Xsan Revealed

MacRetreats 2005

Scott M. Neal
Senseption

Goal of Session

- Learn about SAN fundamentals
- Learn about Xsan, Apple's implementation of SAN technologies
- Let you play with an Xsan setup!

Why SAN?

Direct Attached Storage (DAS)

- VERY simple: A Storage Device (usually a hard drive) directly attached to a machine
- Uses well-defined existing protocols
  - PATA/IDE
  - SATA
  - FireWire
  - SCSI
- Only the machine that is attached to the Storage Device can access files on the device
  - Imagine if you had a hard drive plugged into more than one computer at the same time!

Network Attached Storage (NAS)

- Commonly referred to as "File Sharing": AFP, NFS, SMB/CIFS
- Manufacturers sell dedicated NAS appliances
  - These are not full-fledged servers, but devices dedicated to serving storage over a particular protocol
- Can be accessed by more than one computer at a time
- But what happens when the NAS device becomes disabled or overtaxed?

Market Forces

- Storage requirements have been doubling every 6-12 months (Source: Reuters)
  - Gigabytes seemed unheard of not too long ago
  - Terabytes are becoming the norm for many industries
    - Video
    - Fortune 500 (driven by Sarbannes-Oxley)
- Costs of scaling DAS or NAS solutions can be astronomical
  - Hardware/software upgrades
  - IT resources to maintain

Market Forces

- High-Availability is VERY important
  - Many businesses now insist on "five nines": 99.999% uptime (5 minutes downtime a YEAR)
    - one software update would kill that...
  - IT's focus has shifted from backup to recovery to business continuity
- SAN seems expensive, but compared to the losses incurred with downtime, the cost can be quite reasonable
  - If a customer loses X number of dollars per second of downtime, SAN solutions are often MUCH cheaper than that risk
  - If the customer is feeling enough PAIN, they will pay for SAN--if not, they won't (yet...)

What is SAN?
How does it solve our problems?

Storage Networking: What is it?

- A network whose focus is not general transfer of information, but specific to files and filesystem information
  - optimized for VERY high-speed, robust file transfer--a “built-in drive” for an entire network
- Accessible by more than one client at a time, unlike Direct Attached Storage (DAS)
- Not prone to failover and bandwidth choking issues like Network Attached Storage (NAS)

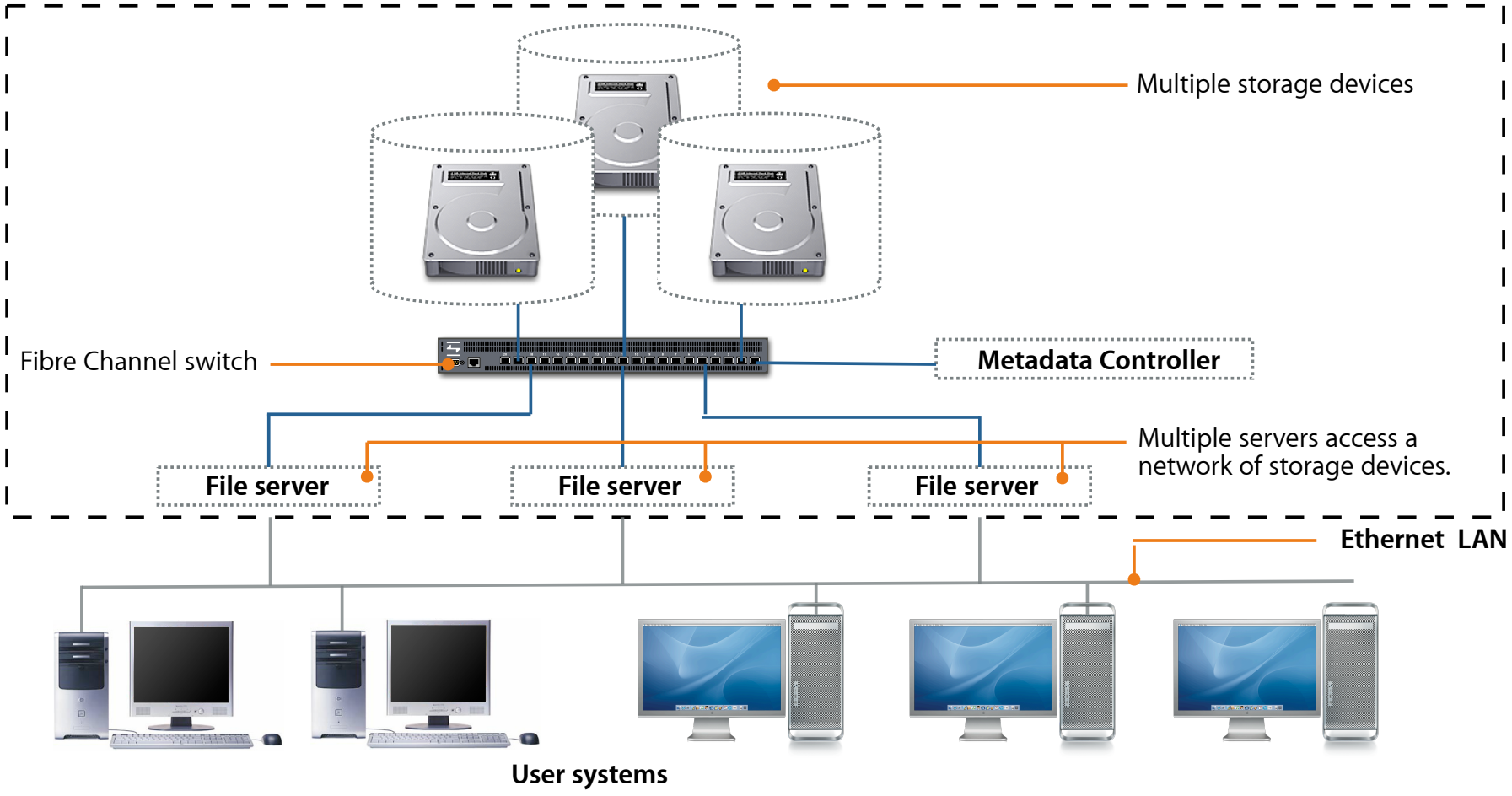
SNIA

- Storage Network Industry Association
- Industry consortium focused on standardizing nomenclature and SAN metrics
- http://www.snia.org
- Formal definition of SAN:
  - "A network whose primary purpose is the transfer of data between computer systems and among storage elements. A SAN consists of a communication infrastructure, which provides physical connections, and a management layer, which organizes the connections, storage elements, and computers systems so that data transfer is secure and robust"

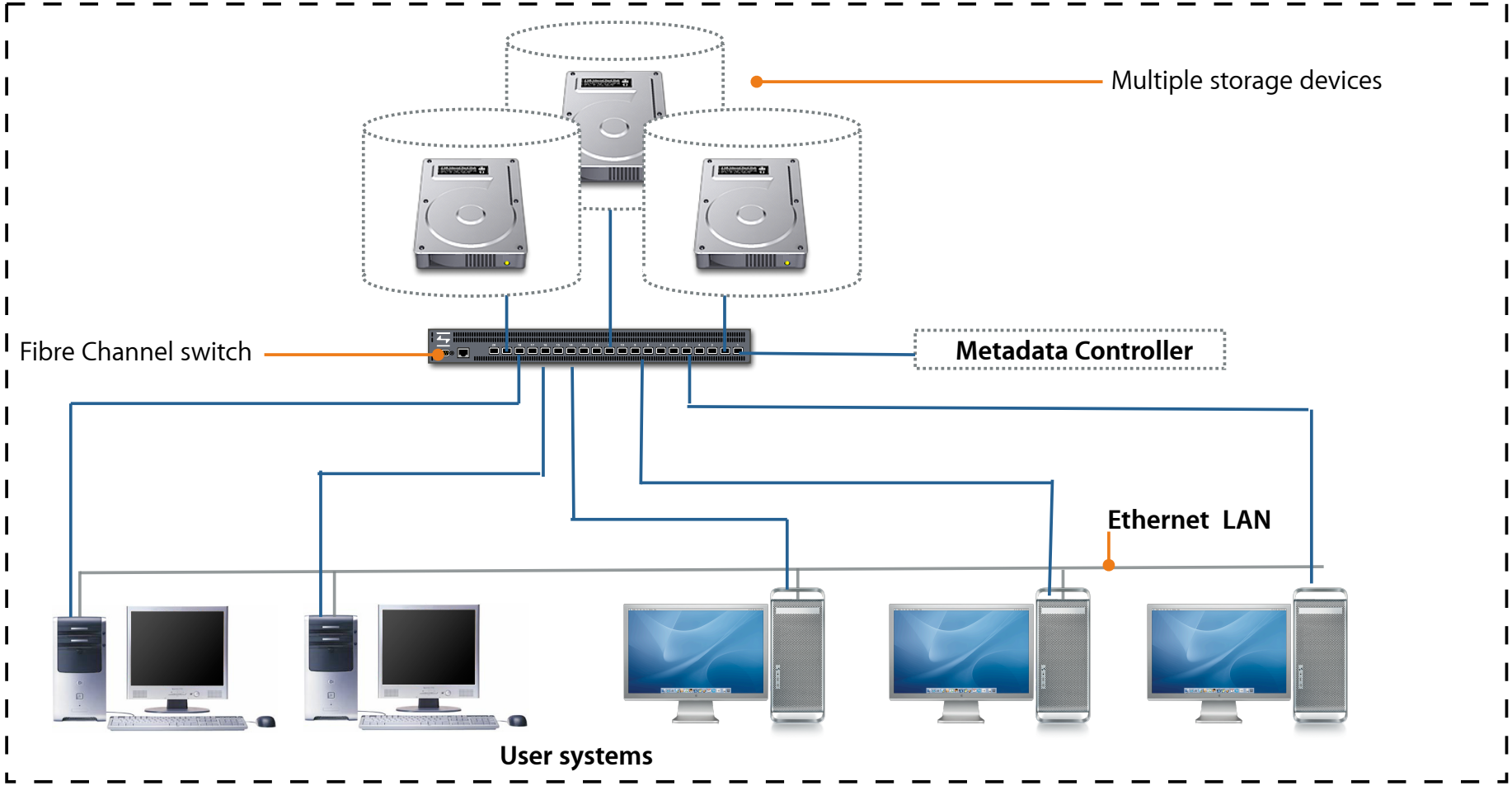
Components of a SAN

- Machines that "speak" SAN
  - Each participant in the SAN requires SAN software
    - User machines themselves
    - File Servers
- SAN Metadata Controller
  - "Brains"
  - may be more than one (failover)
- A dedicated high-speed Fibre Channel network
  - Fibre Channel Switch(es)
  - Fibre Channel Card (HBA) on each SAN host
  - Cabling

High-level SAN View: File Servers as "clients"

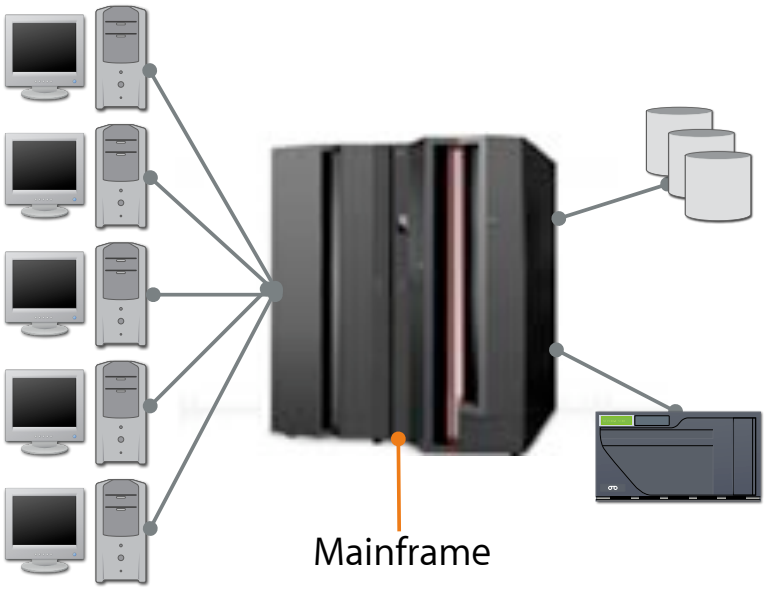


User Machines as "clients"



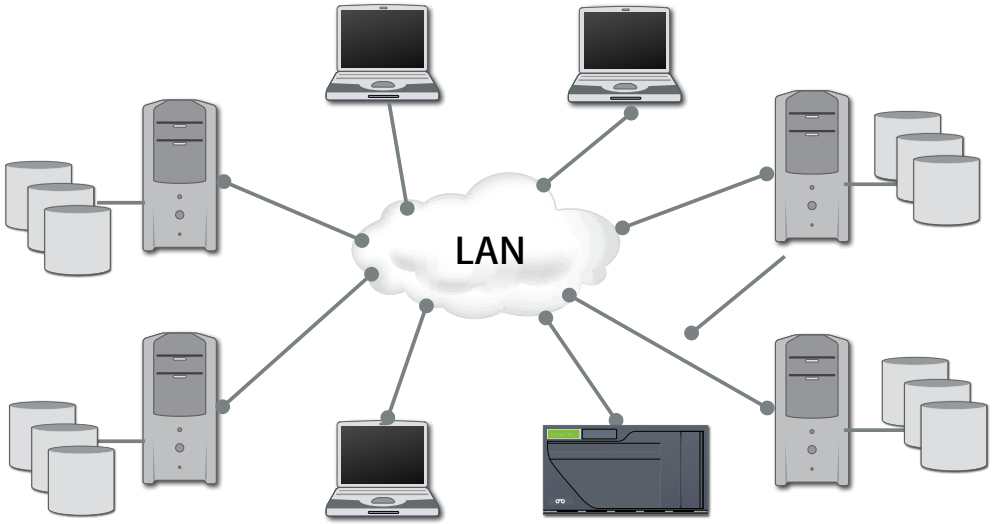
Computing Model 1: Mainframe vs. Departmental

Mainframe model



Proprietary closed system
Limited CPU scalability

Departmental computing model



Open system
Limited storage scalability

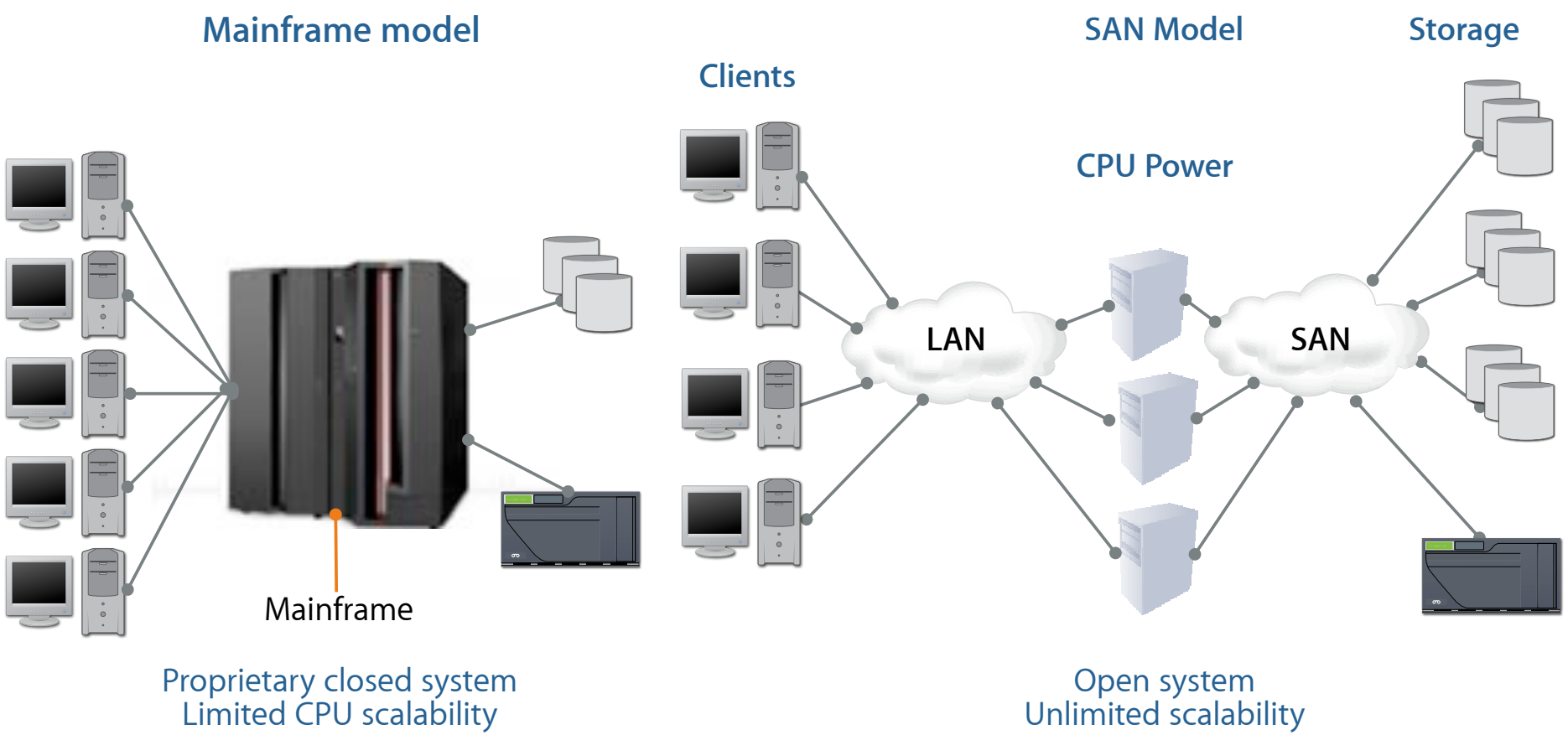
Mainframe

- Mainframe advantages:
  - centralized, sharable storage
  - easy to backup
- Mainframe disadvantages:
  - CPU scalability
  - proprietary technologies

Departmental

- Departmental Advantages:
  - “confederacy”--each computer powerful
- Departmental disadvantages
  - Storage scalability
  - difficult to share files across LAN
  - difficult to easily, automatically backup
  - backups over LAN (slows network down)

Computing Model 2: Mainframe vs. SAN



SAN: Best of All

- Advantages of Mainframe
  - centralized storage
  - easy to backup
- Advantages of Departmental
  - Each computer powerful and autonomous, unlike a standard mainframe "dumb terminal"
- Unique to SAN
  - CPU power and storage independently (and seemingly infinitely) scalable
  - Platform-agnostic: share data across heterogeneous platforms (as long as they all speak SAN)
  - Redundancy: If one SAN file server is down, client machine can roll over to another one

What's the catch?

- So what's the Disadvantage?--"life is a compromise", right...?
  - We need to learn about and use SAN technology
  - Cost is off-putting to many customers (penny-wise, pound-foolish)
  - Lack of industry standardization and metrics
    - SNIA is helping to solve that issue

SAN Markets

Audio/Video

- Usually the first thing people think of when thinking about SAN
- Benefits:
  - Editing
    - Fast access to clips
    - Centralized storage: clips don't get out-of-date
  - Broadcast
    - High bandwidth, low latency
    - Artifacting is NOT acceptable for broadcast

Backup/Restoration

- Remember, the goal for backing up is really RESTORATION
- Benefits
  - LAN-free and serverless backup
  - Reduced backup and restoration times
  - Continuous operation (don't have to take servers down to do backup)
    - copy-on-write snapshots
    - split mirrors snapshots
- Think:
  - Security
  - Sarbannes-Oxley

Database

- Databases ALWAYS benefit from improved access and data integrity
- Benefits:
  - Continuous uptime
  - scalability
    - CPU
    - Storage

Imaging

- Sometimes known as “Service Bureaus” (sounds so exotic!)
- Benefits
  - Fast transfer of HUGE files
  - Heterogeneous sharing

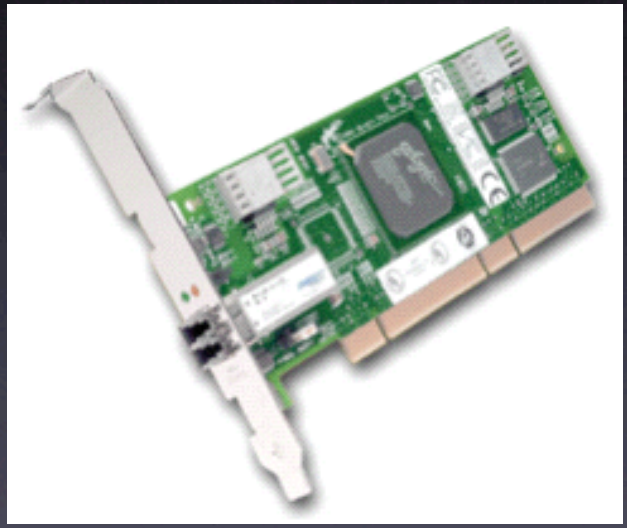
Content Distribution

- Web hosting, Digital Broadcasting, Video-on-Demand
- Benefits
  - Guaranteed Performance (few, if any, latency issues)
  - Continuous uptime (highly available)

Fibre Channel Networking

Host Bust Adaptor (HBA)

- Provides interface from computer's bus (PCI, PCI-X, S-Bus) to Fibre Channel Network
- 1, 2, or 4Gb/sec
- Many differentiators from different vendors



HBA vs. NIC

- HBAs have on-board "smarts" that offload the host CPU
  - 50-80 MB/s over Gig Ethernet requires more than 80% of a server's CPU capacity (1GHz Intel P-III)
  - 95-100 MB/s over Fibre Channel requires less than 10% (same 1 GHz Intel P-III)
  - 2Gb/s HBA provides peak of 200 MB/sec

Apple's Fibre Channel HBA

- CHEAP! ($499)
- Includes 2 SFP-to-SFP cables (Fibre Channel cables are VERY expensive)

Fibre Channel Cable Types

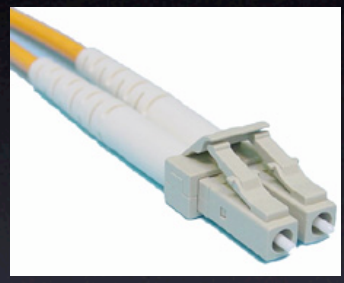
- Copper
  - lower cost
  - short distances (without repeaters)
    - up to 33m
- Optical
  - higher cost
  - long distances (even without repeaters)
    - up to 500m with 50 micron cables
    - up to 120Km (!) using CWDM
- Both copper and optical are equivalent speed-wise for 1, 2, & 4 Gb/s
  - 10Gb/s Fibre Channel requires optical

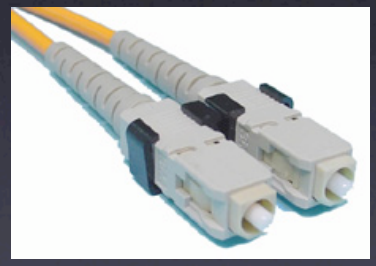
Copper Connectors

- SFP
- HSSDC
- DB9

Optical Connectors

- LC (2Gb)
  - "Lucent Connector"
- Duplex SC (1Gb)
- Transceivers exist
  - copper->optical
  - one optical/copper type to another





Fibre Channel Hubs & Switches

- Hubs
  - Passive wiring
  - Arbitrated Loop Topology (cheaper, not as scalable)
  - Think Ethernet hub vs. switch: cheaper, but not as powerful
- Switches
  - Active wiring ("brains")
  - Switched Fabric Topology (VERY scalable)
- See www.apple.com/xsan/compatibility.html for a list of Apple-certified hubs and switches

FC Bridges and Routers

- Connect Fibre Channel network to non-Fibre Channel Devices
- Typically used to utilize legacy equipment (Tape Drives, etc.) that are not FC (usually SCSI)
- Terminology interchangable/confusing here
  - Sometimes referred to as “Brouters”

WAN Devices

- Interconnect Fibre Channel networks over "phone lines"
- Different "wire" technologies
  - DWDM (Dense Wavelength Division Multiplexing)
  - SONET (ANSI)
  - ATM (Asynchronous Transfer Mode)
- Different protocols
  - Fibre Channel over IP
  - iSCSI

Storage Devices

JBOD

- "Just a Bunch of Disks" (no, I'm NOT making this up!)
  - Utilize either SCSI or SATA drives, chained together internally
  - Fibre Channel connection on the outside (but in simpler, less scalable Arbitrated Loop Topology)
- No brains, just disks connected together
  - RAID technology must be done in Software

RAID

- Redundant Array of Independent Disks
  - A group of SCSI, PATA/SATA, or Fibre Channel disks inside
  - A Fibre Channel connection on the outside
- Dedicated RAID controller provides high level of "brains"
  - Raid types 0, 1, 5 most typical

Tape Devices

- Historically SCSI (hence need for Brouters mentioned earlier)
- Can be quite powerful (and also quite expen$ive)
- www.apple.com/xsan/compatibility.html shows Apple-certified Tape devices

SAN Architecture

SAN is taking the virtual place of DAS

- Recall that DAS is basically a fancy term for plugging a hard drive into a computer
  - We want SAN to be a "miracle network" that is as fast as an internal hard drive, but usable by more than one host
- Protocols for DAS include
  - PATA/IDE
  - SATA
  - FireWire
  - USB/USB2
  - SCSI

What do we mean by DAS "protocols"?

- There are two aspects to DAS connectivity
  - Physical hardware connection
    - Electrons flowing through wires (analog mindset)
  - High-level protocol
    - How blocks get transferred (digital mindset)
    - SCSI and FireWire have their own brains
    - PATA/IDE, USB/USB2 depend on brains of host CPU
    - SATA somewhere in the middle...
- SAN communication is based on a very-well known pre-existing protocol

SCSI !

- Yes, SAN communication uses the SCSI protocol
- Don't dredge up those terminators though, there IS good news:
  - SAN is based on the SCSI Block Transfer Protocol, but NOT the SCSI Physical Hardware Connection--Fibre Channel takes care of that
  - SCSI provides intelligence, so devices can communicate directly with each other w/o a central controller bottleneck
- As previously seen, storage devices are RAIDs or JBODs, which may be PATA/IDE, SATA, or SCSI based
  - This does NOT affect the SAN protocol, it IS SCSI

Metadata Controller

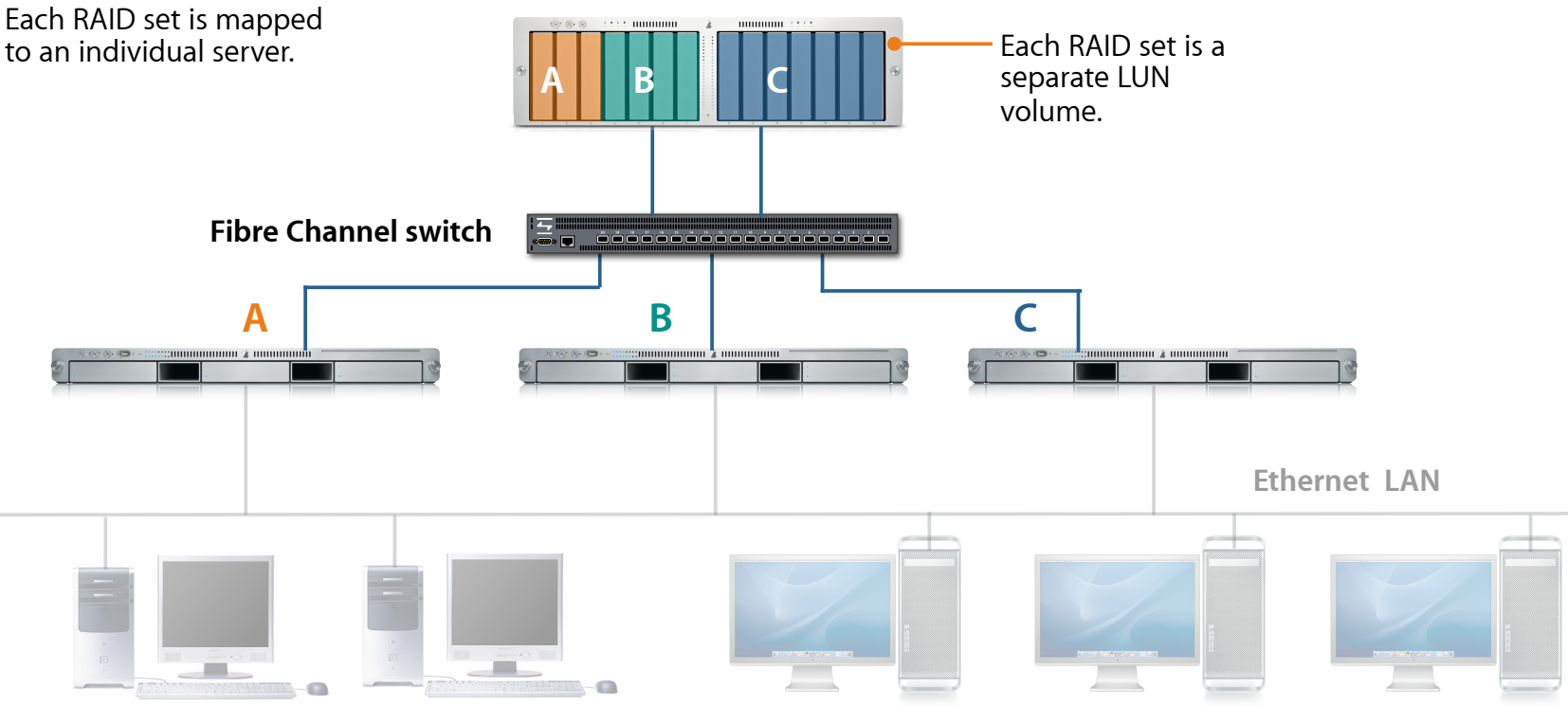
- A host dedicated to keeping track of the file system
- Cognizant of all RAID/JBOD participants
  - You control, when setting up the SAN, how each participant plays a role
- SAN Clients request files from Metadata Controller, Metadata controller responds with directions on where to look
  - Metadata controller does NOT fetch the data itself--more like providing a map so the client can get the data it needs directly
- Should have redundant MetaData Controllers
  - If your MDC goes down, you are, um...

SAN

- Any Questions so far?

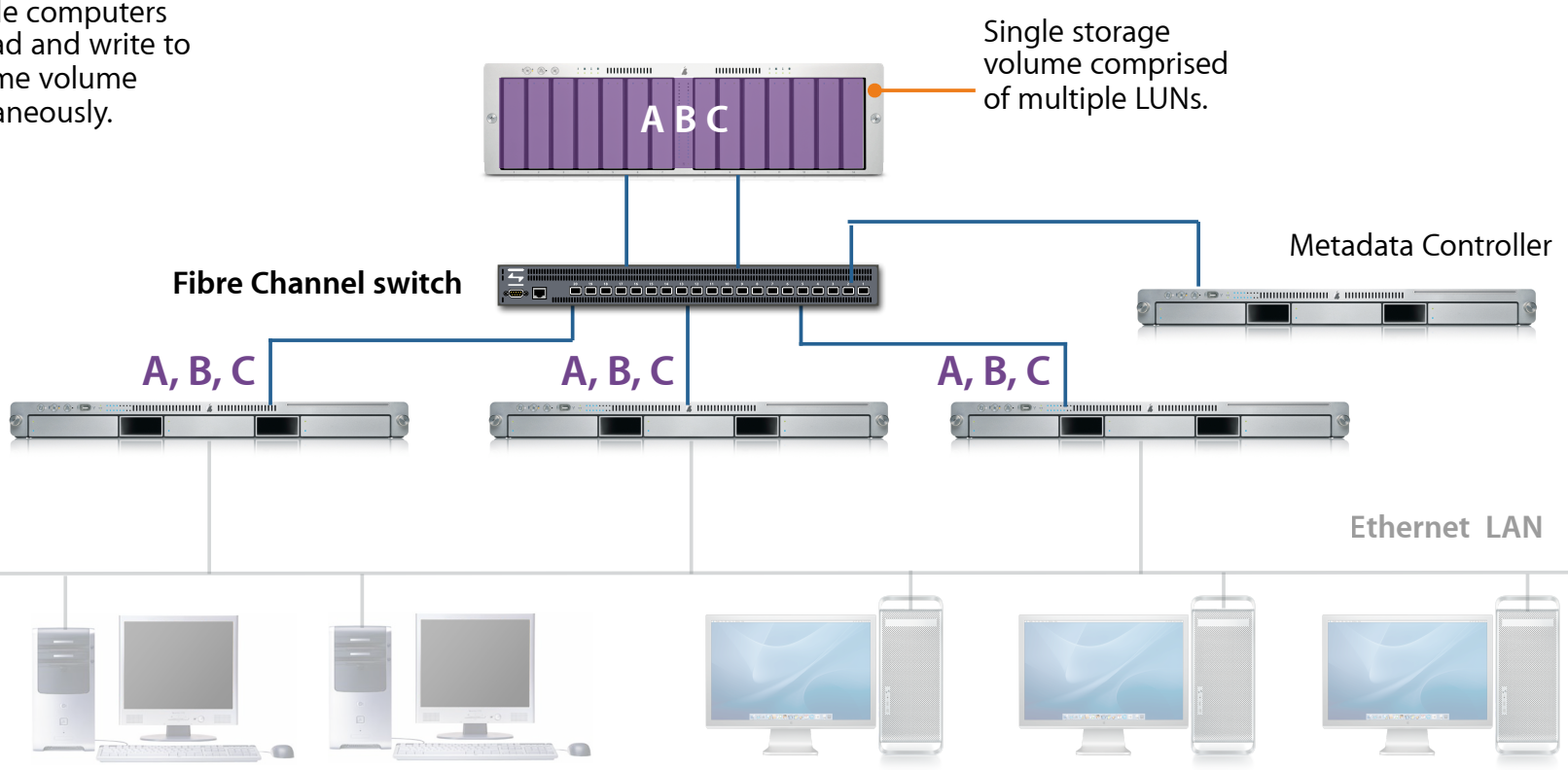
Xsan Storage Virtualization

Before SAN: Xserve Raid Deployment



With Xsan: Xserve Raid Deployment

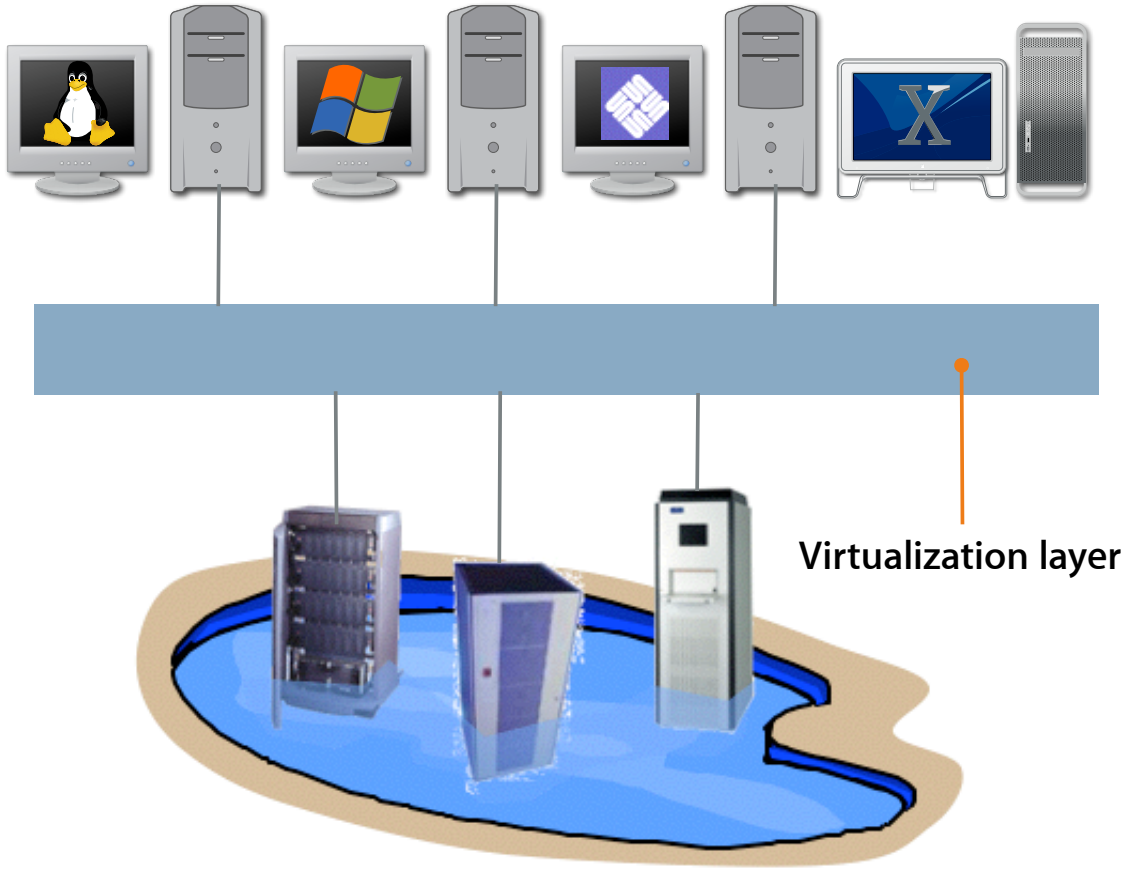
Multiple computers can read and write to the same volume simultaneously.



Organizing Storage Devices with SAN

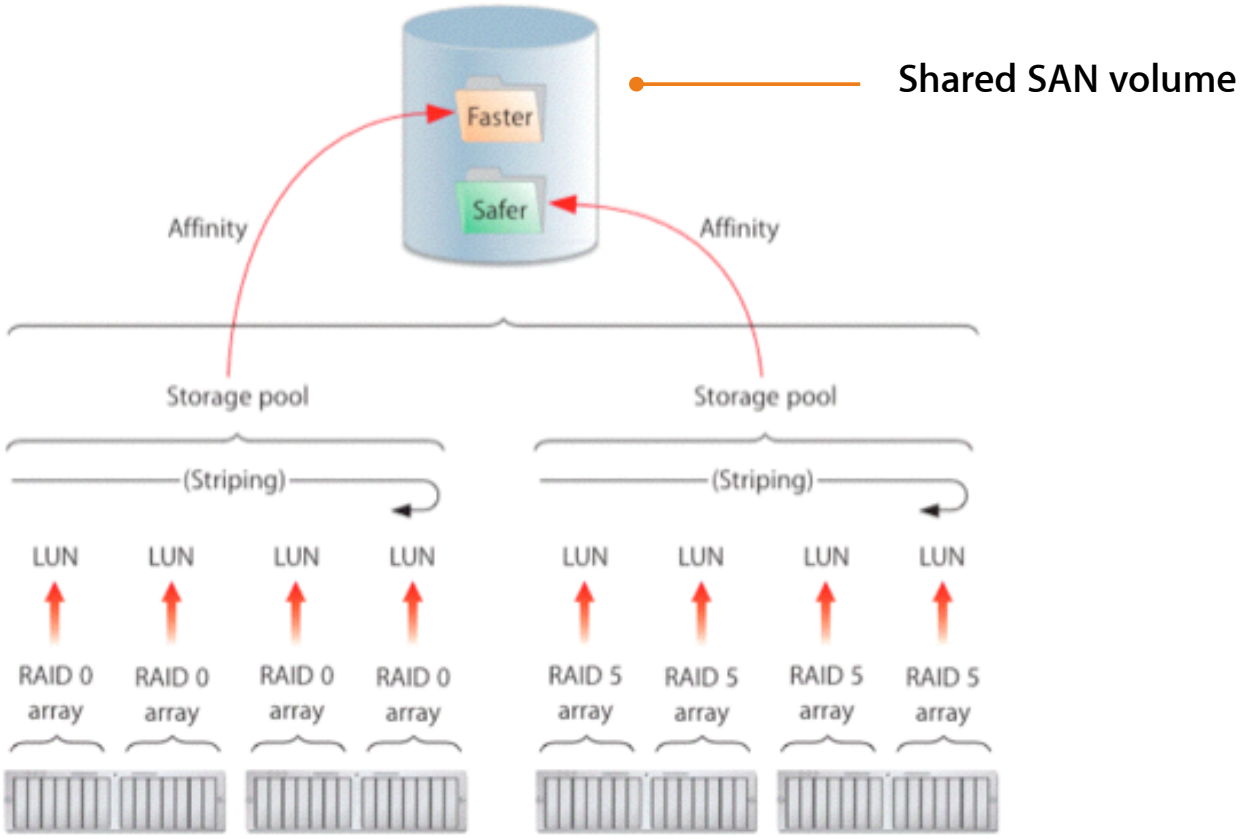
- Each RAID is no longer limited to specific provisioning to a server
- One or more LUNs from a RAID can participate in Storage Pools
  - The Storage Pool becomes the level of storage granularity
- This process is called Storage Virtualization

Storage Virtualization

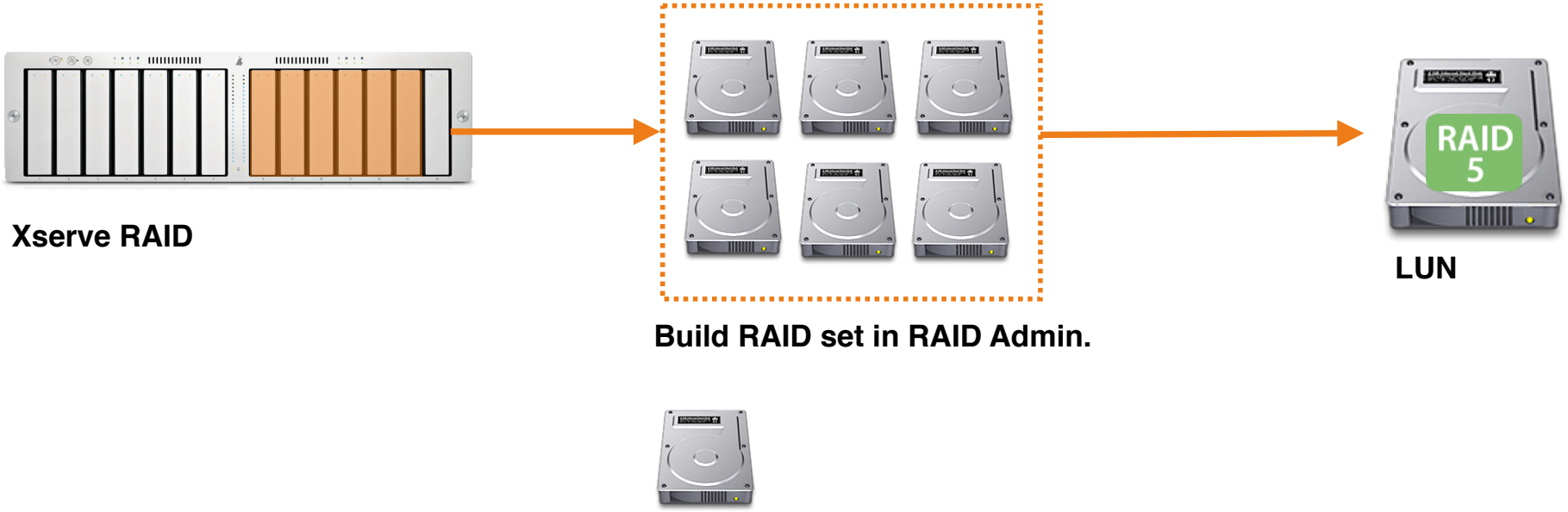


Xsan is a Heterogeneous solution based on ADIC's StorNext, which provides SAN support for multiple platforms

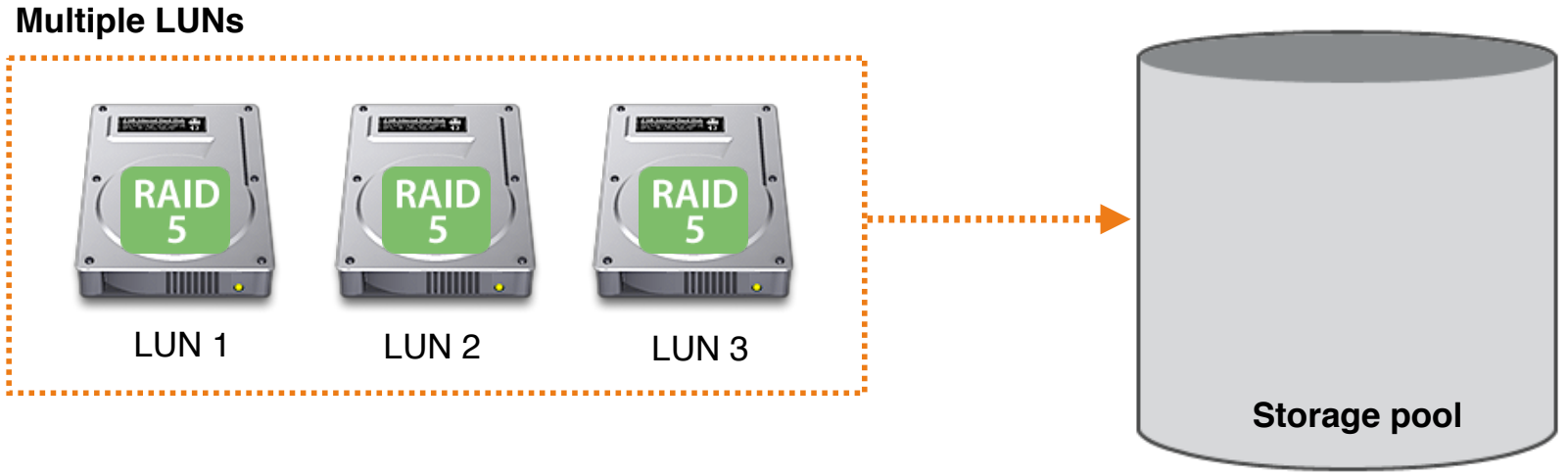
Xsan
Storage Virtualization



Creating an Xsan Volume: Step 1



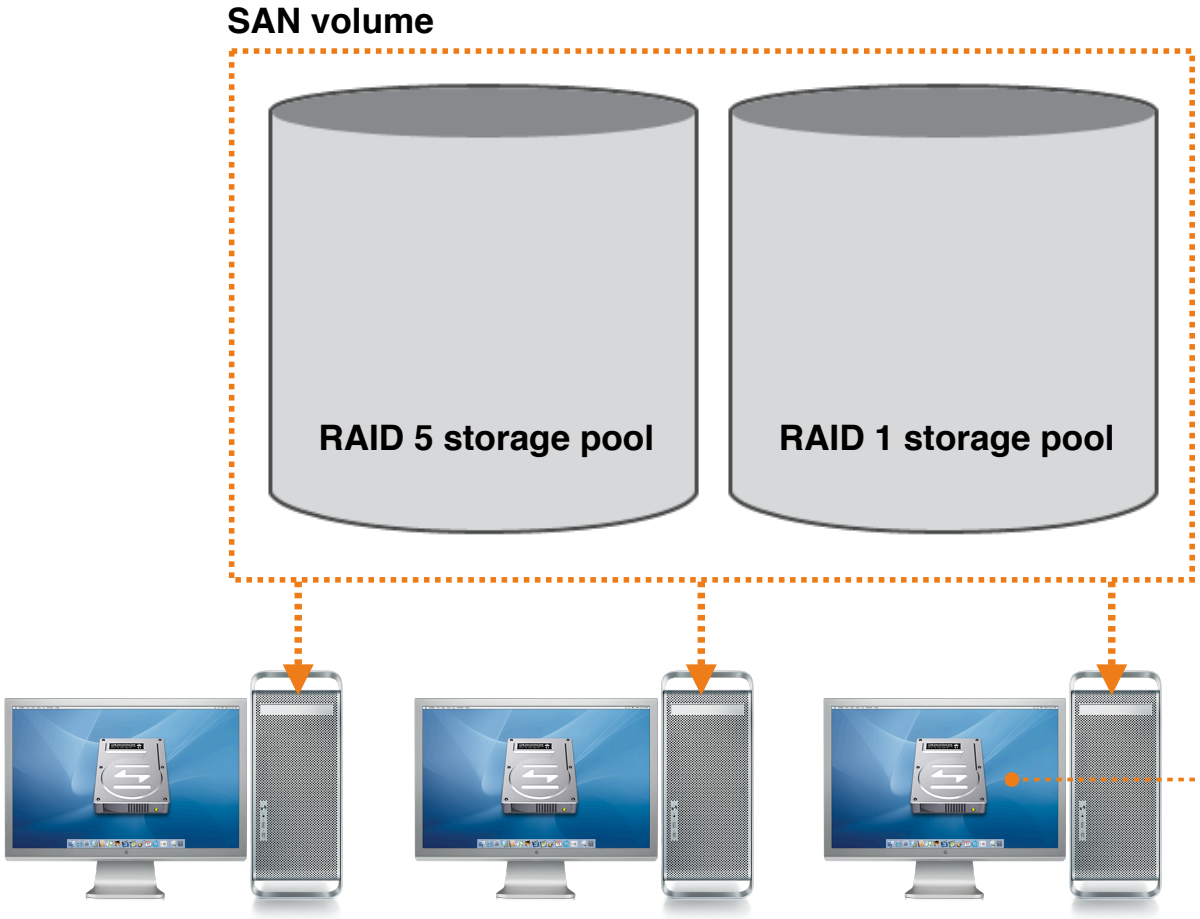
Creating an Xsan Volume: Step 2



Groups LUNs into Storage Pools using Xsan Admin

Storage pools enable grouping of LUNs with the same size and data protection properties. Each LUN is striped across the Storage Pool

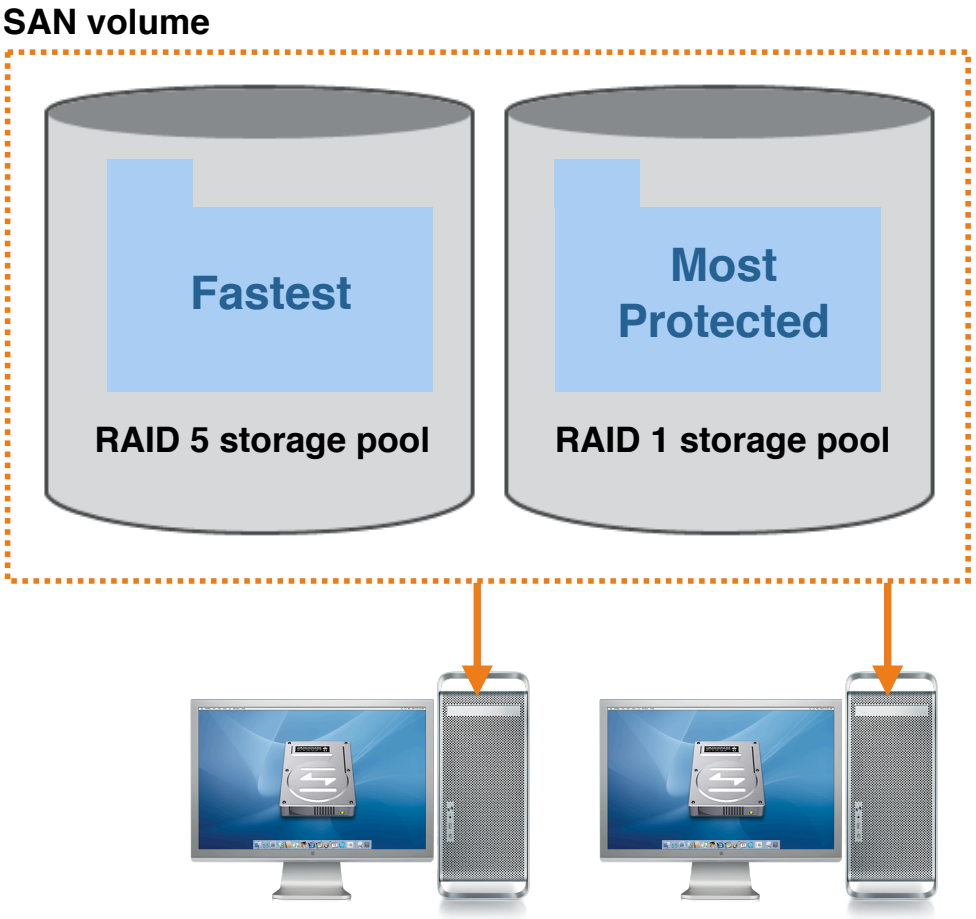
Creating an Xsan Volume: Step 3



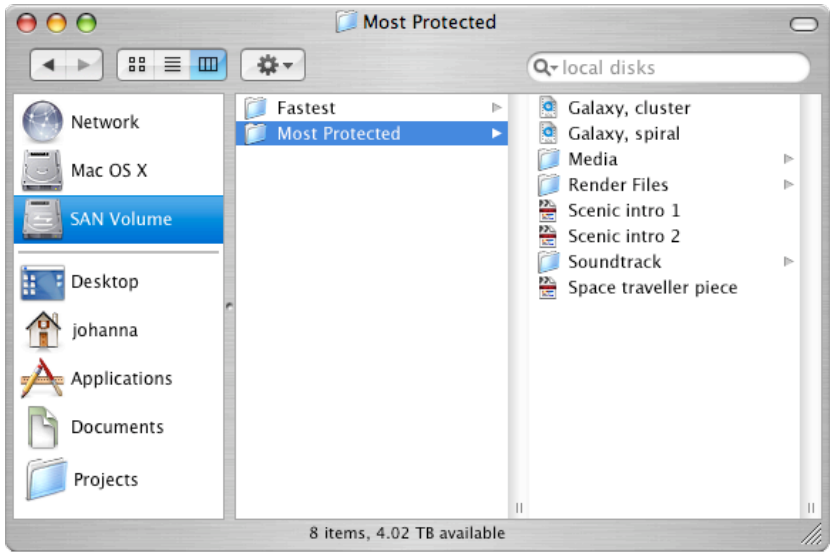
Multiple Storage Pools are combined to create one volume using Xsan Admin

The volume appears as a single hard drive to each SAN client.

Creating an Xsan Volume: Step 4



Affinities are created (optionally) using Xsan Admin to ensure that a named folder has an "affinity" to a particular storage pool



Demo: Creating an Xsan Volume

- Any Questions?

Synopsis

- Xsan is a VERY powerful SAN implementation
- It is also very cost-effective (especially when compared to NOT having it)

Resources

- http://www.apple.com/xsan
- Xsan Admin Guide (downloadable from Apple's OS X Server documentation site)
- Peachpit Xsan Quick-Reference Guide
- Apple training 3-day Xsan course
  - http://train.apple.com

Thank You!
Xsan Revealed

MacRetreats 2005

Scott M. Neal
Senseption