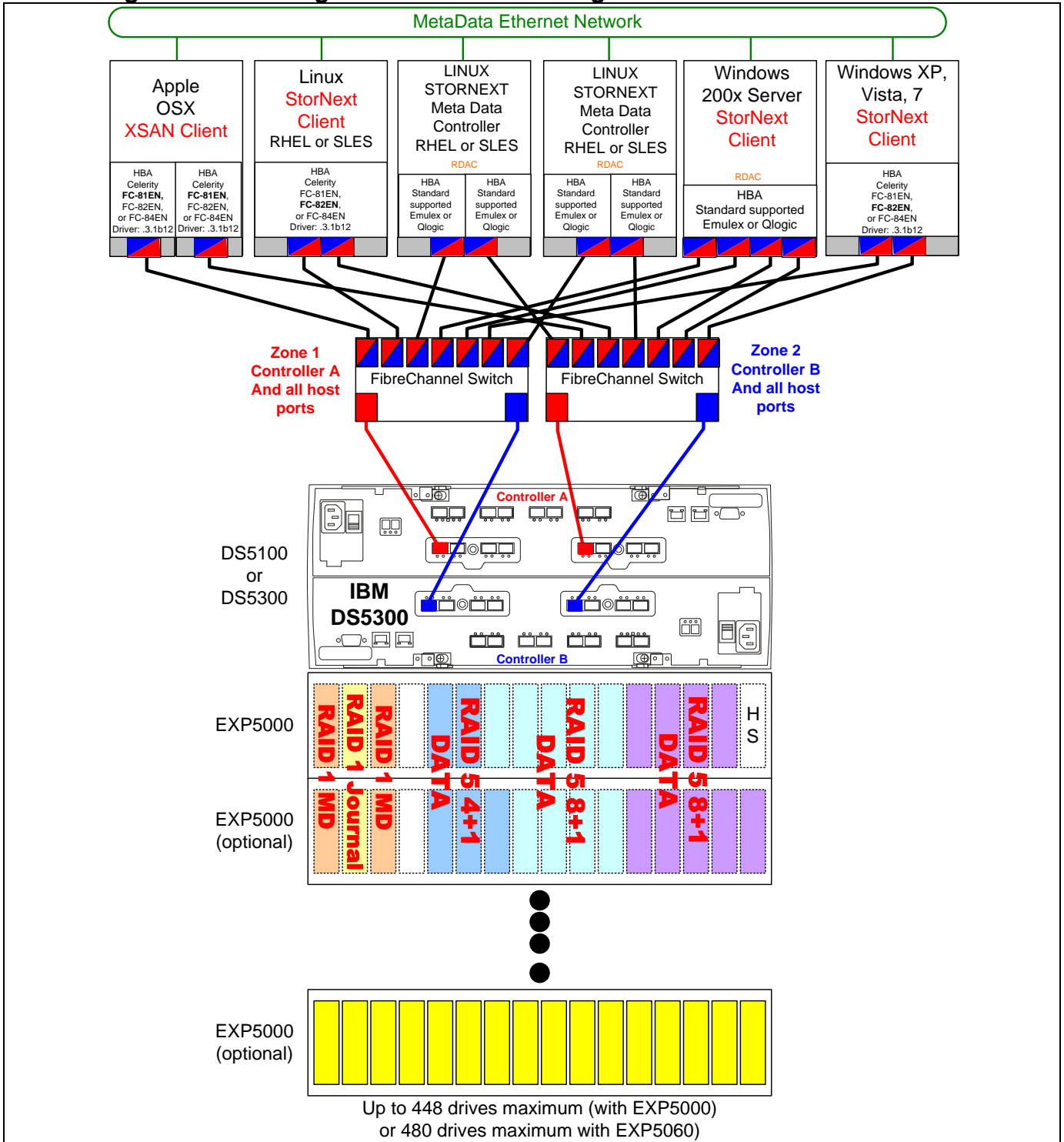
Example 6. Drawing depicts mid to large size and expandable SAN with heterogeneous client connectivity. DS5300 connected to meshed SAN fabric with a minimum of two Fibre Channel switches. For DS5100/DS5300 configurations, a minimum of two switches is recommended. Drawing shows minimal SAN to DS5300 connections to simplify the drawing. However, up to 16 connections is supported.

In this example, the Meta data Controllers each have a single dual ported HBA. Also in this example multiple HIC cards on the DS5300 are used to increase resiliency. This drawing includes RAID Group type and size suggestions. However performance tuning guidelines should be used to plan RAID group layout between trays to maximize performance.

Additional Reference Links

- IBM DS Series Portal: www.ibmdsseries.com
- ATTO Technology, Inc: www.attotech.com/solutions/ibm
- Quantum StorNext: www.stornext.com
- Apple Xsan: http://www.apple.com/xsan/

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information about the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those 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.

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US 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. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of other companies:

Apple, the Apple logo, Mac, Mac OS, and Macintosh, are trademarks of Apple in the United States, other countries, or both.

Quantum, the Quantum logo and are registered trademarks of Quantum Corporation and its affiliates in the United States, other countries, or both.

ATTO, the ATTO logo are trademarks of ATTO Technology Inc, in the United States, other countries, or both.

Microsoft, Windows Server, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

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