Introduction To Storage Area Networks

HP Interex West Symposium
Session # 3147
March 2004

Jim Gursha

High Performance System Solutions, Inc.

1735 York Avenue, 32H

New York, NY 10128

212-831-0291/917-359-2087

jimgursha@sandisks.com

❖Topics Of Discussion

➢ **Storage Area Networks (SANs) Introduction**

- ❖ Basics of Storage Area Networks - What they are comprised of (hardware components) and what type of layered software products are available to make complete SAN solutions
- ❖ Storage Concepts - A discussion of the differences between direct attached storage (DAS), network-attached storage (NAS), and SAN based storage. Storage Virtualization concepts are introduced and contrasted with the different storage paradigms available today.
- ❖ Topics In NAS/SAN Convergence
- ❖ When is a SAN an appropriate solution?
- ❖ SAN Cost Justifications
- ❖ Hardware/Software Cost Savings
- ❖ Storage/System Management Personnel Savings
- ❖ Business Continuity and Disaster Recovery Aspects

Storage Area Networks

➢A SAN is a collection of hosts, and storage…

➢However, there is much more to a SAN…. *increased benefits!*

SAN Architecture
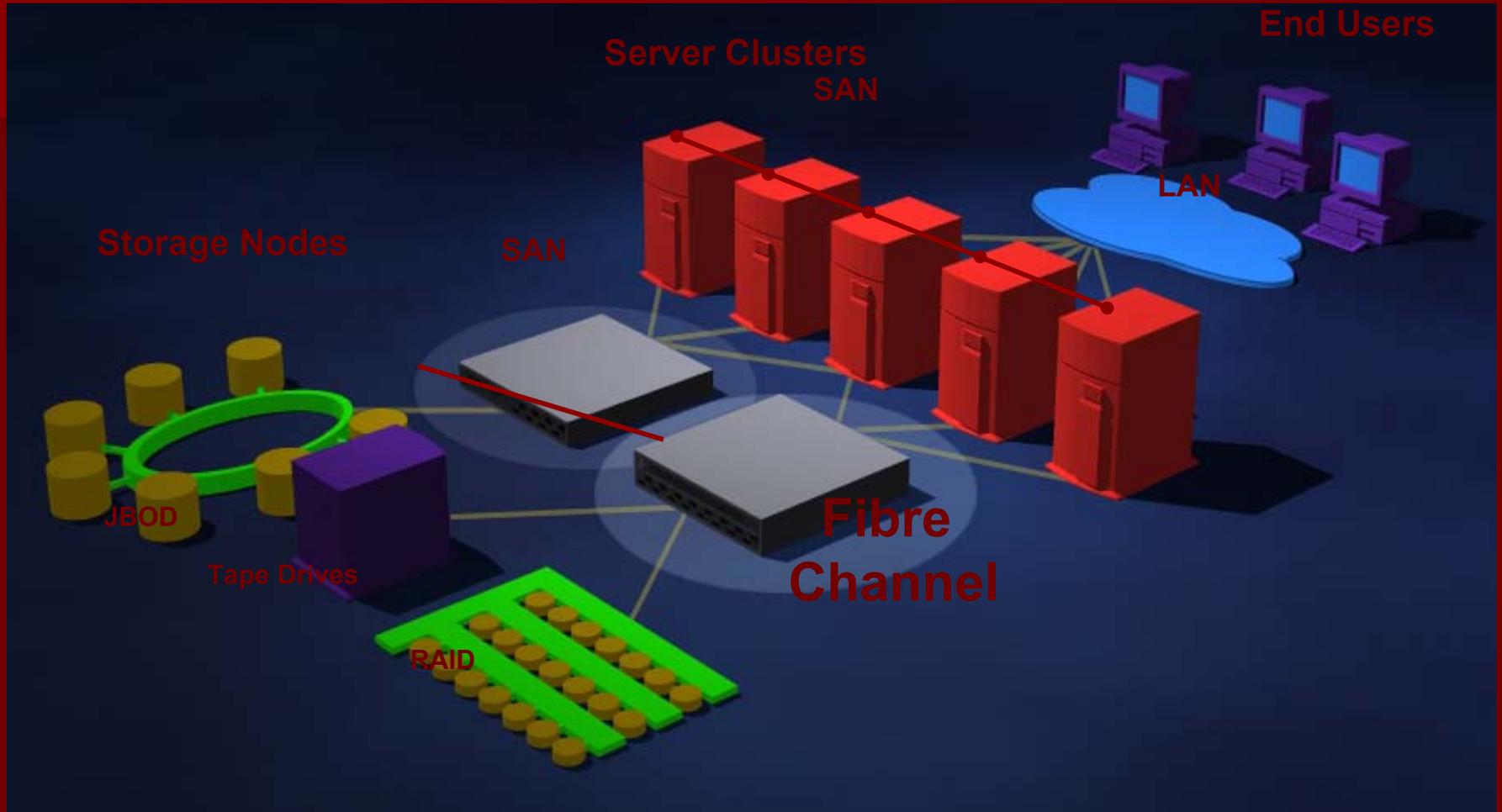
Storage Area Networks

SAN Hardware, Software, and Value Proposition

Host Bus Adapters

- Understanding Individual Component Utilization Is A Necessary Part of SAN Architecture and Implementation.

- Selecting the Right Component Will Lower Overall SAN Costs.

Host Bus Adapters

- Connect the Server to the SAN.
- Alleviate the Server From Some I/O Processing.
- Typically, Assist in the Execution of Parts of Communications Protocol.
- Compatibility Across HBA's.

Emulex HBA's

HBA Selection Guide

Bus Type	PCI Express	PCI-X/PCI	cPCI	SBus	PCI
Bus Speed	x4	133/100/66 MHz	66/33MHz	25MHz	66/33 MHz
Link Speed[1]	2Gb/s	2Gb/s	2Gb/s	2Gb/s	2Gb/s
Midrange[2]		LP1050 LP982			LP952L
Enterprise	LP10000Ex	LP10000 LP9802	LP90002C[3]	LP90002S[3]	LP90002L
Dual Port	LP10000Ex	LP10000DC LP1050DC[2] LP9802DC LP9402DC			LP90002DC

Emulex LP1000 EX

- Single and Dual-channel PCI Express to 2Gb/s Fibre Channel models
- x4 PCI Express Connector, standard
- 4-lane negotiation
- Driver Compatibility Across Emulex Product Line
- Upgradeable Firmware Based architecture
- Centralized HBA management suite, HBAnyware™
- Dynamic load balancing and high availability with MultiPulse™

Emulex HBA's

- **Dual Channel (LP9402DC)**
  - High performance through the use of two Emulex Centaur ASICs,
  - Two 266MIPS onboard processors,high speed buffer memory.
  - Automatic speed negotiation capability which allows complete compatibility between 1GBS and 2 GBS.
  - Switched fabric support,full-duplex data transfers.
  - High data integrity features,support for all Fibre Channel topologies
  - Dual channel HBA. Channels deliver up to 800MB/s link bandwidth

Emulex LP 10000

- Enterprise–class HBAs ideal for large or mixed OS SAN environments
- Single– and dual–channel PCI-X to 2Gb/s Fibre Channel models
- Full fabric support with automatic topology and speed adaptation
- cHBA API support (FC-MI) enables simplified SAN management
- Onboard hardware context cache for superior fabric support and high transaction performance
- Support for use of multiple concurrent protocols (SCSI & IP)
- FC service class 2 and 3 support
- FC-Tape (FCP-2) device support
- Simultaneous on card support for remote boot in x86, Itanium (BootBIOS and EFI), and OpenBoot in Unix environments
- End–to–end parity protection for high data integrity
- Large on card data buffer supports over 50km cabling at full 2Gb/s bandwidth

EMULEX HBA's

- Single Channel (LP9802)
  - -Full duplex 2Gb/s Fibre Channel delivering up to 400MB/s
  - Automatic speed negotiation
  - Automatic topology detection
  - Full fabric support using F_Port and FL_port connections
  - Onboard hardware context cache for superior fabric performance
  - Support for multiple concurrent protocols (SCSI and IP)
  - Full support for both FC service class 2 and 3
  - Full fabric boot support in x86 and SPARC environments to multiple LUNs
  - Support for FC-Tape (FCP-2) devices
  - 66/100/133 MHz PCI-X 1.0a and PCI 2.2 compatibility
  - End-to-end parity protection for high data integrity
  - Buffered data architecture to support over 50km cabling at full 2Gb/s bandwidth
  - Robust suite of software supporting Windows Server 2003, Windows 2000, Windows NT, HP-UX, Linux, NetWare and Solaris
  - Optical small form factor (LC) interface

Emulex GN9000 SI (ISCSI HB A)

- multiple OS/platform support: Windows.NET, Windows 2000 and Linux
- 32/64 bit 33/66MHz PCI v2.2 host bus adapter
- 1Gb/s Ethernet link speed, full duplex
- SCSI architectural model (SAM-2) and SCSI-3 command sets
- SCSI block level data encapsulated in IP (iSCSI)
- supports initiator and target mode
- support for standard TCP/IP-reliable data delivery (TCP), IP routing
- remote data access from Fibre Channel SANs over Ethernet/IP networks through storage routers

FC Hubs and Switches

- **Hubs**
  - 7 port and 12 port models
  - Much lower cost than switches
  - Used in smaller SANs
- **Switches**
  - 8, 16, 64 and 128 port models
  - Better isolation than hubs
  - Switches are more $ than hubs.
  - Scalable to large SANS
  - Cascading, Zoning, Quickloop
  - Higher MTBF, easier repair
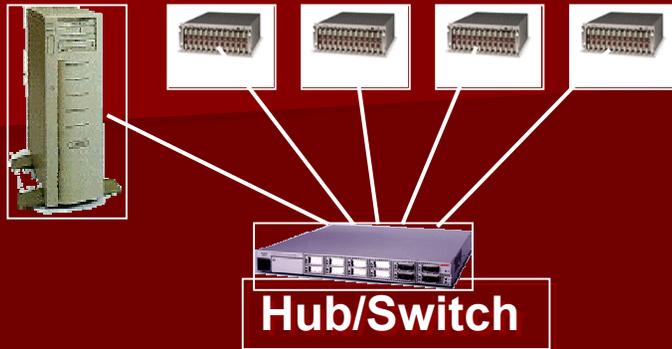- **GBICs (GigaBaud Interface Cards)**
  - Short and long wave variants
  - LW to 100Km
  - SW to 500m

# FC Implementation

Hub/Switch may look the same!
- Electrical Isolation (hot connect)
- Radial Topology
- Hot replacement of GBICs

Hub/Switch

Hubs implement FC-AL
- Logically equivalent to a single path loop
- Bandwidth is 100 MB/Sec
- Nodes see all data

Switches implement FC Fabric
- Switches act like network routers
- Multiple communication paths
- Bandwidth is aggregate of paths
- Switches provide logical isolation
- Significantly better performance
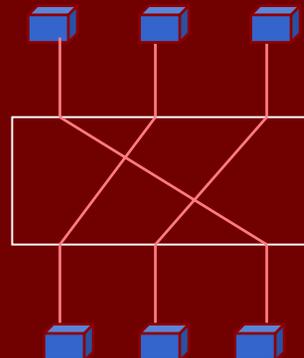
** FC Switches provide non-blocking Dynamic Paths **

Fabrics

- **Scalability means more than adding ports**
  - Supporting data growth while sustaining infrastructure performance, availability, and management capability

- **Enable scalability**
  - Design fabric that supports key environment attributes
    - Examples: Data Store Size and location, Backup, Distance, Applications

- **Fabric design drives product selection**
  - Switched Infrastructure
    - Building block size
    - Availability
    - Link performance
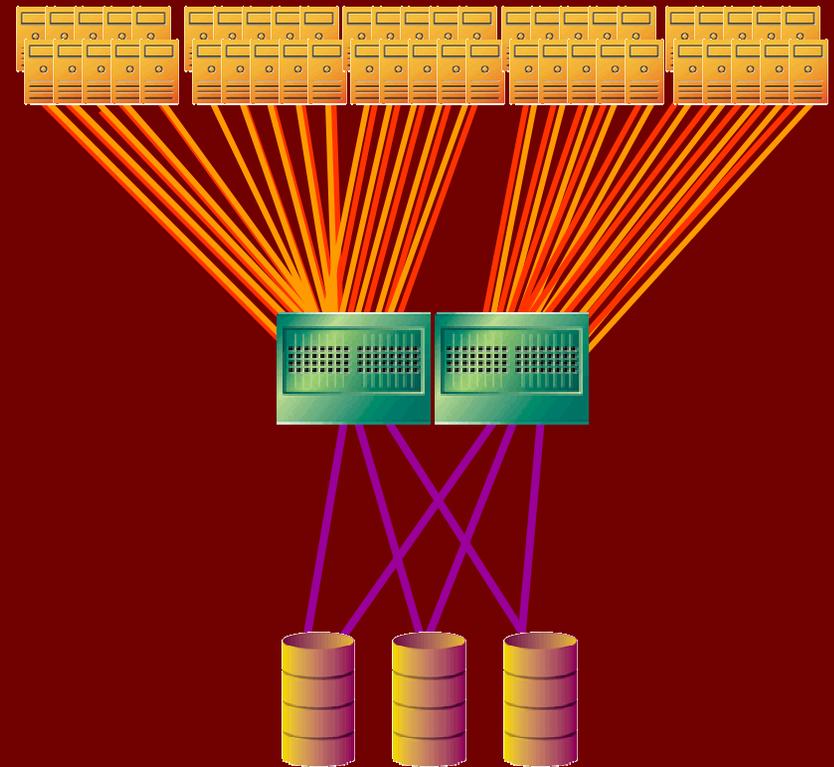 - Management
    - Device to Fabric-wide

Large Fabric Design

Switch Fabric
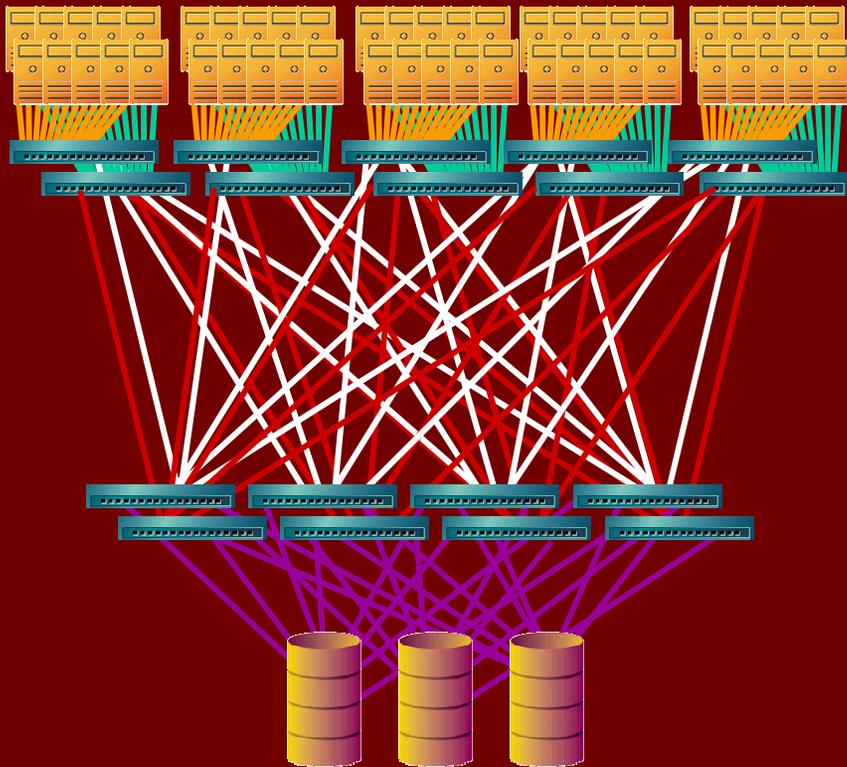
Director Fabric

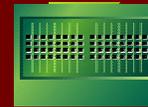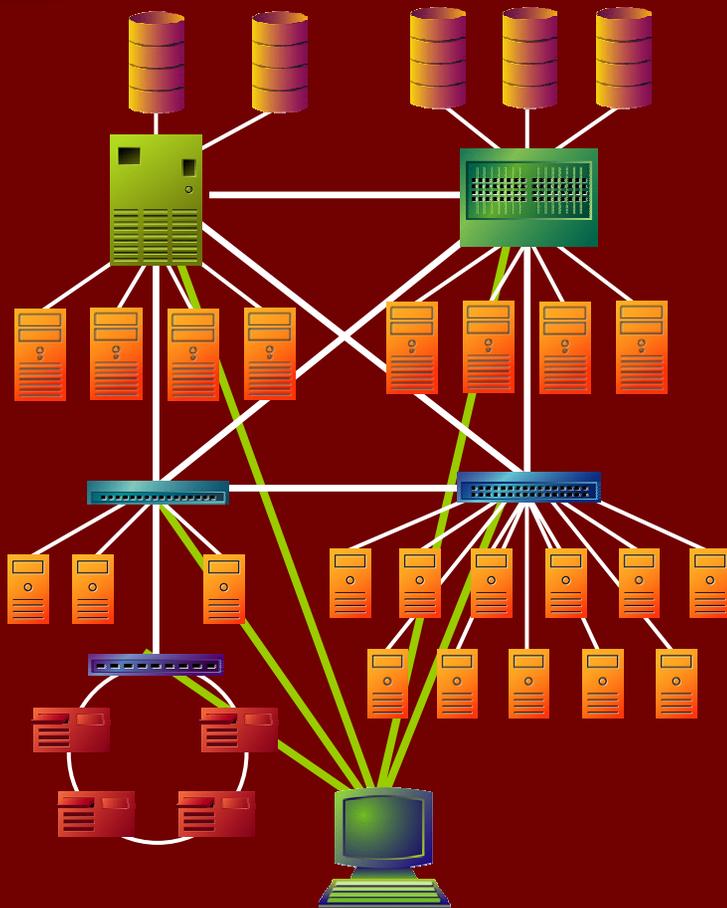**Enterprise Storage**

Enterprise Storage

Fabric Switches vs Directors

	Switches	Directors
Availability (Uptime)	99.9%-99.99%	99.999%
% Throughput affected by a failure	Up to 50%	0%
Management Complexity	18 Devices	2 Devices
Service Down Time	Replacement = 1 hour (best case)	0 minutes
Cost (estimate)	$567,000	$491,000

Fabric Architecture Design

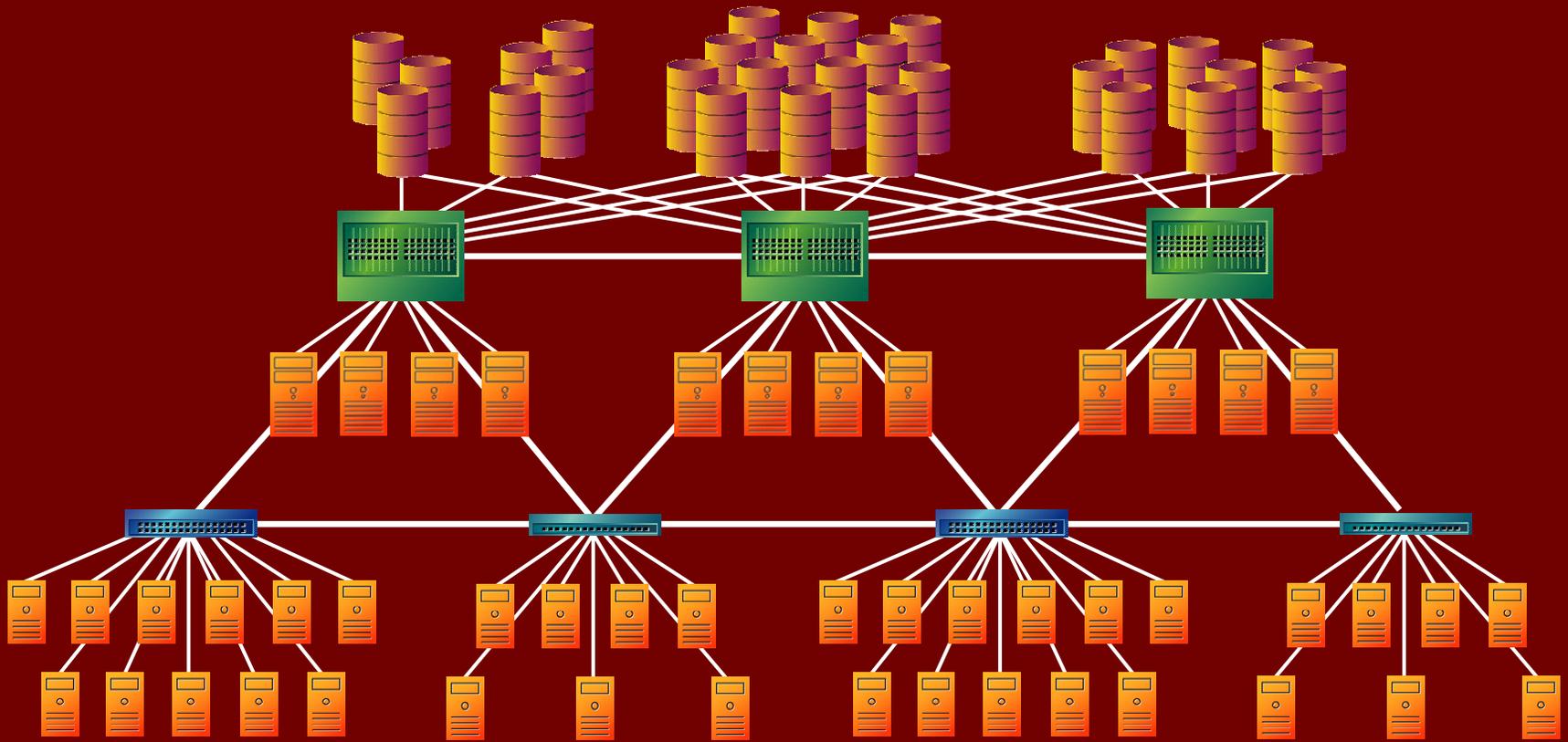Directors
core connectivity
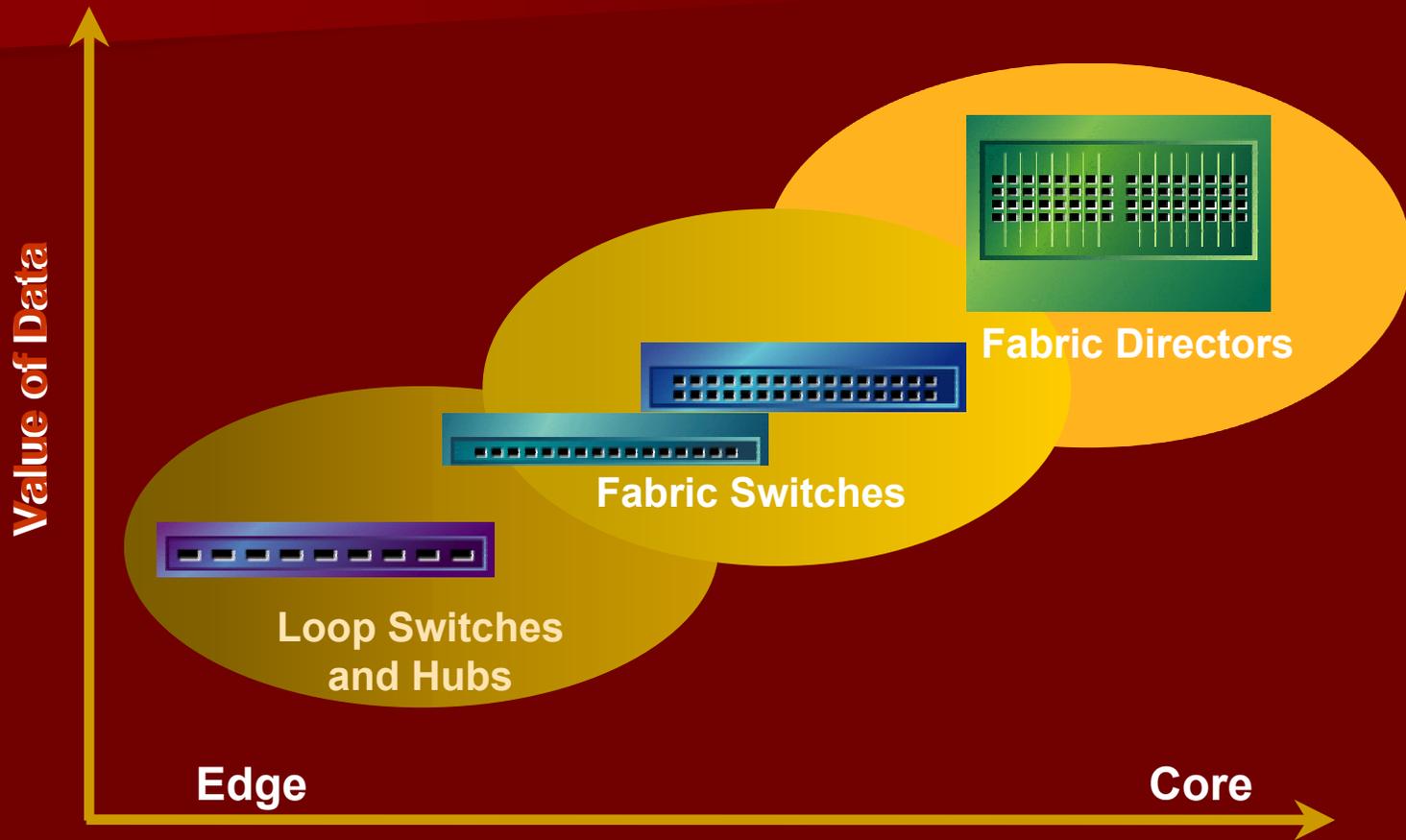
Fabric
Switches

Web
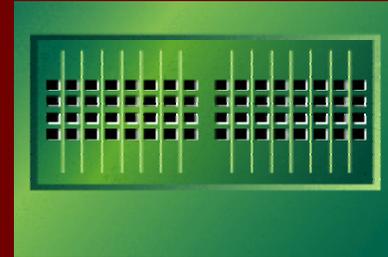Management
Software

Fabric Architecture Design

Fabric Matrix

Directors and Switches

- **Availability**
  - Director: ~99.999%
    - downtime < 5 minutes / year
  - Fabric Switch: ~99.9%
    - downtime 8.8 hours / year
- **Any-to-Any Connectivity**
  - Director: large port count
  - Fabric Switch: small to medium port count
- **Serviceability**
  - Director: non-disruptive
  - Fabric Switch: potentially disruptive
- **Scalability**
  - Director: better fabric scaling
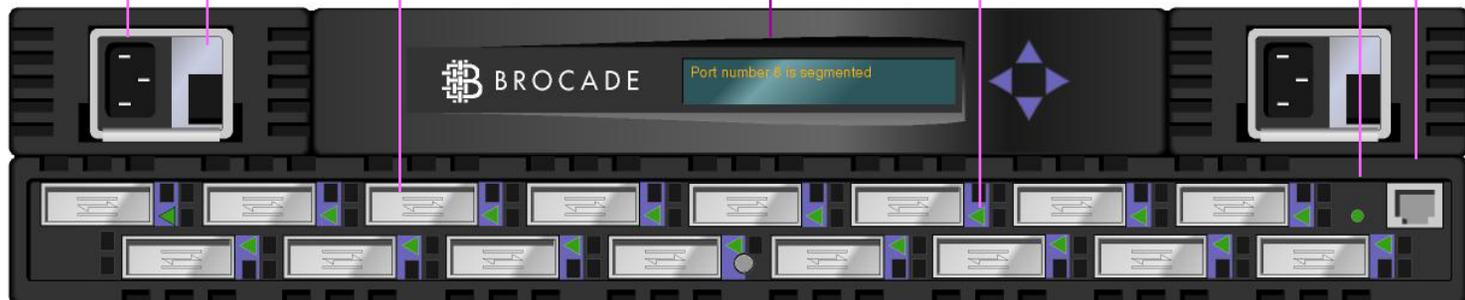 - Fabric Switch: Use Multiple Switches

SilkWorm 2800 Front Panel

1" fibre channel disk drives

- dual-ported 2Gb/sec FC-AL
- dual FC-AL interface allows up to 120 drives to be supported per FC-AL pair
- 36GB and 72GB 10K rpm; 36GB 15K rpm
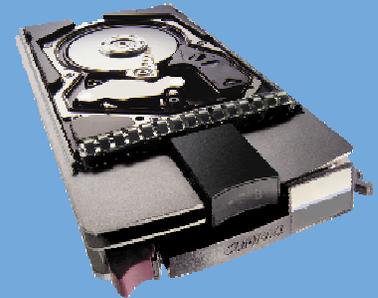
capacity	36GB	72GB	36GB
spindle speed	10K rpm	10K rpm	15K rpm
rotational latency	2.99 msec	2.99 msec	2.0 msec
average read/write	4.7 / 5.2 msec	4.7 / 5.2 msec	3.6 / 4.2 msec
seek time track-to-track read/write seek time	0.6 / 0.9 msec	0.6 / 0.9 msec	0.3 / 0.4 msec

Universal Hot Plug Carrier

AlphaServers™
(DS20E, ES40)

ProLiant™ **Servers**

CL380
Cluster Solution

StorageWorks™ **4314**
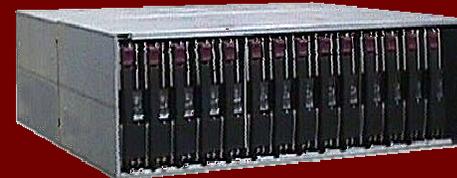*StorageWorks™* **4354**

Modular Array Series

RA 4100

hp online storage portfolio

XP

VA   EVA

MA/EMA

MSA1000

VA7100

EVA

va7410

ACS
v8.7

business enterprise

- monolithic or modular
- high connectivity
- high scalability
- high efficiency
- high performance
- lower TCO
- highest disaster tolerance
- ease of administration
- universal connectivity and heterogeneity

business midrange

- scalable modularity
- heterogeneity
- ease of administration
- price/performance
- moderate functionality

business entry

- moderate scalable modularity
- investment protection:
  - DAS to SAN (DtS)
- moderate scalability
- price sensitive
- faster life cycles

hp nearline storage systems

10/180-20/700 series

ESL series

2/20-10/100 series

MSL series

1/8

1/16

standalone drives

business entry
- low cost
- traditional backup okay
- perfect for DAS
- departmental workgroup
- >8 hour backup window

business midrange
- high storage density, scalable products
- supports Ultrium, SDLT, and DLT
- short backup window <8 hours
- unpredictable growth- need ability to scale
- heterogeneous environments
- medium to large enterprises
- entry-level SAN

business enterprise
- high capacity and performance
- supports Ultrium, SDLT, DLT, and 9840
- very short backup window <4 hours
- advanced levels of high availability, performance, scalability, and capacity
- heterogeneous environments
- enterprise data center
- large scale SANs

SAN Ancillary Products Overview

HP StorageWorks Layered Products Overview

Proactive Services

Network View (w/ Performance Management)

Open SAN Manager

Automation Manager

Resource Monitor

Provisioning Management

SANworks™ SUM Appliance

Element Managers

Storage Allocation Reporter

Data Replication Management Interface

Enterprise Volume Manager

Storage Resource Manager

Secure Path Management Interface

hp storage management software

storage area management

proactive services

OpenView storage area manager suite
- storage node manage
- allocater
- optimizer
- builder
- accountant
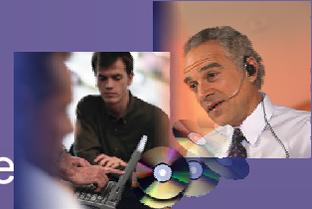
data protection and management

- data protector
- partner solutions
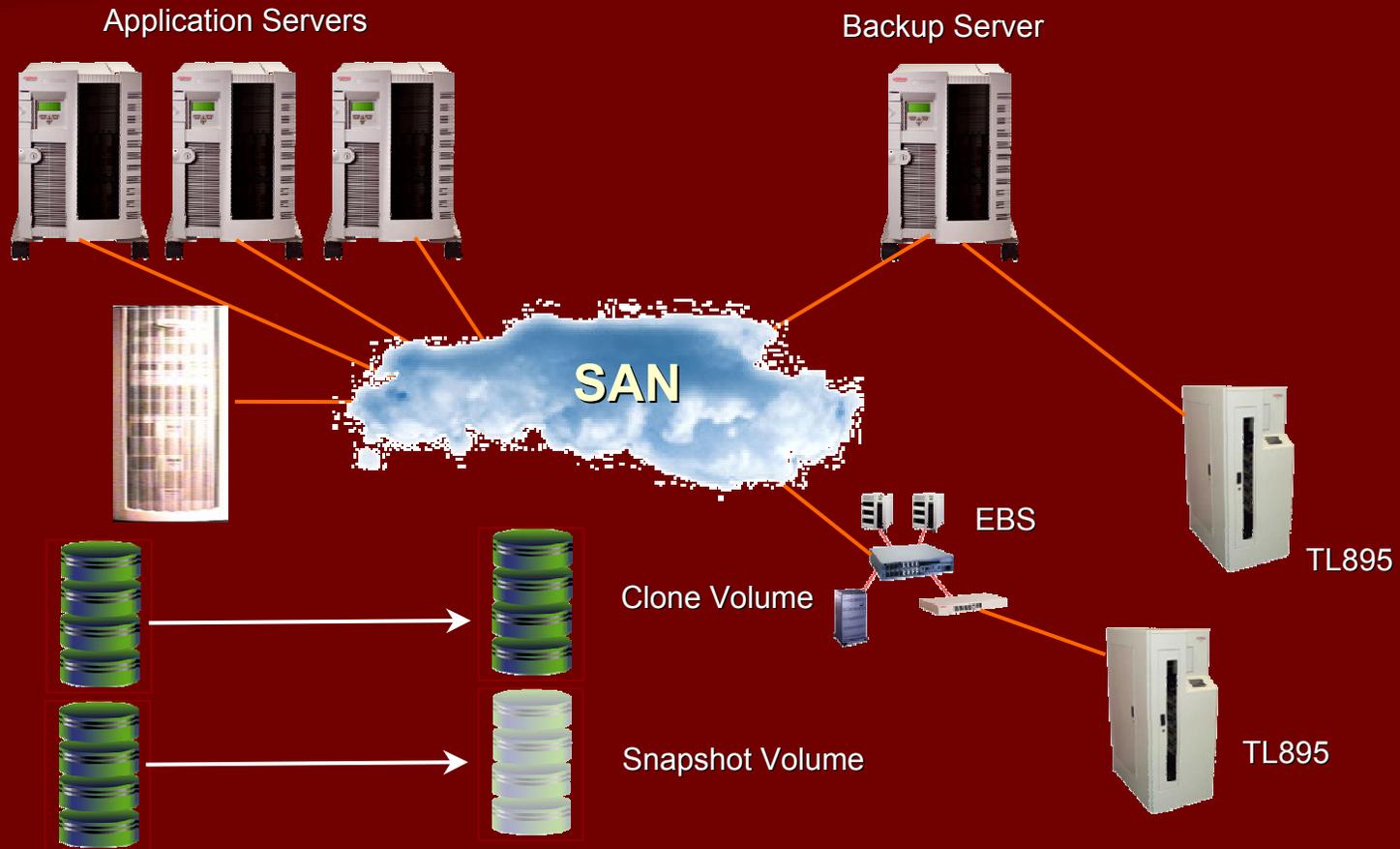
new services
- provisioning and planning

- remote services

HA and replication

- enterprise volume manager
- data replication manager
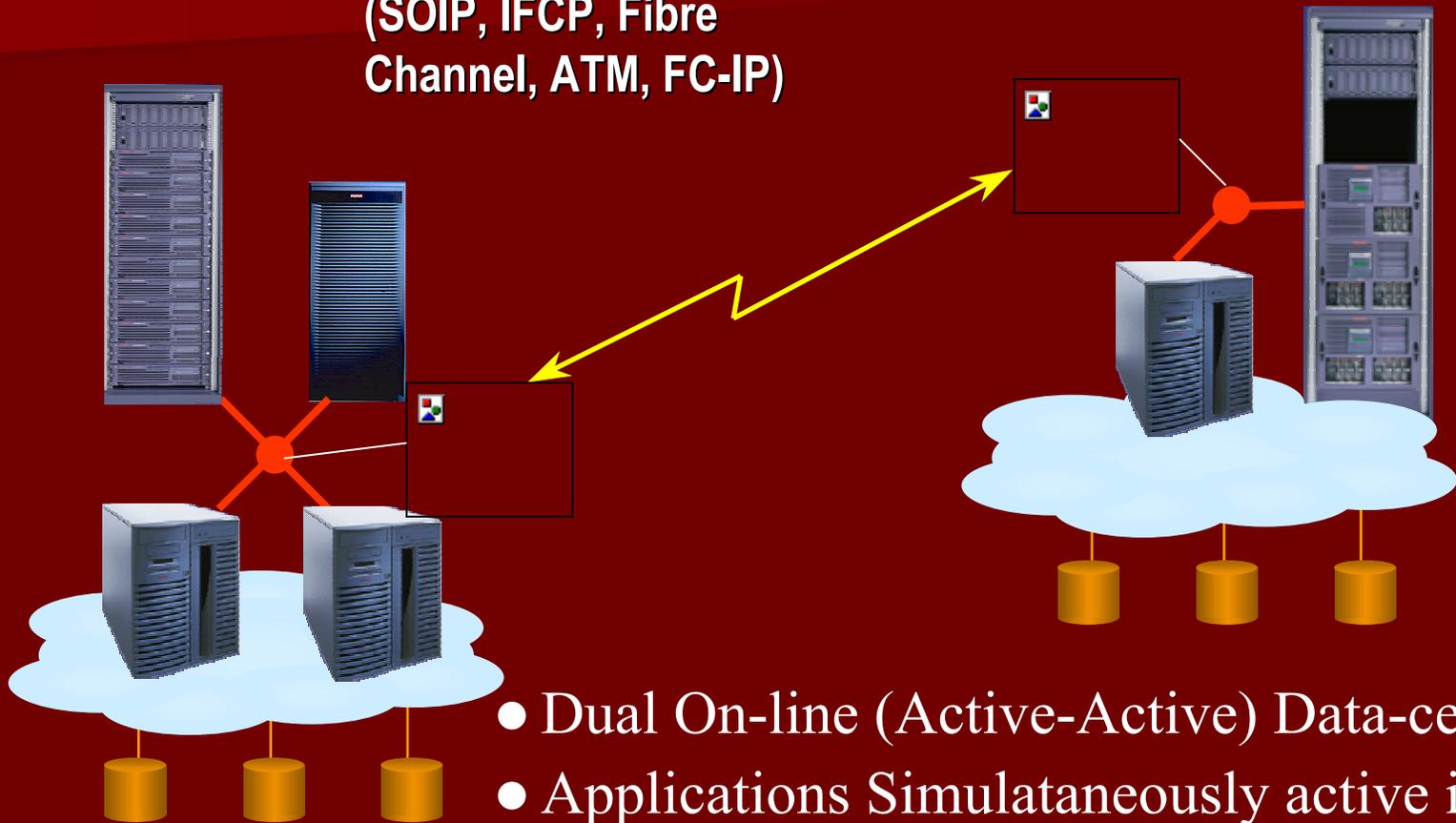- secure path/auto path
- array-based software

centralized management

Point-in-Time Copy Enterprise Volume Manager

Application Servers

Backup Server

SAN

EBS

Clone Volume

Snapshot Volume

TL895

TL895

Highly Available Wide Area Data

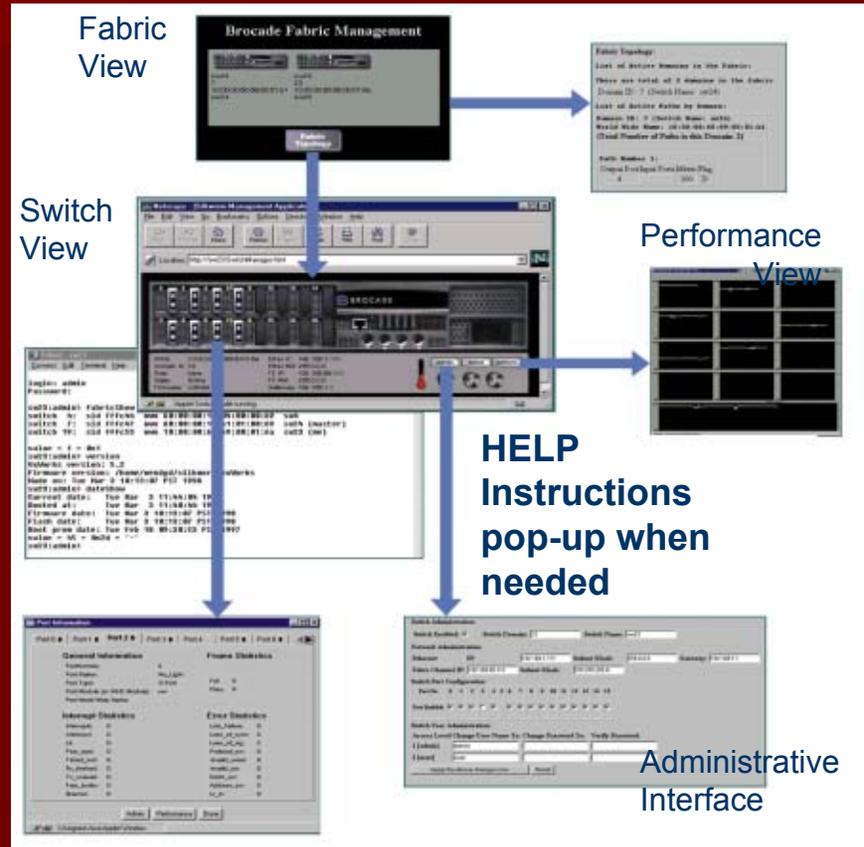**Wide Area Network Links (SOIP, IFCP, Fibre Channel, ATM, FC-IP)**

- Dual On-line (Active-Active) Data-centers
- Applications Simulataneously active in both Data-centers
- Data Constantly Updated Between Data-centers

Web Tools:
Switch Management GUI

- **Platform independent management**
  - JAVA Based
- **Scalable across the entire fabric**
- **In band data from switch to switch**
- **Detailed data & management**
  - Statistics
  - Zoning
  - Name Services

Fabric View

Fabric Topology View

Switch View

Performance View

Telnet Interface

HELP Instructions pop-up when needed

Portal Detail View

Administrative Interface

Brocade® Web Tools

HP SANworks™ Network View

- Simplified SAN administration from one console
- Automated monitoring of device availability and performance and notification when issues arise
- Increases system administrator capabilities
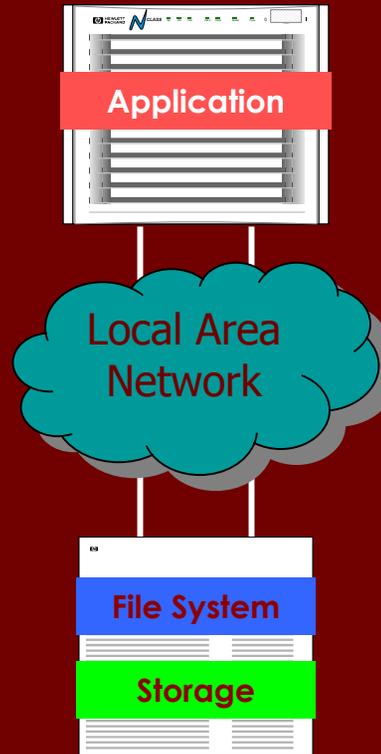- EMC Symmetrix support
- Brocade and McData support
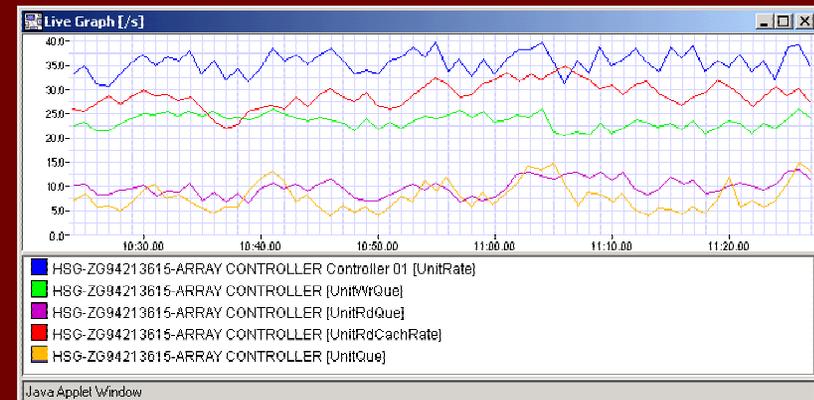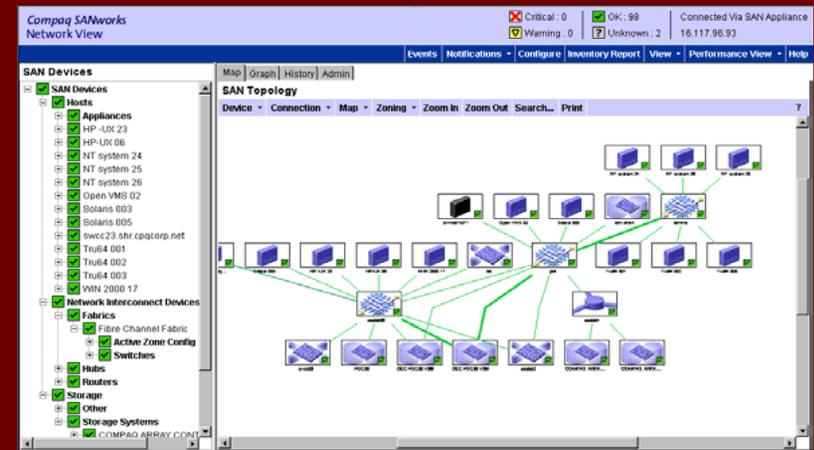
Zoning

- Provides OS & storage isolation
- Store multiple zone configurations
- Zones based on port # or device WWN
- Updates distributed dynamically across the fabric
- Overlapping zones allowed
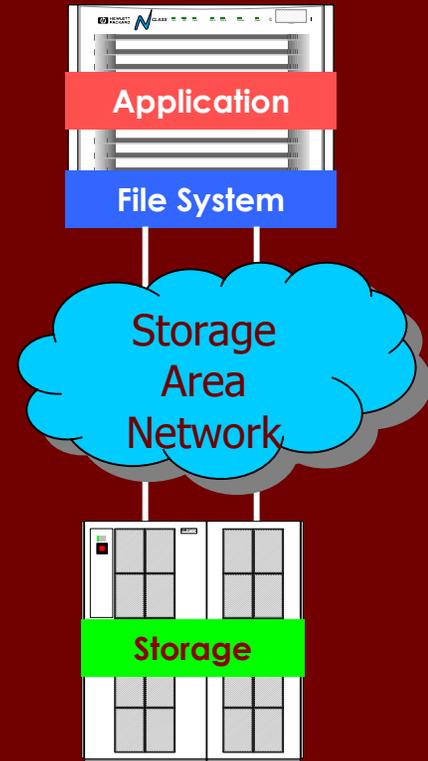- No logical limit on the # of zones

Fibre Channel Fabric

Unix Zone 3

Unix Zone 2

NT Zone 1
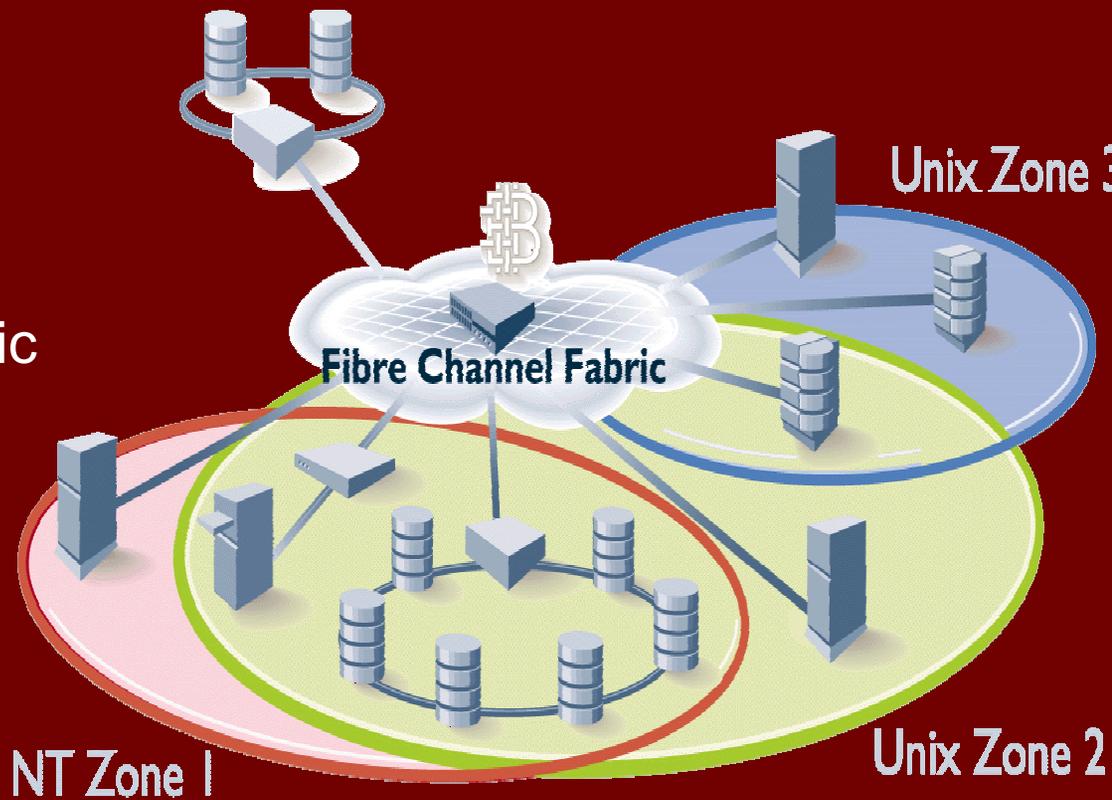
Storage Connect Architectures

Application | **File System**

Storage

Direct Attach Storage

(DAS)

Application

Local Area Network

File System

Storage

Network Attach Storage
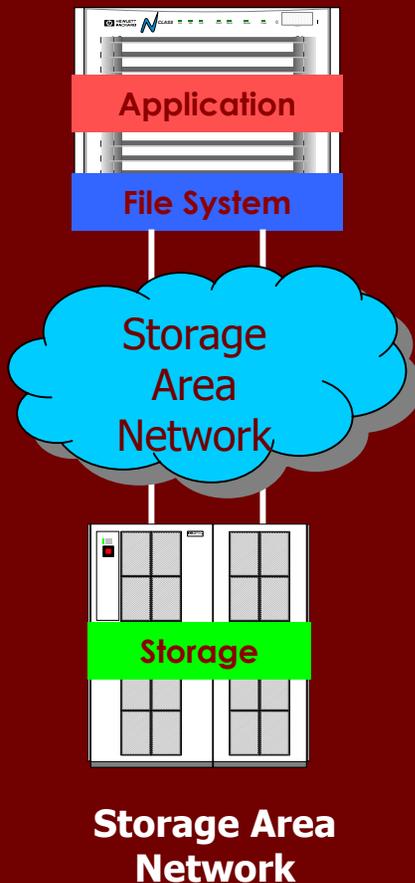
(NAS)

Application | **File System**

Storage Area Network

Storage

Storage Area Network

(SAN)

Networked Storage: SAN

Storage Area Network - Dedicated multi-protocol data network typically implemented with Fibre Channel

Application

File System

Storage Area Network

Storage

Storage Area Network

- Typically block level access as opposed to file level access
- Server oriented
- File system on application server
- Provides more distance than DAS
- Overcomes connectivity of DAS
- **Great for Database support**

Networked Storage: NAS

Network attached storage that is accessed by clients and servers via standard network interface

Application

Local Area Network

File System
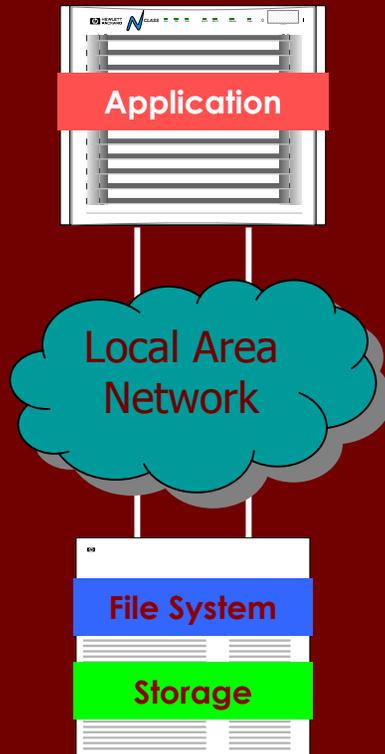
Storage

Network Attach Storage

- Optimized for a file serving
- Designed to move files
- Easy Installation and Monitoring
- Network access by workstations
- Utilize existing network
- Unix/Windows file sharing

The Need For Networked Storage

- Networked Storage Drivers

  – The explosion of data & network traffic

  – Increasing value of data & information

  – Requirement for non-stop information access

  – Transforming data to information; to knowledge; to competitive advantage

  – Accelerating pace of technology change

  – World Wide Web — explosion

  – Shortage of skilled personnel

Why Networked Storage?

- Moving to networked storage provides
  - Increased data and network availability
  - Consolidation of storage and practices
  - Increased heterogeneous storage utilization
  - Consolidated storage acquisition costs
  - Simplified and centralized management
  - Non-disruptive storage/server scalability
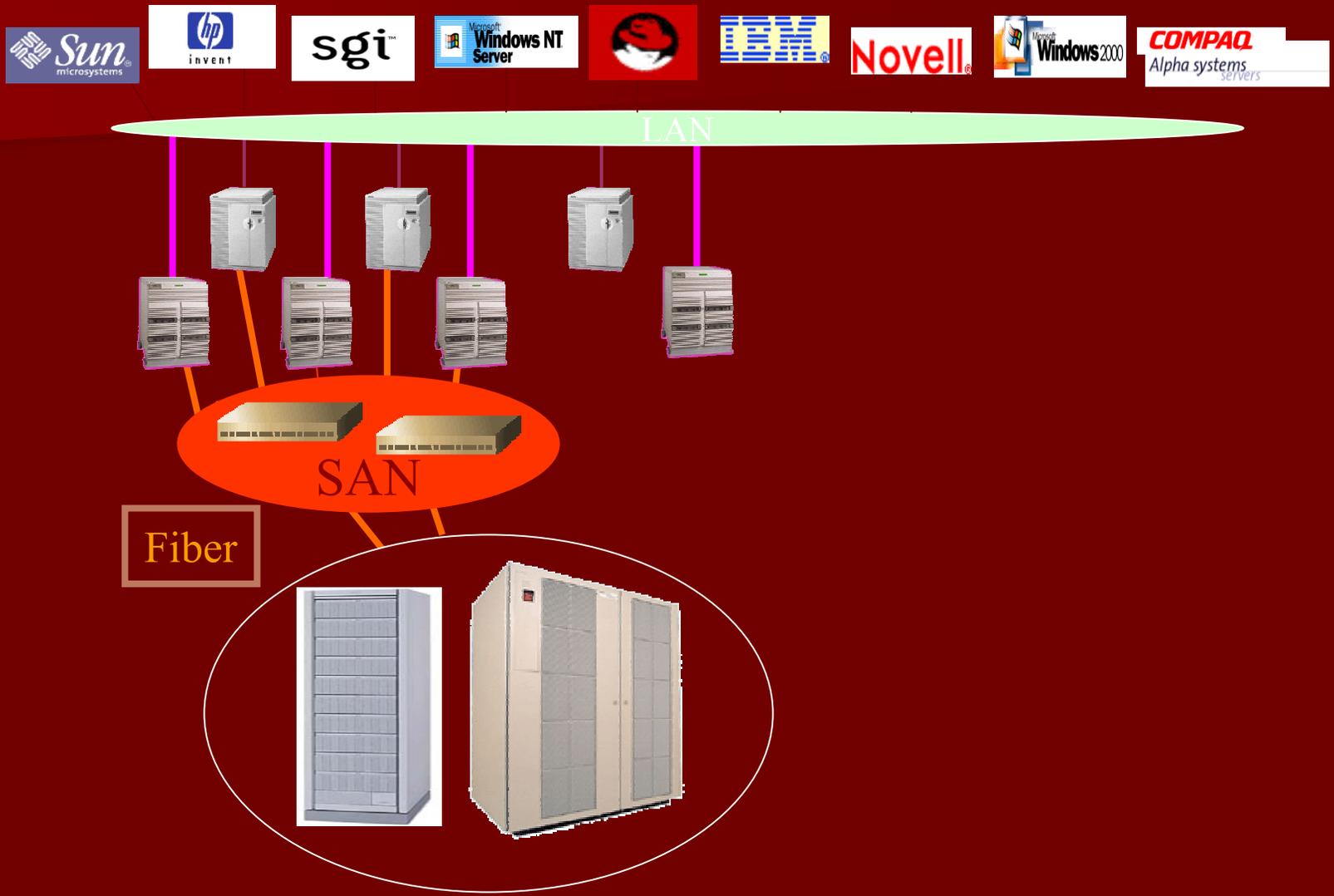
Why Network Attached Storage?

- Networked Storage: NAS
  - Multiple Protocol support
    - Microsoft Clients (SMB/CIFS)
    - UNIX Clients (NFS)
    - HTTP - (Web Browsers)
  - Removes server I/O bottlenecks and overhead
  - Speeds increase as bandwidth increases
    - 10Mb>100Mb>1Gb
  - Easy Installation and Manageability
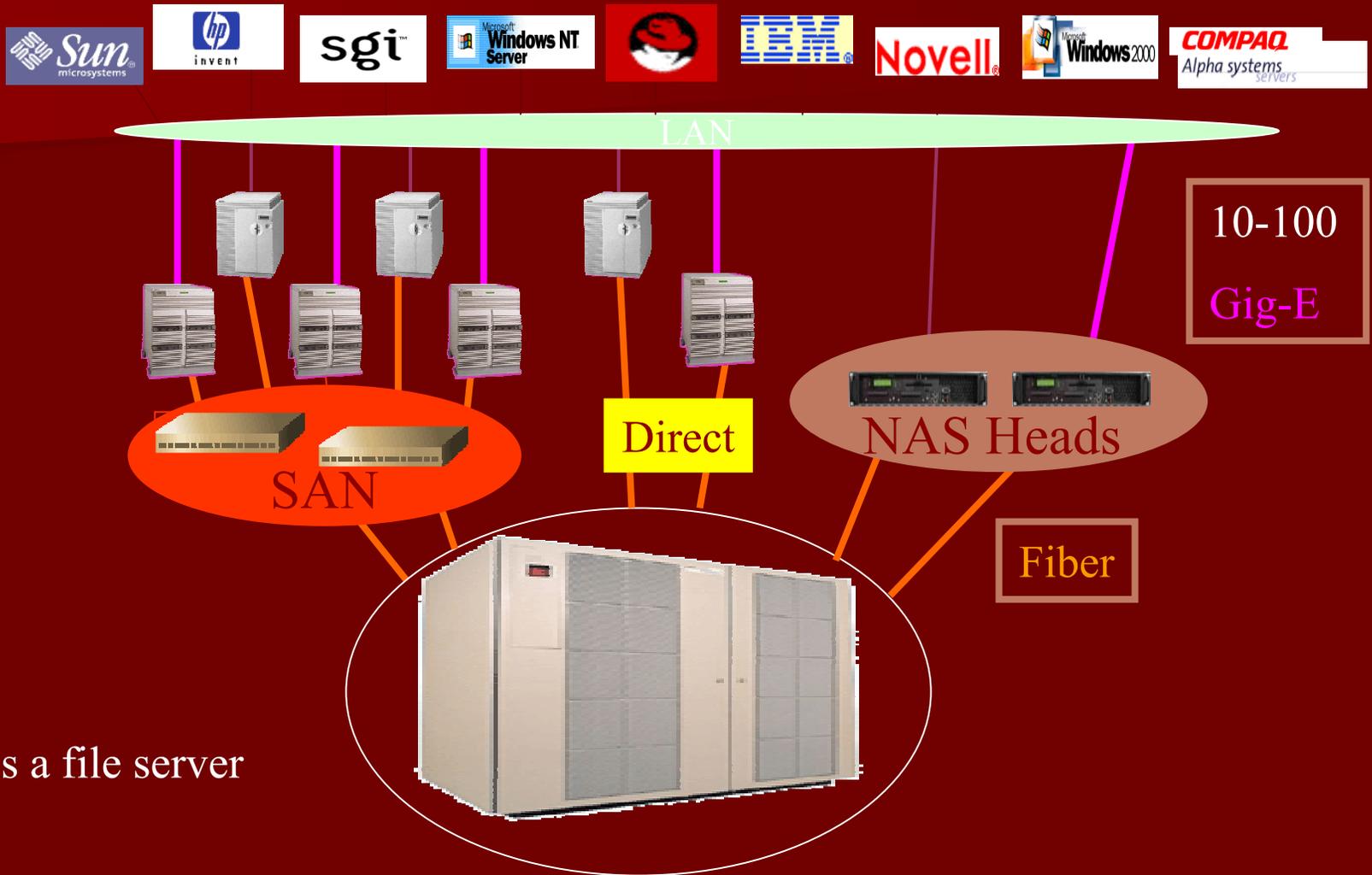
Direct Attachment

LAN

Direct Attachment

Disk Arrays
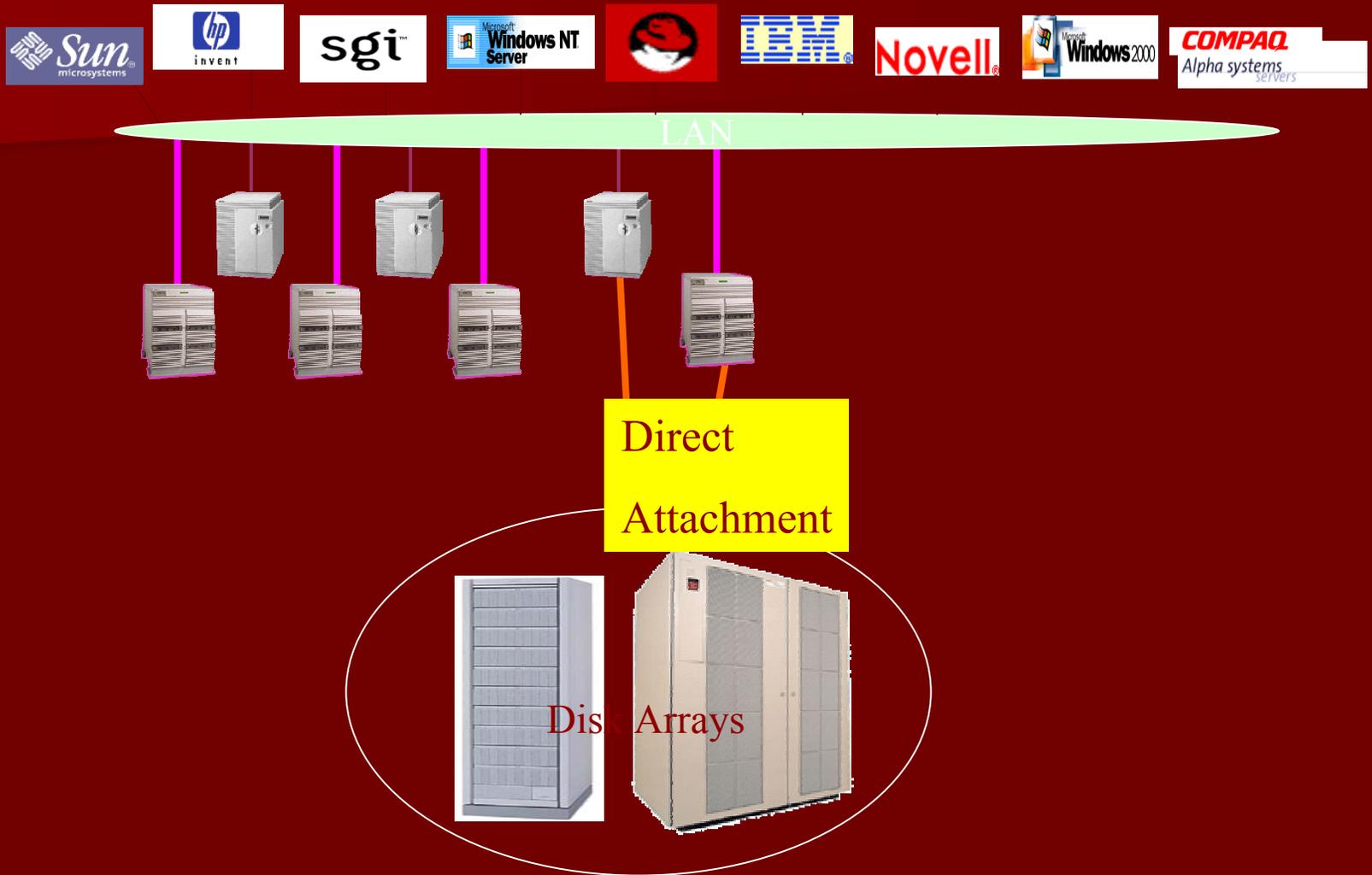
SAN Attachment

NAS Attachment

LAN

10-100

Gig-E

Direct

NAS Heads

SAN

Fiber

NAS is a file server

NAS Applications

- File serving(relieve NT servers)
- Web serving
- Multimedia
- Video streaming
- Graphic design
- CAD/CADAM
- Engineering
- Imaging

Flat File Oriented

Focusing On NAS

- Provides multi-server access over a **_shared_** network using network FILE (NFS/CIFS) protocols

- NAS is a server independent appliance

- Reduces server bottlenecks and overhead

- NAS uses existing network infrastructure

- Easy installation and management

NAS Advantages

- Networked Storage: NAS
  - Multiple Protocol support
    - Microsoft Clients (SMB/CIFS)
    - UNIX Clients (NFS)
    - HTTP - (Web Browsers)
  - Removes server I/O bottlenecks and overhead
  - Speeds increase as bandwidth increases
    - 10Mb>100Mb>1Gb
  - Easy Installation and Manageability

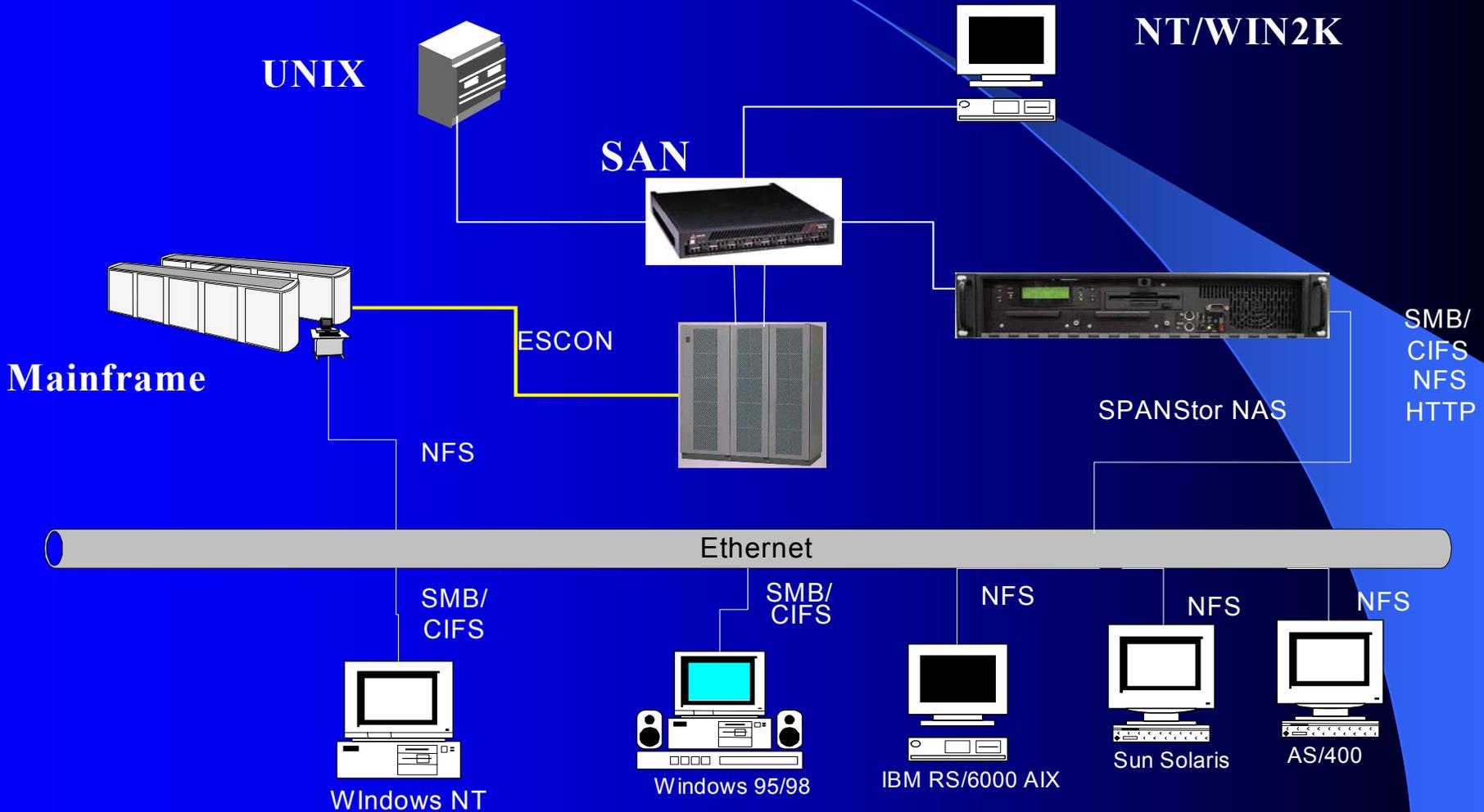| NAS	SAN
NFS protocol	SCSI protocol
Globally over a net	Over a storage interconnect that looks like a network
Slower NFS speed	*Faster SAN SCSI speed*
NFS overhead greater	*Lower SAN SCSI overhead*
Inexpensive, easy complicated	Prices dropping,
NAS established today	SAN is now a Standard
Shared storage	Shared storage
Shared data	Promises shared data

NAS and SAN

	NAS	SAN
Type of Storage	Shared files	Virtualized, pooled blocks
Protocols, Transport	TCP/IP, Ethernet	Fibre Channel
Consumer	Clients or servers	Server
Capacity	100s of TB	100s of PB-EB
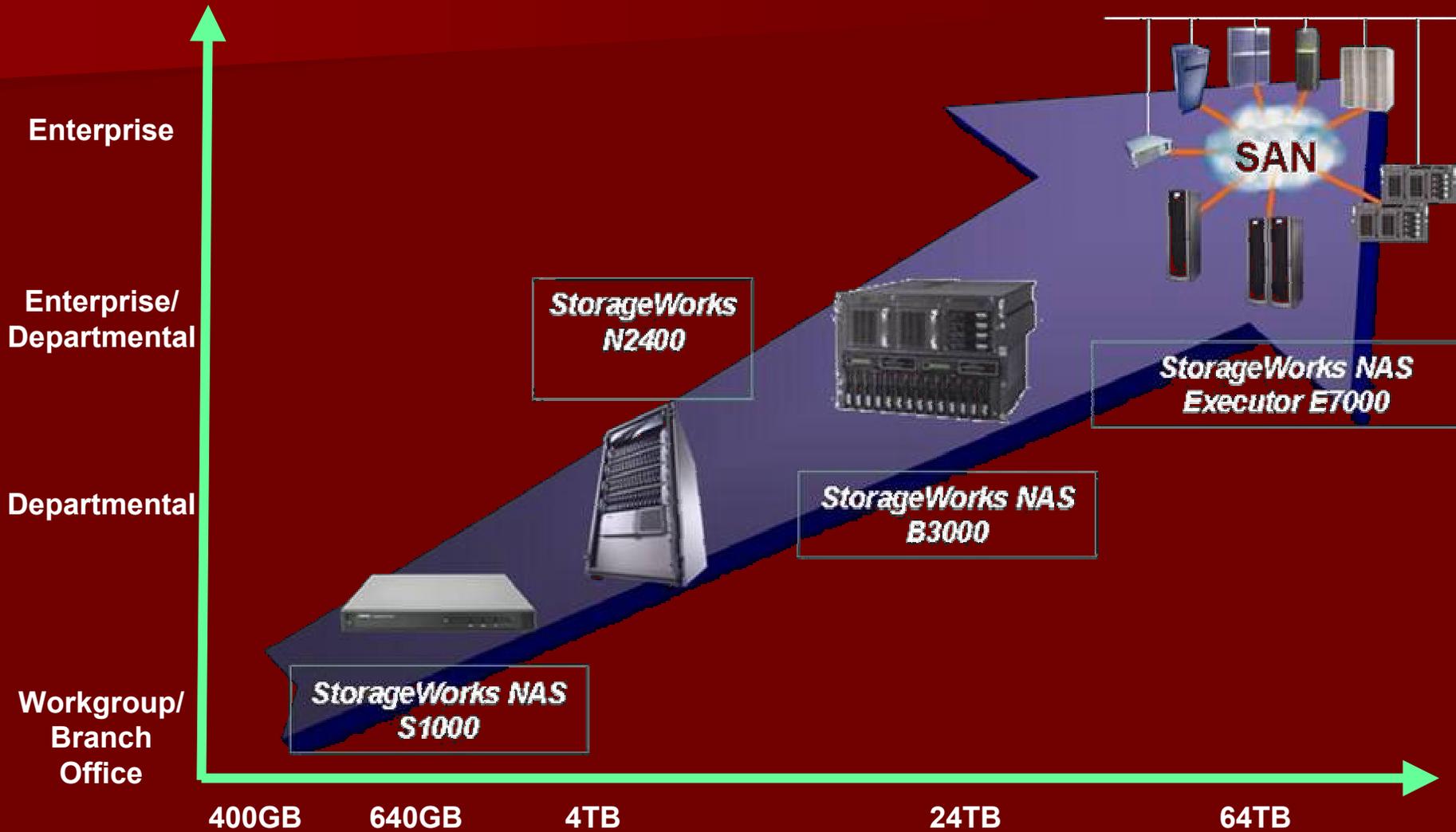
SAN/NAS Integration

UNIX

NT/WIN2K

SAN

Mainframe

ESCON

SPANStor NAS

SMB/
CIFS
NFS
HTTP

NFS

Ethernet

SMB/
CIFS

SMB/
CIFS

NFS

NFS

NFS

WIndows NT

Windows 95/98

IBM RS/6000 AIX

Sun Solaris

AS/400

| NAS	SAN
NFS protocol	**SCSI protocol**
Globally over a net	**Over a storage interconnect that looks like a network**
Slower NFS speed	*Faster SAN SCSI speed*
NFS overhead greater	*Lower SAN SCSI overhead*
Inexpensive, easy complicated	**Prices dropping,**
NAS established today	**SAN is now a Standard**
Shared storage	**Shared storage**
Shared data	**Promises shared data**

Virtualization Technology Introduction

Virtualization: Storage

- **Enables vast amount of physical capacity scattered across the enterprise to appear as a single large pool of data for presentation to application servers.**
    - **Storage pool is a single manageable entity**
    - **Virtualization can occur at any of three levels.**
    - **All levels can be used together, or independently, to maximize the benefits to customers:**
        - ***Server level*: Ideal for small entry level needs**
        - ***SAN Fabric level*: SAN-wide virtualization**
            - **increase efficiency in SAN development, management, and service**
        - ***Storage System level:* Ideal for large volumes without sacrificing performance or reliability**

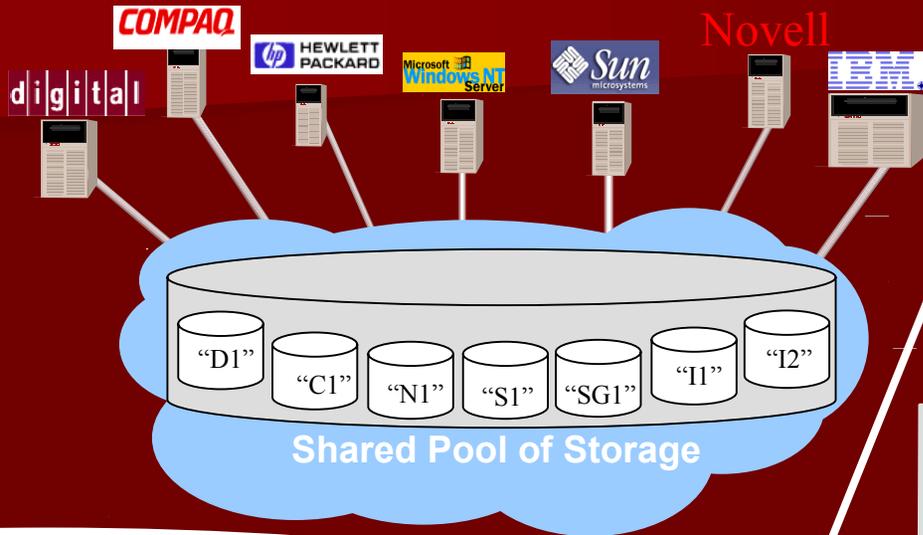# Virtualization — Abstraction of Storage at the Block Level
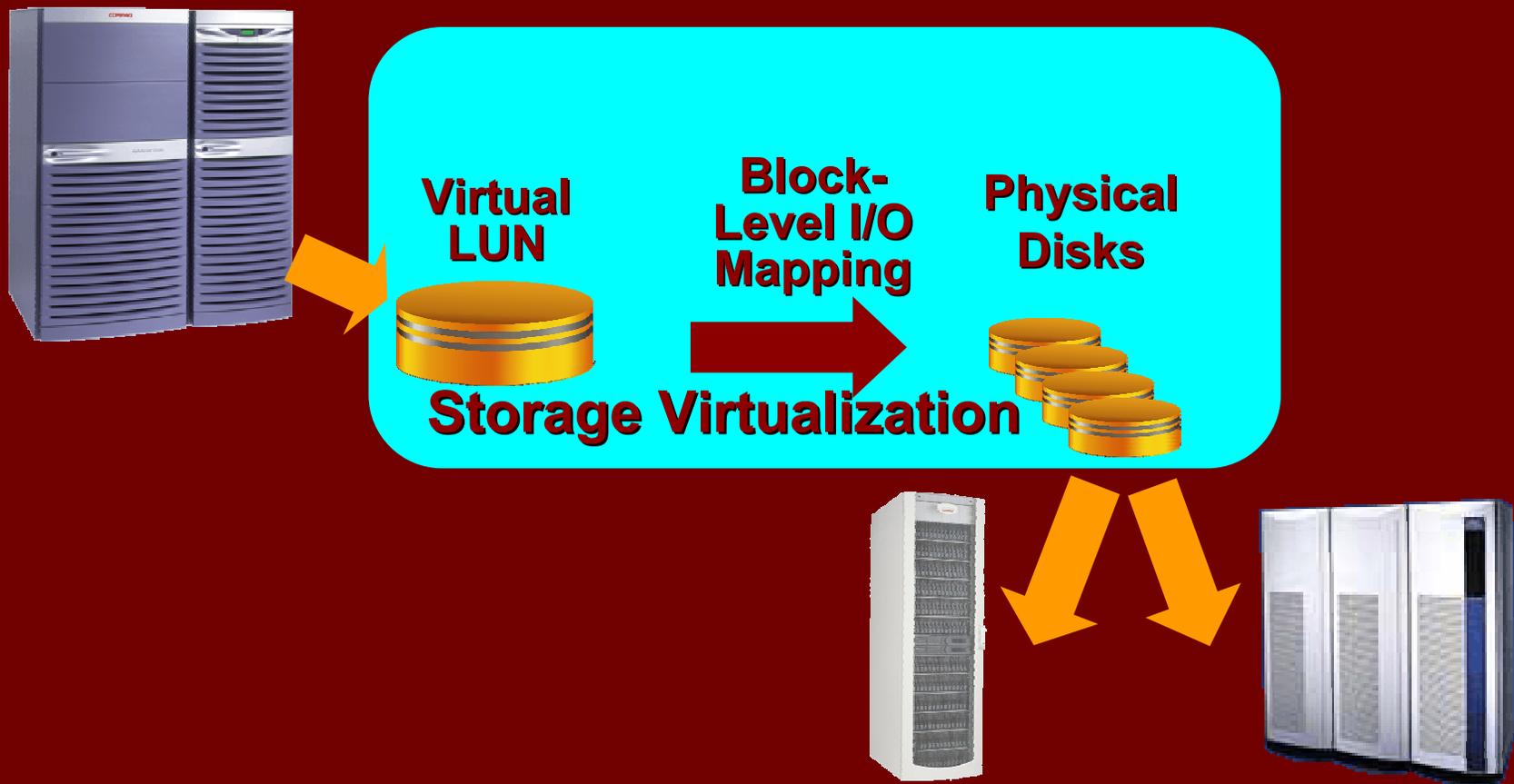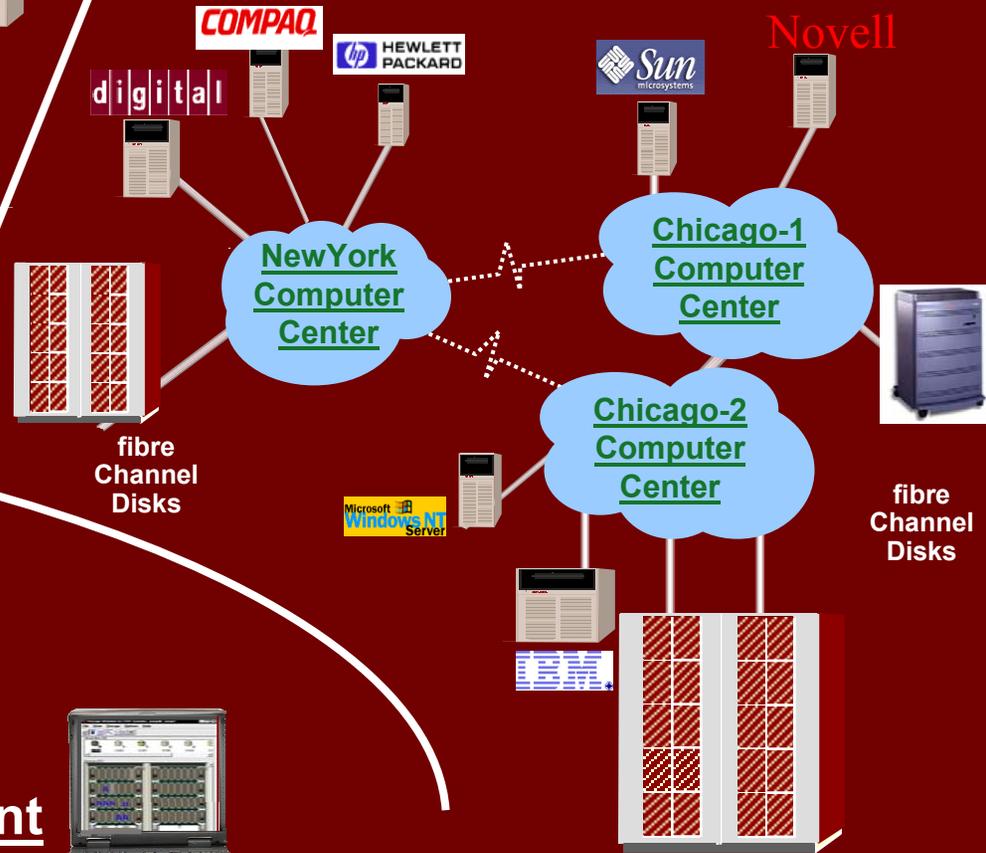
Why is Virtualization Important Anyway ?

- Virtualization will:
  - Provide investment protection for non-virtualized arrays
  - Provide a Common Storage Pool that will allow for greater efficiency and not a load of stranded capacity in servers or on SAN RAID volumes/arrays
  - Data Migration that is transparent to the OS for upgrades, changes, etc to dataset for applications
  - Reduction of ownership costs

The Future - Open Virtual Storage

SAN Wide Virtual View

Shared Pool of Storage

Disks: "D1" "C1" "N1" "S1" "SG1" "I1" "I2"

Actual Geography

NewYork Computer Center

Chicago-1 Computer Center

Chicago-2 Computer Center

fibre Channel Disks

fibre Channel Disks

Individual System View

"D1"

Global Virtual Management

StorageWorks Enterprise Virtual Array

Server Level

- Homogeneous SAN and non-SAN deployment
- Limited interoperability with HWare or Sware
- SANworks Virtual Replicator – MS Windows

Fabric Level

- Open SAN environments
- Conventional and virtualized storage systems
- SANworks VersaStor – Asymmetric (future)

System Level

- Independent of the server host
- Virtually Capacity-Free Snapshot on RAID
- StorageWorks Enterprise Virtual Array

EVA Virtualization Primer

- Principles of EVA virtualization
  - "Disk blocks are disk blocks"
    - All disk blocks are equivalent*
  - Foundation object is the "Virtual Disk"
    - A SCSI LUN
      - Attributes
      - Data
    - It doesn't matter where or how data actually stored

# EVA Virtualization Primer (con't)

- Data location independence
  - Makes new benefits possible
  - BUT - to realize those benefits, we must change some old habits
    - Backend disk access pattern no longer correlated with host access pattern
    - Stop using configuration as "organizing" tool
      - Use EVM capabilities instead
    - Resist temptation to micromanage
    - Avoid temptation to utilize every last byte

Virtual Disk Mapping

● Controller assigns the logical blocks of a virtual disk to specific physical disk blocks via a dynamic "map" that the controller manages.

SCSI LUN

Disk Drives

LBA 0
LBA 1

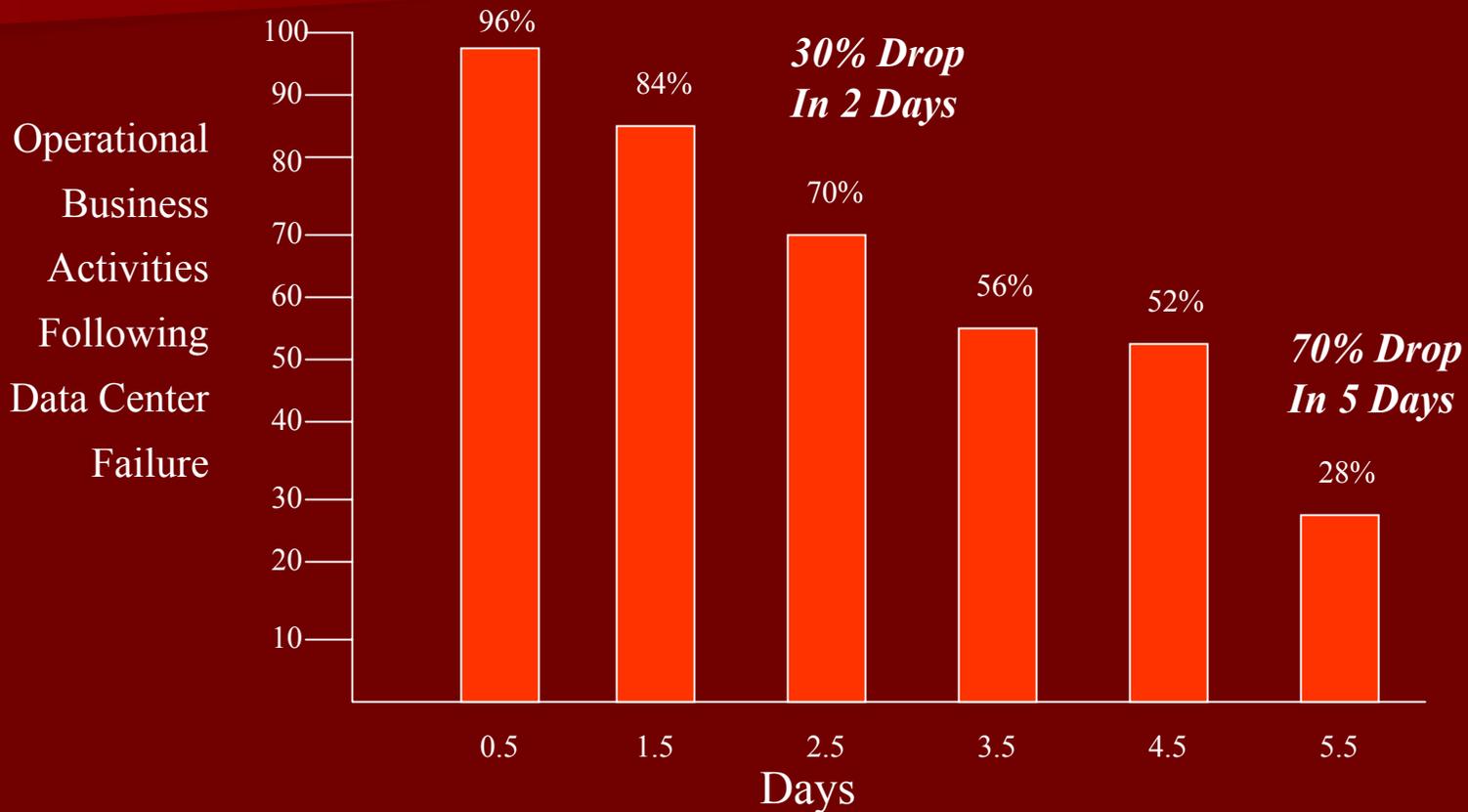MAP

LBA n
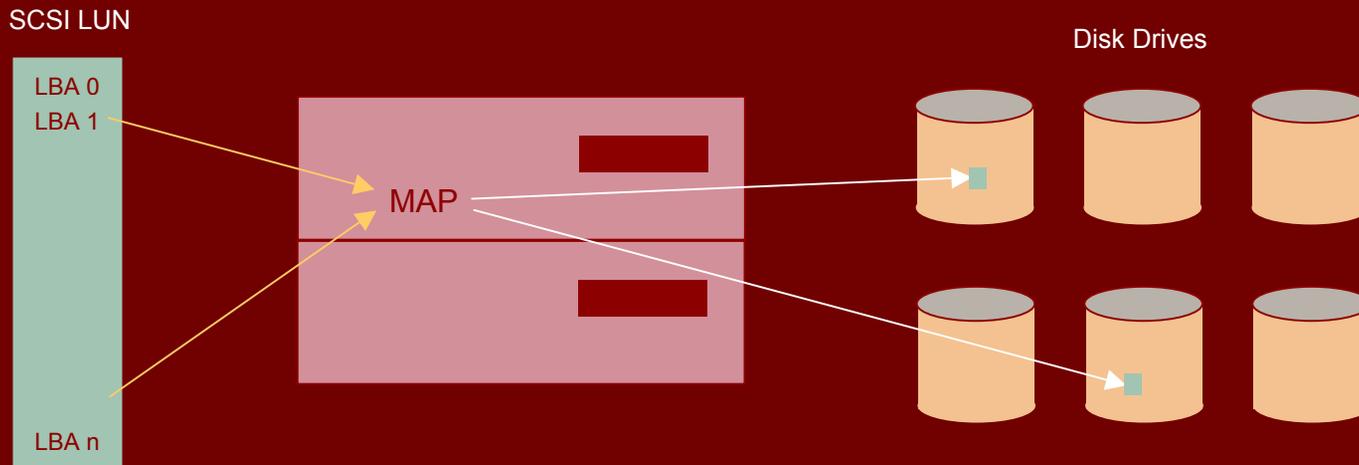
Virtual Disk Mapping (con't)
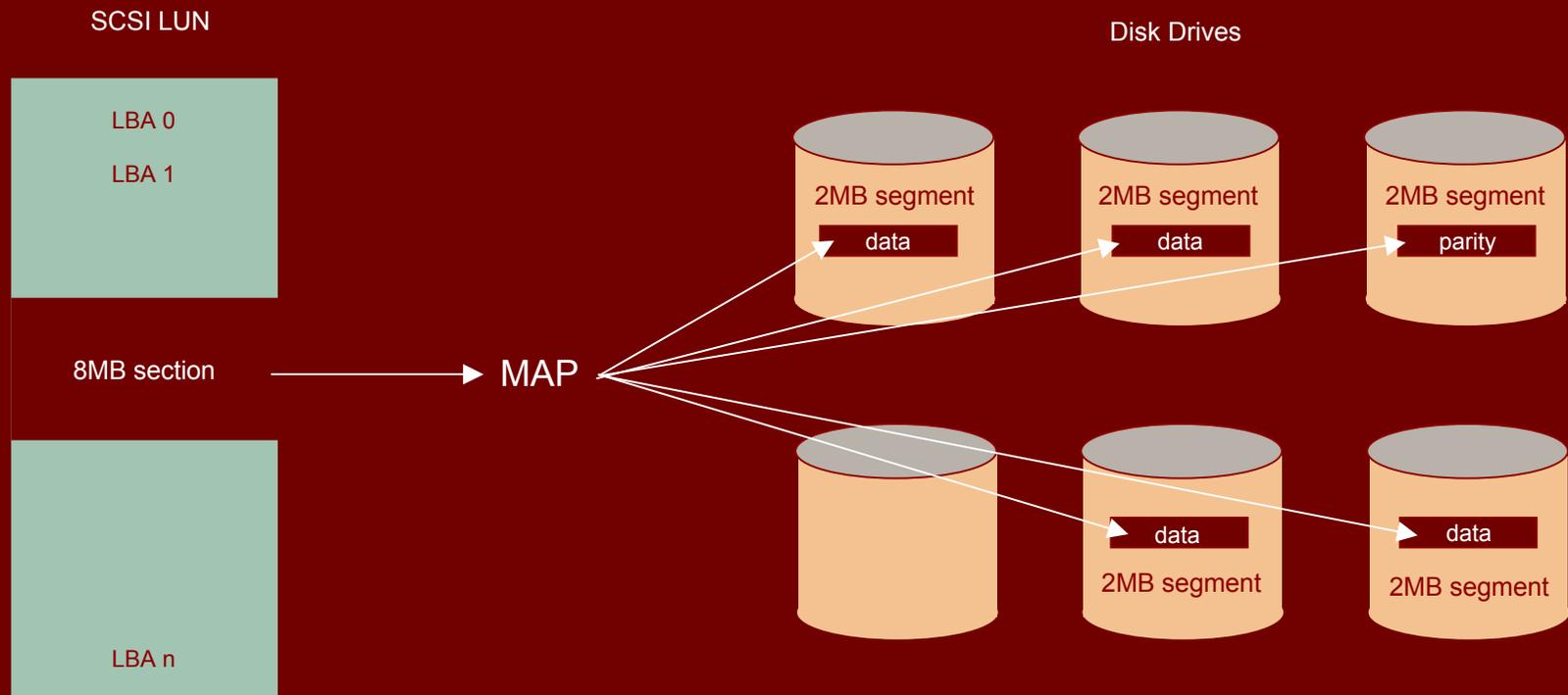
Virtual Disk Mapping (con't)

● Physical segments for a given virtual disk

- can come from any of the disks in the disk group

- can and will change over time

- may be "reserved" but not "assigned" until later

- may be "assigned"

-- when unit is created

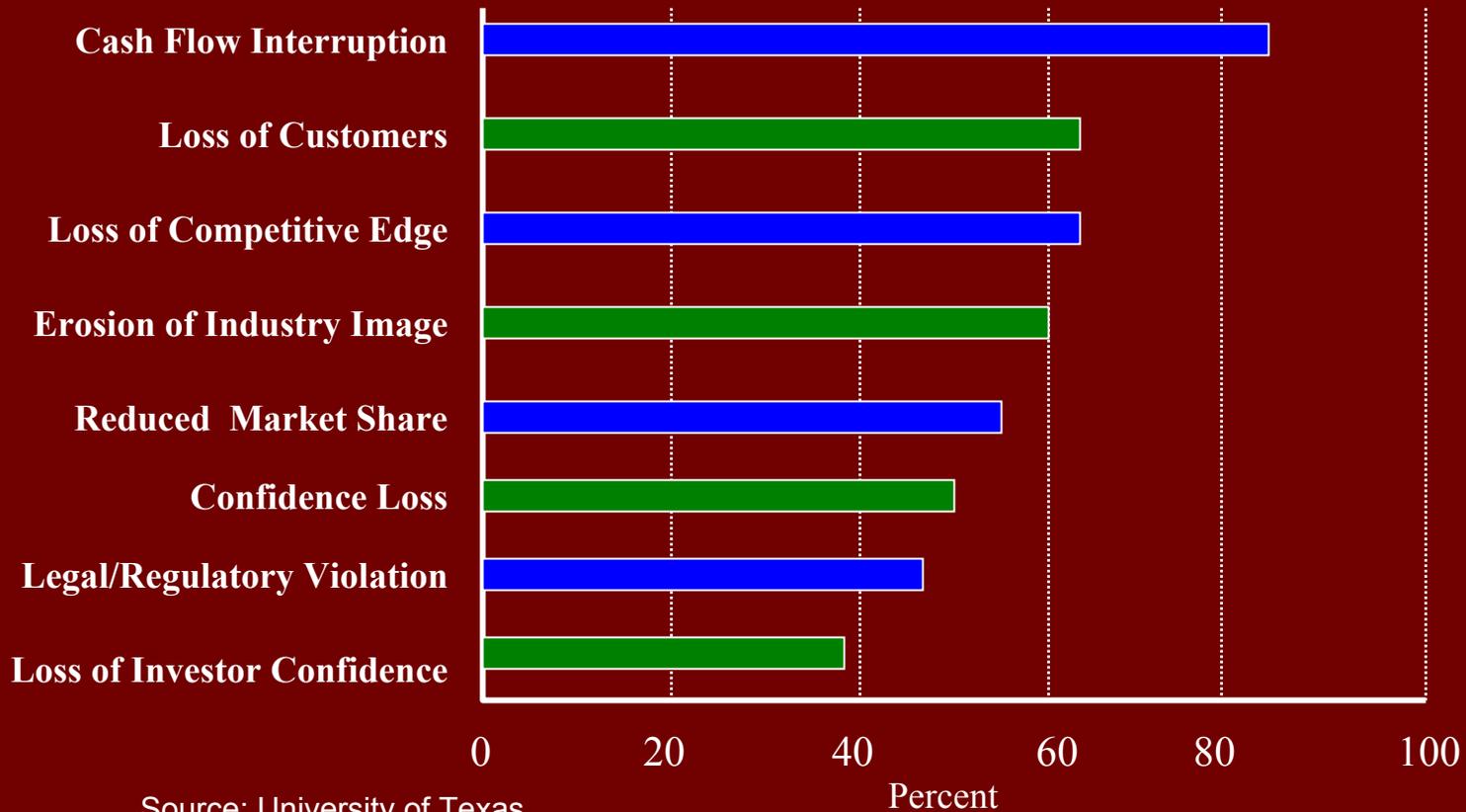-- when needed

Business Continuity & Disaster Recovery

Loss of Capability to Perform Basic Business Functions

Operational Business Activities Following Data Center Failure

96% — 0.5 Days
84% — 1.5 Days
70% — 2.5 Days
56% — 3.5 Days
52% — 4.5 Days
28% — 5.5 Days

30% Drop In 2 Days

70% Drop In 5 Days

Days

No longer just a computer disaster after 1 or 2 days

Intangible Costs of Computer Outages

From a survey of 160 MIS Managers

Cash Flow Interruption
Loss of Customers
Loss of Competitive Edge
Erosion of Industry Image
Reduced Market Share
Confidence Loss
Legal/Regulatory Violation
Loss of Investor Confidence

0   20   40   60   80   100

Percent

Source: University of Texas

The Effects of Data Loss

- When Companies lose data in a disaster, 50% never reopen and 90% are out of business in two years
  - Source: University of Texas
- 93% of companies that suffer a significant data loss are out of business within five years
  - Source: U.S. Bureau of Labor
- 43% of U.S. businesses never reopen after a disaster and 29% close within two years
  - Source: University of Wisconsin
- 30% of computer users say the spend the equivalent of one week per year reconstructing lost data.
  - Source: 3M Corporation

Estimated Value of Data

Website corporate page	1-10 percent of daily sales
Customer support call center	1-10 times an average invoice
E-mail message database	$1000 per employee
Competitive intelligence	1-5 percent of annual sales

Terms

- Fault Tolerant
- Disaster Tolerant
- Disaster Recovery
  - RPO, RTO, RSO
- Business Continuity
- Contingency Planning

Fault Tolerant

- Applied to CPUs and some subsystems
- No *single point of failure* (spof)
- A "local phenomenon"

Disaster Tolerance

- *Disaster Tolerance* is Fault Tolerance where the single point of failure you are trying to eliminate is the <u>datacenter</u>.
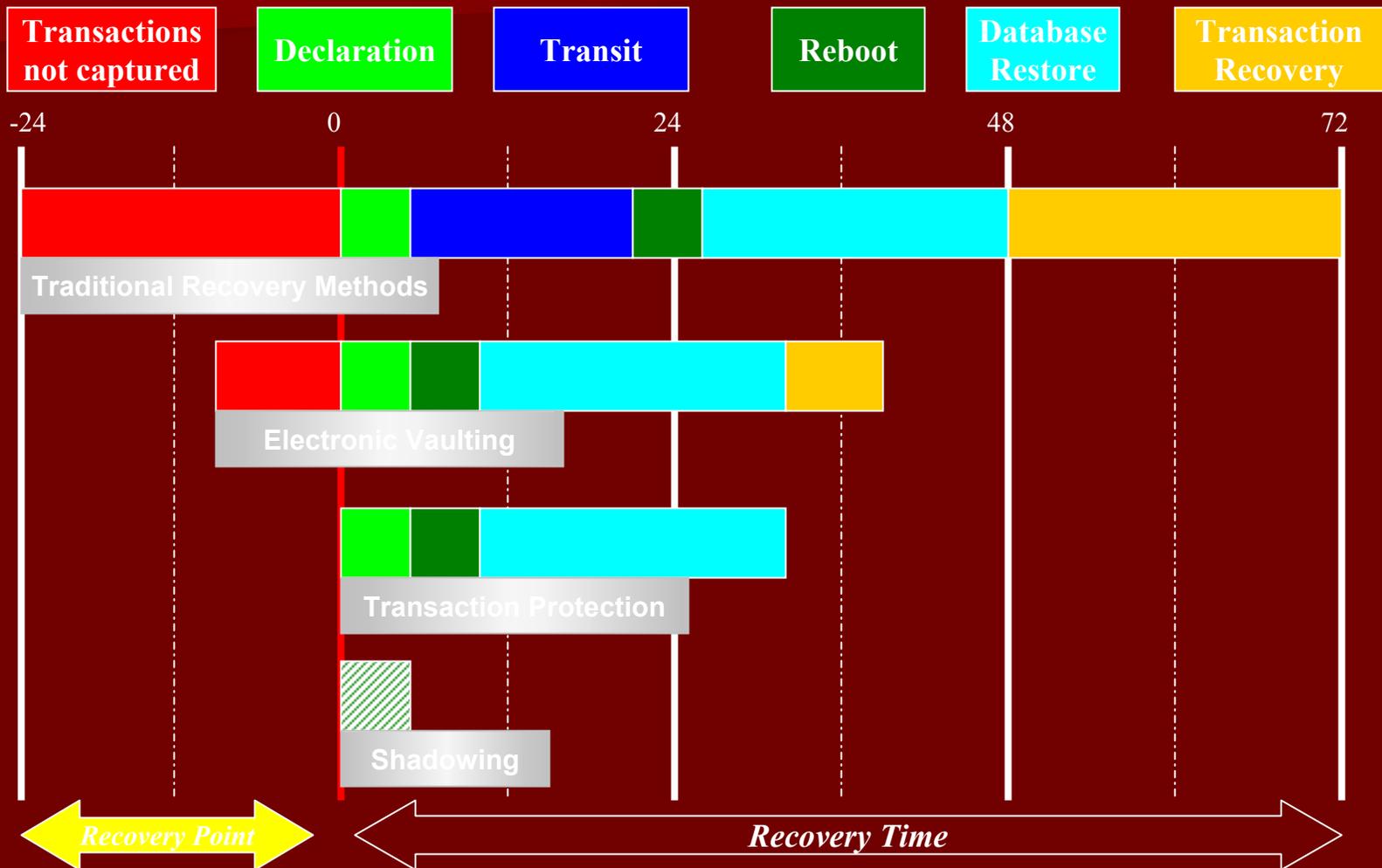
- Achieved through redundancy and distance.

Disaster Recovery

- Usually refers to physical disasters, i.e., earthquake, fire, flood, terrorist attack
- Includes more than just computing infrastructure
- Associated with three objectives
  - Recovery **Point** Objective
  - Recovery **Time** Objective
  - Recovery **Service** Objective

Objectives
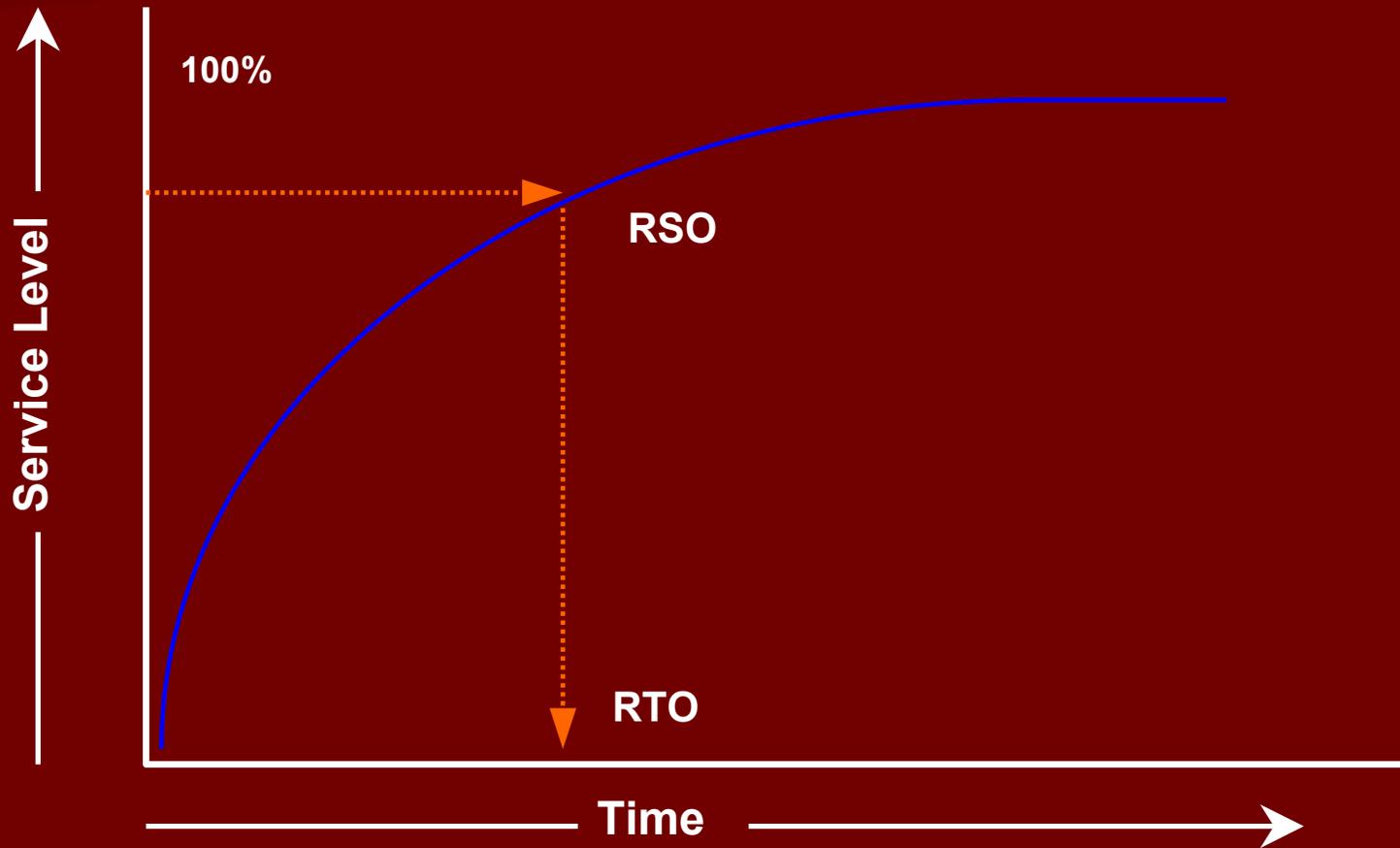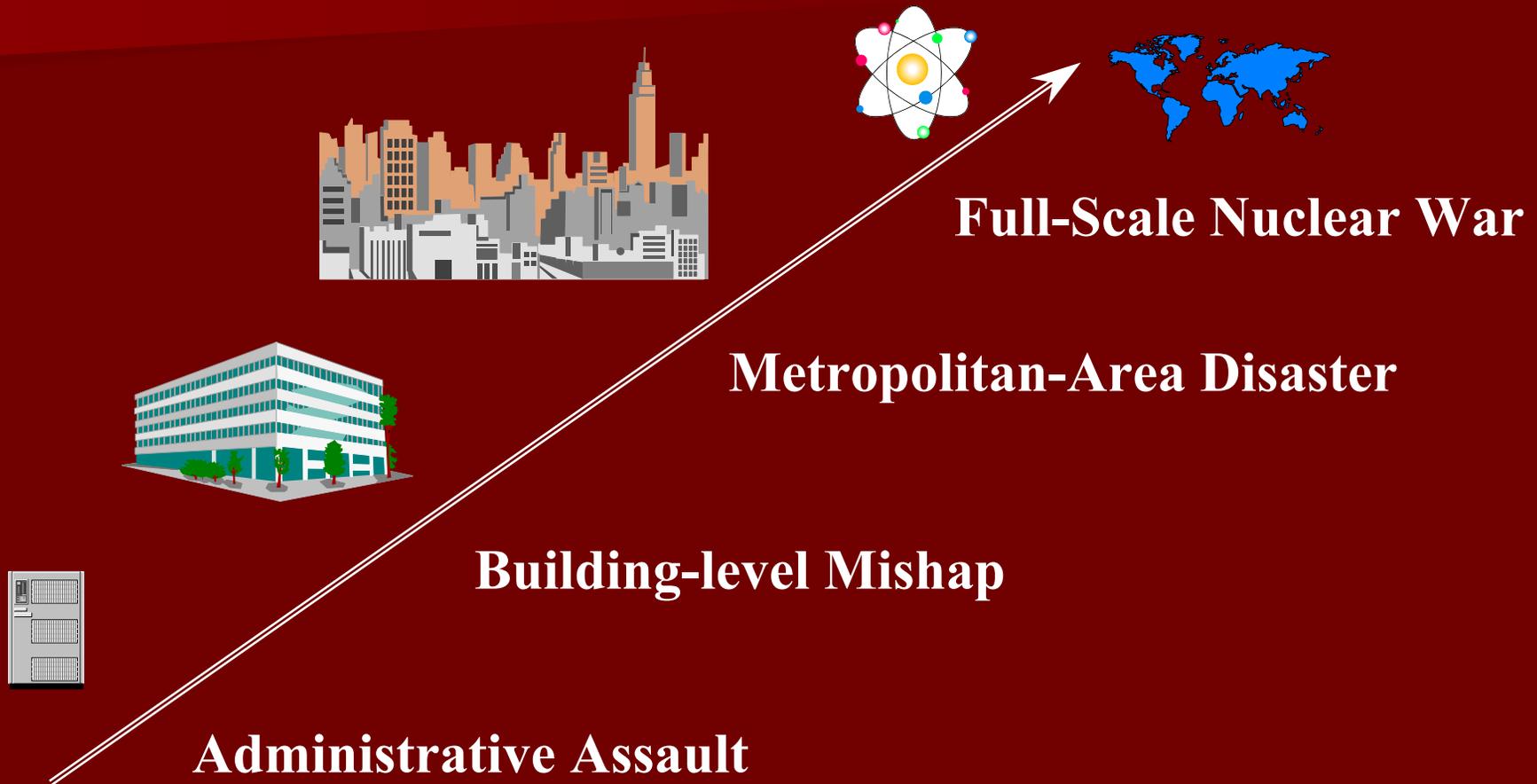
Disaster Recovery Timelines

BC/CP

- Business Continuity/Contingency Planning
  - Usually refers to Disaster Recovery **plus** non-physical phenomena such as product tampering, failure of a supplier, strike, biohazard.

Implementing Disaster Recovery

- Two major questions:
  1. What disaster are you trying to survive?

Implementing Disaster Tolerance

Full-Scale Nuclear War

Metropolitan-Area Disaster

Building-level Mishap

Administrative Assault

Implementing Disaster Recovery

■ Two major questions:

1. What disaster are you trying to survive?
2. Have you done a Business Impact Analysis?

Business Impact Analysis

- If you haven't done a BIA,
  you don't know how much
    time,
    effort, and
    money
  to spend on your business continuity plan.

Disaster Tolerance Basics

- TANSTAAFL
  - Redundancy is needed
    - 2N
    - N+1
  - Need to manage the redundancy
  - Vigilance
    - Errors
    - Configuration changes
    - What's normal

Disaster Tolerance Basics

- Distance
  - Bandwidth
  - Latency
  - Partitioning

Bandwidth

- Affects
  - Large transfers
  - Backups
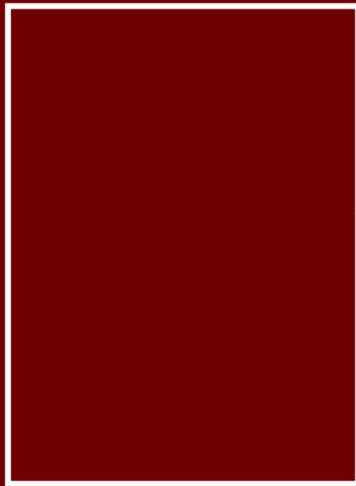 - Recovery

Latency

SPEED LIMIT

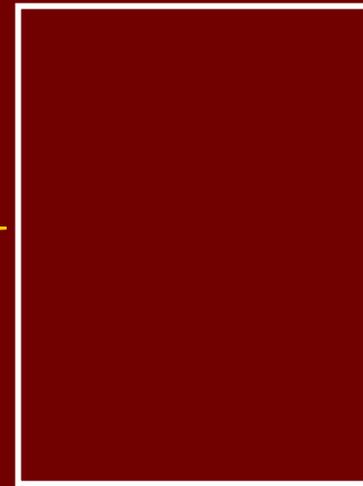3×10^8 m/s

STRICTLY ENFORCED

Latency

- Affects
  - Request rate
  - Small transfers

- 1ms/100mi

- Buffering to "decrease" latency is a compromise masquerading as a solution

Partitioning

Site 1

Site 2

Communications Link

Partitioning

- If site 1 fails,
  then no messages are received by site 2
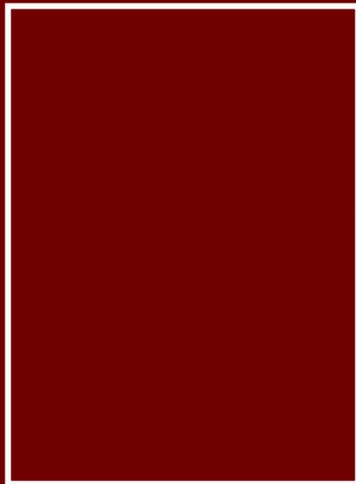- Must be time to make site 2 operational

Site 1

Site 2

Communications
Link

Partitioning

- If the communications link fails, then site 2 receives no messages
- It is NOT time to make site 2 operational

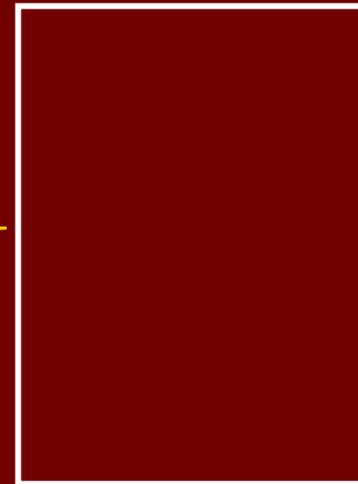**Site 1**

Site 2

Communications
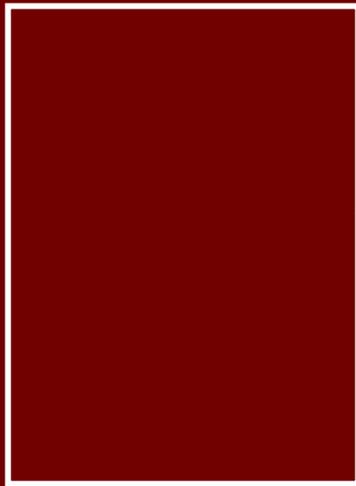Link

Partitioning

Site 1

Site 2

Communications

Link

Partitioning

Site 1

Site 2

Communications
Link

Selling It

- Is not easy.
  - No one wants to discuss doom and gloom
  - No heroes
  - Urgency is hard to instill (or maintain)
  - It is selling insurance
  - ROI is hard to identify
- BIA can open the door
- Key phrase to listen for "We can't let that happen."

Summary

- If it were easy, everybody would do it.